ÁREA TEMÁTICA: Administração Pública, Governo e Terceiro Setor
CATEGORIA: ENANGRAD Pleno

DA PREDIÇÃO À PREVENÇÃO DA EVASÃO NO ENSINO SUPERIOR: REVISÃO SISTEMÁTICA SOBRE INTELIGÊNCIA ARTIFICIAL E APRENDIZADO DE MÁQUINA NO ENFRENTAMENTO

Resumo

A evasão no ensino superior representa um desafio de proporções globais, com taxas que superam 40% e, no Brasil, atingem 34,9% na modalidade a distância. Embora os avanços em Inteligência Artificial e Aprendizado de Máquina tenham ampliado as possibilidades de predição do abandono estudantil, a literatura ainda carece de sínteses que articulem modelos preditivos a ciclos institucionais de intervenção e de prevenção efetiva. Este artigo tem como objetivo analisar como essas tecnologias têm sido aplicadas à previsão e ao apoio à prevenção da evasão no ensino superior. Para isso, foi conduzida uma Revisão Sistemática da Literatura, seguindo o protocolo PRISMA, com buscas nas bases Scopus e Web of Science, resultando em um corpus final de 24 artigos, analisados por meio de um protocolo de extração padronizada e de uma matriz de síntese. Os resultados indicam que a literatura tem avançado na identificação de estudantes em risco, especialmente por meio de modelos preditivos baseados em variáveis acadêmicas de trajetória e estratégias de avaliação adequadas a bases desbalanceadas. Esse avanço, contudo, ainda se concentra predominantemente na classificação de risco, enquanto a transição para sistemas de alerta precoce e intervenções institucionais integradas permanece incipiente. Conclui-se que a capacidade técnica de predição avançou mais rapidamente do que a capacidade institucional de converter previsões em ações sistemáticas de prevenção. Como contribuição central, o artigo propõe um framework conceitual orientado por evidências, destinado a integrar dados, modelos preditivos, alertas, decisões humanas e intervenções institucionais em um ciclo contínuo de apoio à permanência estudantil.

Palavras-chave: Evasão discente. Ensino superior. Inteligência artificial. Aprendizado de máquina. Modelos preditivos.

Abstract

Dropout in higher education is a global challenge, with rates exceeding 40% and reaching 34.9% in distance education in Brazil. Although advances in Artificial Intelligence and Machine Learning have expanded the

possibilities for predicting student dropout, the literature still lacks syntheses that articulate predictive models in relation to institutional cycles of intervention and effective prevention. This article aims to analyze how these technologies have been applied to predicting and supporting the prevention of dropout in higher education. To this end, a Systematic Literature Review was conducted following the PRISMA protocol, with searches in the Scopus and Web of Science databases, yielding a final corpus of 24 articles, which were analyzed using a standardized extraction protocol and a synthesis matrix. The results indicate that the literature has advanced in identifying students at risk, especially through predictive models based on academic trajectory variables and evaluation strategies better suited to imbalanced datasets. However, this progress remains predominantly focused on risk classification, while the transition toward early warning systems and integrated institutional interventions remains incipient. It is concluded that the technical capacity for prediction has advanced more rapidly than the institutional capacity to convert predictions into systematic prevention actions. As its central contribution, the article proposes an evidence-based conceptual framework that integrates data, predictive models, alerts, human decisions, and institutional interventions into a continuous cycle of support for student retention.

Keywords: Student dropout. Higher education. Artificial intelligence. Machine learning. Predictive models.

1. Introdução

A evasão no ensino superior representa um desafio de proporções globais: taxas de abandono superiores a 40% em diversos países (Arévalo-Cordovilla; Peña, 2025) e de 34,9% na modalidade a distância no Brasil (Tete *et al*., 2022) evidenciam a necessidade de estratégias sistemáticas de monitoramento capazes de identificar sinais de risco antes que se convertam em desligamentos definitivos.

Diante desse cenário, a evasão deixa de ser apenas um problema de diagnóstico retrospectivo e passa a exigir mecanismos capazes de antecipar trajetórias de risco ao longo do percurso acadêmico. Embora a literatura reconheça a multicausalidade do fenômeno (Fernández-Martín *et al*., 2019), este artigo não se propõe a revisar exaustivamente suas causas gerais, mas sim a analisar abordagens tecnológicas de predição e de suporte à prevenção, fundamentadas em Inteligência Artificial, Aprendizado de Máquina, Mineração de Dados Educacionais e Learning Analytics.

A justificativa para esta revisão sistemática parte de uma lacuna estrutural identificada na literatura: embora os avanços técnicos em modelos preditivos sejam expressivos, os estudos tendem a tratar a aplicação da

Inteligência Artificial no ensino superior de forma fragmentada, com ênfase nas métricas de desempenho dos algoritmos e menor atenção à explicabilidade, à utilidade prática para gestores e educadores e à articulação entre predição, monitoramento e intervenção institucional (Arévalo-Cordovilla; Peña, 2025; Baneres; Rodríguez-González; Guerrero-Roldán, 2023; Tete *et al*., 2022).

É nessa distinção que reside o problema central desta pesquisa: prever a evasão não equivale, por si só, a mitigá-la. A existência de um modelo preditivo preciso é condição necessária, mas insuficiente, se a instituição não dispuser de capacidade operacional para traduzir dados em intervenções humanizadas, oportunas e contextualizadas. A partir dessa lacuna, este estudo é orientado pela seguinte questão de pesquisa: de que maneira a literatura científica tem utilizado Inteligência Artificial e Aprendizado de Máquina para prever e apoiar a prevenção da evasão no ensino superior, e em que medida os modelos desenvolvidos avançam da predição à proposição de intervenções institucionais efetivas? Especificamente, busca-se identificar técnicas e algoritmos predominantes, mapear os dados e as variáveis utilizados, caracterizar as métricas de avaliação, examinar a presença de sistemas de alerta, indicadores de risco e intervenções práticas, bem como identificar lacunas recorrentes para pesquisas futuras.

A investigação delimita-se à análise de estudos científicos publicados entre 2016 e 2025 que abordam a aplicação de Inteligência Artificial, Aprendizado de Máquina, Mineração de Dados Educacionais e Learning Analytics à predição, ao monitoramento ou ao apoio à prevenção da evasão no ensino superior. Assim, não se pretende desenvolver ou validar empiricamente um novo modelo preditivo, nem realizar uma revisão abrangente de todas as causas da evasão universitária. O foco recai sobre a produção científica que articula técnicas computacionais, variáveis, métricas, sistemas de alerta, indicadores de risco e possibilidades de aplicação institucional voltadas à permanência estudantil.

A contribuição esperada deste artigo é dupla. No plano analítico, a revisão organiza criticamente como a IA e o Aprendizado de Máquina têm sido aplicados à evasão no ensino superior, sistematizando técnicas, variáveis, métricas e aplicações práticas em categorias comparáveis. No plano propositivo, busca delinear elementos mínimos que um modelo orientado à prevenção — e não apenas à predição — deveria contemplar, contribuindo para uma base conceitual aplicável ao desenvolvimento futuro de indicadores ou scores de risco no contexto institucional brasileiro.

Este artigo está organizado em cinco seções. Após esta Introdução, a Fundamentação Teórica apresenta os principais conceitos que sustentam a pesquisa. A Metodologia descreve o percurso da revisão sistemática da literatura. A seção de Resultados e Discussão apresenta uma síntese crítica

dos estudos selecionados. Por fim, as Considerações Finais retomam a questão de pesquisa e os objetivos, sintetizam os achados, discutem as limitações e indicam caminhos para pesquisas futuras.

2. Fundamentação Teórica

2.1 Evasão no ensino superior e permanência estudantil

A evasão no ensino superior constitui fenômeno complexo, multicausal e de amplo impacto, com efeitos individuais, institucionais, sociais e econômicos, resultante da combinação de fatores acadêmicos, socioeconômicos, institucionais, comportamentais e pessoais, que variam conforme o contexto educacional, a modalidade, o curso, o perfil discente e as condições de permanência (Fernández-Martín *et al*., 2019; Quintana Quispe *et al*., 2024; Rabelo; Zárate, 2025; Tete *et al*., 2022). Por isso, não deve ser compreendida apenas como evento final de desligamento ou como uma trajetória progressiva de vulnerabilidade, o que desloca o foco das análises retrospectivas para abordagens capazes de identificar sinais antecipados de risco e subsidiar ações direcionadas por gestores, docentes e equipes de apoio antes que o abandono se materialize (Baneres; Rodríguez-González; Guerrero-Roldán, 2023; Rabelo; Zárate, 2025).

2.2 Inteligência Artificial, modelos preditivos e sistemas de alerta precoce

A aplicação de Inteligência Artificial, Aprendizado de Máquina, Mineração de Dados Educacionais e Learning Analytics ampliou a capacidade de analisar e gerir a evasão no ensino superior, ao permitir o processamento de dados educacionais, a identificação de padrões complexos e o apoio à tomada de decisão institucional (Arévalo-Cordovilla; Peña, 2025; Rabelo; Zárate, 2025). Nesse campo, modelos supervisionados, como árvores de decisão, regressão logística, Random Forest, Support Vector Machine, K-Nearest Neighbors, redes neurais e algoritmos de boosting, são utilizados para estimar, com base em dados históricos e atuais, a probabilidade de abandono do curso, da instituição ou de atividades acadêmicas. Esses modelos demandam variáveis acadêmicas, demográficas, socioeconômicas, administrativas e comportamentais, como desempenho, frequência, créditos, reprovações, situação financeira, forma de ingresso e interações em ambientes virtuais (Rabelo; Zárate, 2025; Rovira; Puertas; Igual, 2017; Tete *et al*., 2022). Embora parte dos estudos enfatize métricas como acurácia, precisão, recall, F1-score e AUC, há avanço para scores de risco, dashboards, monitoramento em tempo real e sistemas de alerta precoce, cuja efetividade depende da qualidade dos dados, da interpretabilidade dos resultados e de fluxos institucionais capazes de converter alertas em ações concretas de permanência estudantil (Baneres; Rodríguez-González; Guerrero-Roldán, 2023; Tete *et al*., 2022; Villar; Andrade, 2024).

2.3 Da predição à prevenção: explicabilidade, decisão humana e intervenção institucional

Um ponto central na literatura sobre IA e evasão no ensino superior é a distinção entre prever e prevenir. Embora modelos preditivos identifiquem estudantes em risco, essa indicação só contribui para a permanência quando seus resultados são compreensíveis, acionáveis e integrados a estratégias institucionais de acompanhamento. Nesse sentido, a explicabilidade torna-se fundamental, pois modelos precisos, mas pouco interpretáveis, podem limitar a confiança de gestores e educadores. Técnicas como importância de variáveis, SHAP e LIME ampliam a transparência e apoiam decisões mais informadas, mas não substituem a ação humana. Assim, scores de risco, alertas e dashboards devem estar vinculados a protocolos de atendimento, responsáveis definidos, comunicação adequada, encaminhamentos pedagógicos e psicossociais, monitoramento dos resultados e revisão contínua dos critérios utilizados (Martín-Caro Álamo, 2024; Rabelo; Zárate, 2025).

2.4 Síntese conceitual da pesquisa

A fundamentação teórica desta pesquisa sustenta-se em quatro pressupostos articulados: a evasão no ensino superior é um fenômeno multicausal e progressivo; IA (Aprendizado de Máquina), Mineração de Dados Educacionais e Learning Analytics oferecem recursos relevantes para sua análise preditiva; a qualidade técnica dos modelos é necessária, mas insuficiente para garantir utilidade institucional; e a prevenção efetiva requer integração entre predição, explicabilidade, monitoramento e intervenção humana contextualizada. Com base nesses pressupostos, este artigo analisa a literatura científica recente buscando compreender não apenas quais algoritmos, variáveis e métricas têm sido utilizados, mas também em que medida os estudos avançam na direção de sistemas de alerta, indicadores de risco e estratégias institucionais capazes de transformar dados em ações orientadas à permanência estudantil.

3. Metodologia

Este artigo caracteriza-se como uma revisão sistemática da literatura, de abordagem qualitativa, com apoio de procedimentos quantitativo-descritivos para a organização dos registros recuperados, a seleção dos estudos e a sistematização dos achados. A revisão foi desenvolvida com o objetivo de mapear, analisar e sintetizar estudos sobre o uso de Inteligência Artificial, Machine Learning, Educational Data Mining, Learning Analytics e abordagens correlatas na previsão, no monitoramento e no apoio à prevenção da evasão no ensino superior. O percurso metodológico foi orientado pelas etapas de identificação, triagem, elegibilidade e inclusão dos estudos, em consonância com a lógica do protocolo PRISMA, adotado como referência para conferir

maior transparência, rastreabilidade e reprodutibilidade ao processo de revisão.

As buscas bibliográficas foram realizadas em 28 de maio de 2026 nas bases de dados Scopus e Web of Science, selecionadas por sua ampla cobertura de publicações científicas e por sua utilização recorrente em revisões sistemáticas. O recorte temporal abrangeu o período de 2016 a 2025, considerando apenas os anos completos de indexação. Foram incluídos exclusivamente artigos científicos publicados em inglês, espanhol ou português. A estratégia de busca foi estruturada em três blocos temáticos: termos relacionados à evasão discente, descritores referentes ao ensino superior e palavras-chave associadas a tecnologias preditivas e analíticas. As expressões foram adaptadas à sintaxe de cada base, mas mantiveram o mesmo escopo conceitual, de modo a recuperar estudos relacionados a modelos preditivos, sistemas de alerta precoce, indicadores de risco e aplicações de IA, ML, EDM ou Learning Analytics no contexto da evasão universitária. A busca retornou 257 artigos na Scopus e 179 na Web of Science, totalizando 436 registros.

Tabela 1 – Estratégias de busca adotadas

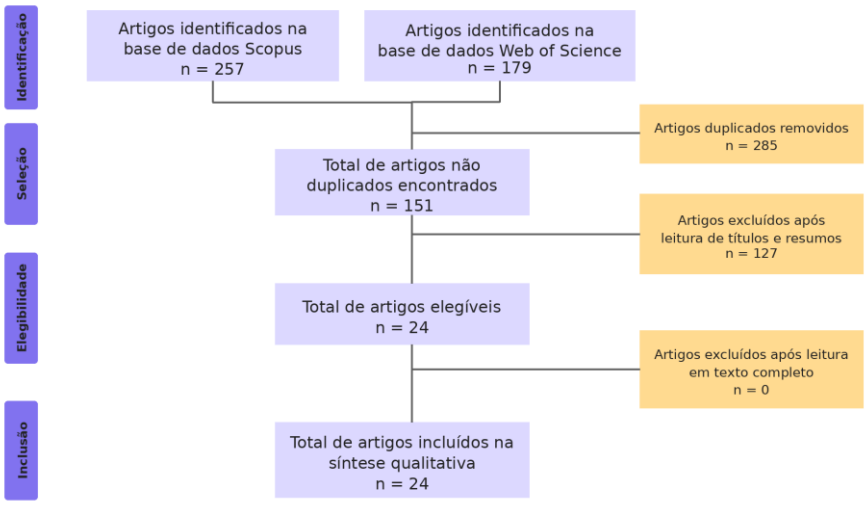
Base de dados	String	Número de artigos
Scopus	TITLE-ABS-KEY (("higher education dropout" OR "university dropout" OR "college dropout" OR "student dropout" OR "student attrition" OR "student desertion" OR "academic dropout" OR "university attrition") AND ("higher education" OR universit* OR college* OR undergraduate*) AND ("artificial intelligence" OR "machine learning" OR "educational data mining" OR "data mining" OR "learning analytics" OR "predictive model*" OR "predictive analytic*" OR "dropout prediction" OR "student dropout prediction" OR "early warning system*" OR "risk prediction" OR "risk indicator*" OR "risk score*" OR "classification model*")) AND PUBYEAR > 2015 AND PUBYEAR < 2026 AND (LIMIT-TO (DOCTYPE, "ar")) AND (LIMIT-TO (LANGUAGE, "English") OR LIMIT-TO (LANGUAGE, "Spanish") OR LIMIT-TO (LANGUAGE, "Portuguese"))	257
Web of Science	TS=(("higher education dropout" OR "university dropout" OR "college dropout" OR "student dropout" OR "student attrition" OR "student desertion" OR "academic dropout" OR "university attrition") AND ("higher education" OR universit* OR college* OR undergraduate*) AND ("artificial intelligence" OR "machine learning" OR "educational data mining" OR "data mining" OR "learning analytics" OR "predictive model*" OR "predictive analytic*" OR "dropout prediction" OR "student dropout prediction" OR "early warning system*" OR "risk prediction" OR "risk indicator*" OR "risk score*" OR "classification model*")) AND PY=(2016 OR 2017 OR 2018 OR 2019 OR 2020 OR 2021 OR 2022 OR 2023 OR 2024 OR 2025) AND DT=(Article) AND LA=(English OR Spanish OR Portuguese)	179

Fonte: elaborada pelos autores.

Os critérios de inclusão e exclusão foram definidos para manter o corpus alinhado ao objetivo da revisão. Foram incluídos artigos que abordassem a evasão no ensino superior e apresentassem o uso, a análise ou a discussão de técnicas de Inteligência Artificial, Machine Learning, Educational Data Mining, Data Mining, Learning Analytics, modelos preditivos, sistemas de alerta precoce ou indicadores de risco. Foram excluídos estudos voltados a outros níveis de ensino, trabalhos em que a evasão aparecia apenas de forma periférica, textos sem relação com abordagens preditivas ou analíticas, revisões sem aderência direta ao foco tecnológico da pesquisa e documentos sem informações suficientes para a extração de objetivos, técnicas utilizadas, variáveis, resultados ou contribuições.

Após a realização das buscas, os registros foram exportados no formato RIS e importados no Zotero, utilizado para a organização das referências e a verificação de duplicidade. Dos 436 registros identificados, 285 duplicados foram removidos, restando 151 artigos para triagem. Em seguida, procedeu-se à leitura de títulos e resumos com base nos critérios previamente definidos. Nessa etapa, 127 artigos foram excluídos por baixa aderência ao foco da revisão, restando 24 estudos para avaliação de elegibilidade. Após a leitura e conferência dos textos elegíveis, não houve novas exclusões, de modo que os 24 artigos compuseram o corpus final da revisão sistemática. O fluxo completo de identificação, triagem, elegibilidade e inclusão foi organizado conforme o protocolo PRISMA.

Figura 1 – Fluxograma PRISMA do processo de identificação, triagem, elegibilidade e inclusão dos estudos.



Fonte: elaborada pelos autores.

Para a análise dos dados, foi adotado um protocolo de extração padronizada, operacionalizado por meio de uma matriz de síntese dos estudos incluídos. Cada artigo foi registrado individualmente, incluindo informações sobre autoria, ano, contexto institucional ou geográfico, objetivo, tipo de pesquisa, dados utilizados, técnicas ou modelos empregados, variáveis consideradas, métricas de avaliação, resultados, aplicações práticas, limitações e contribuições para esta revisão. Esse procedimento permitiu organizar os estudos de forma comparável e reduzir a dispersão analítica típica de revisões com objetos metodologicamente heterogêneos.

Tabela 2 – Campos utilizados na extração e síntese dos estudos incluídos na revisão

Campo utilizado na extração	**Finalidade na revisão**
Autor, ano e contexto	Identificar a autoria, o período de publicação e a localização ou o contexto institucional do estudo.
Objetivo do estudo	Delimitar o foco de cada artigo analisado e sua relação com a evasão, a predição ou a prevenção.
Tipo de pesquisa	Caracterizar o desenho metodológico adotado nos estudos incluídos no corpus.
Dados utilizados	Identificar as bases, as amostras, os períodos, as fontes institucionais e os tipos de dados analisados.
Técnicas/modelos empregados	Mapear algoritmos, modelos e abordagens de IA, Machine Learning, EDM ou Learning Analytics.
Variáveis consideradas	Identificar fatores preditores utilizados nos modelos, tais como dados acadêmicos, socioeconômicos, demográficos, comportamentais e institucionais.
Métricas de avaliação	Comparar as medidas de desempenho preditivo, como acurácia, precisão, recall, F1-score, AUC-ROC e matriz de confusão.
Aplicações práticas	Verificar usos institucionais, sistemas de alerta, indicadores de risco, painéis ou apoios à decisão.
Relação entre predição e mitigação	Avaliar se o estudo avança da previsão do risco para propostas de prevenção, de intervenção ou de permanência estudantil.
Lacunas/limitações	Identificar os limites metodológicos, técnicos, éticos, institucionais ou de generalização apontados no corpus.

Fonte: elaborada pelos autores.

A partir da matriz de síntese, os estudos foram analisados transversalmente para identificar padrões, convergências, diferenças e lacunas no uso de Inteligência Artificial e de Machine Learning aplicados à evasão no ensino superior. A análise foi estruturada em seis dimensões: técnicas e algoritmos utilizados; variáveis e dados considerados; métricas de avaliação; sistemas de alerta e aplicações práticas; relação entre predição e mitigação da evasão; e lacunas recorrentes na literatura. Essa organização permitiu que a revisão não se restringisse à descrição isolada dos artigos, mas também subsidiasse uma síntese crítica sobre o modo como os modelos preditivos têm sido utilizados e em que medida suas aplicações se aproximam de ações institucionais voltadas à prevenção da evasão e à permanência estudantil.

4. Análise e Discussão dos Resultados

O corpus final desta revisão revela a consolidação recente do uso de Inteligência Artificial (IA), Aprendizado de Máquina (ML), Educational Data Mining (EDM) e Learning Analytics (LA) no enfrentamento da evasão no ensino superior. Predominam estudos empíricos voltados à construção e à comparação de modelos preditivos, sobretudo a partir de 2020, em diferentes contextos institucionais e nacionais. Em termos gerais, a literatura demonstra avanço técnico consistente na capacidade de identificar estudantes em risco, mas ainda evidencia fragilidade na conversão dessa capacidade em políticas, fluxos e práticas institucionais de prevenção.

Do ponto de vista metodológico, sobressaem-se algoritmos de classificação supervisionados, como Regressão Logística, Árvores de Decisão, Random Forest, Support Vector Machine, K-Nearest Neighbors, Naive Bayes e, mais recentemente, modelos ensemble e algoritmos de boosting, como XGBoost, LightGBM e CatBoost (Agrusti; Bonavolontà; Mezzini, 2019; Rabelo; Zárate, 2025; Villar; Andrade, 2024). Embora a literatura reporte bons níveis de desempenho, o principal interesse gerencial desses modelos não se limita à acurácia, mas também à utilidade de identificar oportunamente estudantes vulneráveis. Por isso, métricas como recall, F1-score e AUC-ROC tornam-se particularmente relevantes, sobretudo em bases desbalanceadas, nas quais o custo institucional de não identificar um estudante em risco tende a ser maior do que o de acompanhar preventivamente casos sinalizados em excesso.

As variáveis mais recorrentes concentram-se em dados acadêmicos e administrativos, como desempenho, médias, reprovações, frequência, créditos cursados e aprovados, situação de matrícula, trancamentos e cancelamentos, frequentemente combinados com informações demográficas, socioeconômicas, financeiras e comportamentais, incluindo interações em ambientes virtuais de aprendizagem (Al Hashmi *et al.*, 2025; Baneres; Rodríguez-González; Guerrero-Roldán, 2023; Berka; Marek, 2021). Esse padrão sugere que, no contexto brasileiro, uma parcela expressiva dos insumos necessários à predição já está disponível nos sistemas institucionais. A principal barreira, portanto, parece residir menos na inexistência de dados do que na fragmentação entre sistemas, na ausência de padronização, na qualidade dos registros e na necessidade de uma governança informacional compatível com a LGPD.

A análise do corpus também indica um deslocamento conceitual relevante: a evasão deixa de ser tratada apenas como desfecho final e passa a ser compreendida como um processo progressivo de vulnerabilidade acadêmica, social e institucional. Nessa perspectiva, sinais como baixo desempenho inicial, irregularidade acadêmica, não entrega de atividades, queda no engajamento, dificuldades financeiras e enfraquecimento do vínculo

institucional tornam-se indicadores úteis para uma lógica de monitoramento contínuo e de intervenção precoce. Esse movimento aproxima-se da predição de uma função gerencial mais ampla, na qual os modelos atuam como instrumentos de apoio à tomada de decisão e à priorização dos esforços institucionais.

Esse deslocamento conceitual, da evasão como desfecho para a evasão como processo, tem implicações diretas sobre o tipo de contribuição que os estudos buscam oferecer: se a evasão é um processo, monitorá-la exige mais do que classificar corretamente quem evadiu; exige também estruturas capazes de agir sobre o risco enquanto ainda está em formação. É justamente essa variação no tipo de contribuição que a Tabela 3 evidencia.

Tabela 3 – Aplicabilidade institucional dos modelos preditivos no corpus

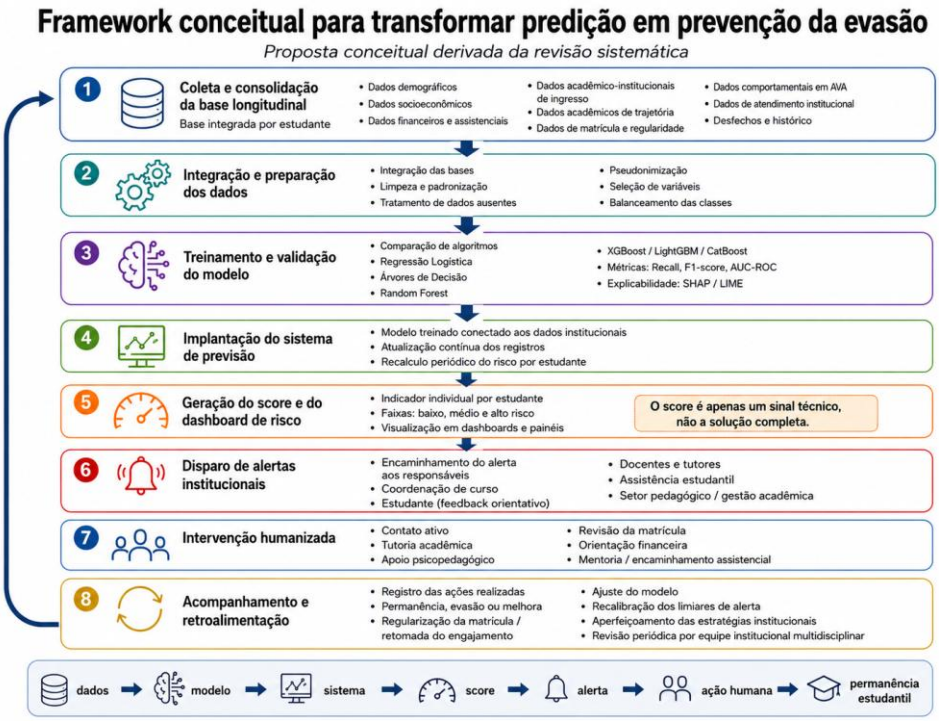
Tipo de contribuição identificada	Critério de classificação	Quantidade de estudos
Modelagem preditiva empírica	Estudos que desenvolveram, testaram ou compararam modelos preditivos próprios com dados estudantis	18
Revisão, síntese ou discussão estratégica	Estudos que revisam literatura, tecnologias, fatores ou estratégias de retenção sem desenvolver modelo empírico próprio	5
Análise empírica de retenção com discussão de IA/ML	Estudo com análise empírica e discussão sobre IA/ML, sem sistema preditivo individual operacional	1
Sistema, aplicação, dashboard ou ferramenta de apoio à decisão	Estudos que associam o modelo à ferramenta, sistema, aplicação, dashboard, visualização ou apoio operacional à decisão	6
Monitoramento em tempo real ou predição diária	Estudos que trabalham com atualização frequente do risco durante o percurso do estudante	1
Explicabilidade formal do modelo	Estudos que incorporam técnicas explícitas de XAI, como SHAP ou LIME	3
Predição vinculada a ações institucionais de apoio	Estudos que relacionam o risco predito a tutorias, apoio pedagógico, assistência, bem-estar, suporte personalizado, contato institucional ou chatbot	9
Avaliação longitudinal da intervenção após alerta	Estudos que implementam alerta, intervenção institucional e avaliação posterior da redução da evasão	0
Governança operacional dos limiares e estratégias	Estudos com equipe ou comitê institucional formalmente responsável pela revisão periódica dos limiares de risco e das estratégias de intervenção	0

Fonte: elaborada pelos autores

As categorias apresentadas na Tabela 3 não são mutuamente excludentes, pois cada estudo foi classificado com base nas diferentes contribuições identificadas no corpus, e não por um único rótulo. Assim, um mesmo artigo pode aparecer em mais de uma categoria quando, além de desenvolver ou comparar modelos preditivos, também apresenta um sistema, um dashboard, explicabilidade, monitoramento contínuo ou vínculo com ações institucionais. Por isso, as quantidades indicam ocorrências por critério analítico e não a soma absoluta dos 24 estudos. A tabela evidencia que, embora a modelagem preditiva empírica seja predominante, o número de estudos que avançam para aplicações institucionais estruturadas é menor. Esse resultado reforça a lacuna central da revisão: a literatura tem avançado mais na classificação do risco de evasão do que na consolidação de mecanismos institucionais para enfrentá-lo. Portanto, a predição só se aproxima da prevenção quando os alertas são incorporados a fluxos institucionais com responsáveis definidos, protocolos de atendimento, encaminhamentos e acompanhamento dos resultados.

Nessa direção, propõe-se um framework conceitual orientado por evidências, estruturado em oito etapas organizadas em um ciclo institucional de prevenção da evasão (Figura 2).

Figura 2: Framework



Fonte: elaborada pelos autores

As etapas 1 e 2 de coleta e consolidação da base longitudinal e de integração e preparação dos dados sustentam-se na recorrência de variáveis

acadêmicas, socioeconômicas, demográficas e comportamentais identificadas no corpus, bem como na necessidade de tratamento, padronização e pseudonimização desses dados. As etapas 3 e 4 de treinamento e validação do modelo, bem como a implantação do sistema de previsão, refletem a predominância de algoritmos supervisionados e de métricas como recall, F1-score e AUC-ROC, que recalculam periodicamente o risco de cada estudante. A etapa 5 de geração do score e do dashboard de risco converte esse resultado em indicadores individuais e faixas de risco, sempre como sinal técnico de apoio, não como solução completa. As etapas 6 e 7 de disparo de alertas institucionais e de intervenção humanizada direcionam o risco identificado às coordenações, aos docentes, aos tutores e aos setores de assistência estudantil, bem como ao próprio estudante, traduzindo-o em ações como tutoria, apoio psicopedagógico ou orientação financeira. Por fim, a etapa 8 de acompanhamento e retroalimentação registra os resultados das intervenções e retroalimenta o modelo, preferencialmente sob revisão periódica de uma equipe institucional multidisciplinar. Assim, o framework articula dados, tecnologia e ação institucional em um ciclo contínuo, coerente com a gestão universitária e com o uso mais estratégico dos recursos públicos.

5. Conclusão e Contribuições

Este artigo teve como objetivo analisar como a Inteligência Artificial e o Aprendizado de Máquina vêm sendo aplicados à previsão e ao apoio à prevenção da evasão no ensino superior. A revisão sistemática mostrou que a literatura tem utilizado predominantemente essas tecnologias para transformar dados acadêmicos, administrativos, socioeconômicos e comportamentais em modelos de classificação, indicadores e alertas de risco. Os achados confirmam, contudo, que o avanço técnico da predição não foi acompanhado, com a mesma intensidade, de um avanço institucional na prevenção.

Os estudos analisados revelam predominância de modelos supervisionados, centralidade de variáveis acadêmicas e administrativas e crescimento recente do interesse por sistemas de alerta precoce, dashboards, monitoramento contínuo e abordagens explicáveis. Ainda assim, a principal fragilidade da literatura reside na baixa articulação entre os modelos preditivos e os processos institucionais de intervenção, acompanhamento e avaliação de resultados. Assim, IA e ML mostram-se promissoras como instrumentos de apoio à decisão, mas sua efetividade depende da integração com os fluxos administrativos, as equipes multidisciplinares e as políticas de permanência estudantil.

A contribuição teórica do artigo consiste em organizar criticamente a literatura sobre IA e ML aplicadas à evasão no ensino superior, evidenciando padrões, convergências e lacunas. A contribuição prática e gerencial reside na proposição de um framework conceitual voltado à gestão da permanência,

capaz de orientar gestores, coordenações de curso, docentes, tutores, assistência estudantil e setores pedagógicos no uso responsável dos dados para apoiar intervenções preventivas. Em termos sociais e públicos, a revisão reforça o potencial dessas tecnologias para qualificar a gestão universitária, reduzir desperdícios de recursos educacionais e fortalecer políticas de equidade e de permanência.

Como limitações, este estudo restringiu-se à análise de artigos científicos publicados entre 2016 e 2025, recuperados nas bases de dados Scopus e Web of Science, além de não realizar validação empírica própria do framework proposto. Soma-se a isso a necessidade de adaptação contextual do modelo às especificidades institucionais, considerando as diferenças nos sistemas de informação, nas políticas de permanência, na disponibilidade de dados e na capacidade operacional de intervenção. Pesquisas futuras devem avançar na validação do framework em instituições brasileiras, no desenvolvimento de indicadores explicáveis, na condução de estudos longitudinais, na avaliação dos efeitos das intervenções institucionais e na análise dos desafios éticos, legais e operacionais relacionados à governança de dados e à LGPD. Em síntese, o desafio contemporâneo não se limita a aprimorar a capacidade preditiva dos modelos, mas também envolve transformar essas predições em ações institucionais éticas, oportunas e efetivas de apoio à permanência estudantil.

Referências

- AGRUSTI, F.; BONAVOLONTÀ, G.; MEZZINI, M. University dropout prediction through educational data mining techniques: a systematic review. Journal of E-Learning and Knowledge Society, v. 15, n. 3, p. 161-182, 2019. DOI: 10.20368/1971-8829/1135017.
- AL HASHMI, R. A. M.; ZERVOPOULOS, P. D.; ELMEHDI, H. M.; OZTURK, I. Predicting dropout in MENA STEM higher education using explainable AI: a machine learning approach. Emerging Science Journal, v. 9, n. Especial, p. 268-286, 2025. DOI: 10.28991/ESJ-2025-SIED1-016.
- ARÉVALO-CORDOVILLA, F. E.; PEÑA, M. AI-driven predictive models and chatbots for early intervention and student success in higher education: a systematic review. International Journal of Engineering Education, v. 41, n. 6, p. 1473-1488, 2025.
- BANERES, D.; RODRÍGUEZ-GONZÁLEZ, M. E.; GUERRERO-ROLDÁN, A. E. A real-time predictive model for identifying course dropout in online higher education. IEEE Transactions on Learning Technologies, v. 16, n. 4, p. 484-499, 2023. DOI: 10.1109/TLT.2023.3267275.
- BERKA, P.; MAREK, L. Bachelor's degree student dropouts: who tend to stay and who tend to leave? Studies in Educational Evaluation, v. 70, art. 100999, 2021. DOI: 10.1016/j.stueduc.2021.100999.

- CUZME ROMERO, M. G.; MEZA, J.; GARCÍA, R. Predicting student dropout risk using machine learning: a case study at the Technical University of Manabí. Kasetsart Journal of Social Sciences, v. 46, n. 2, art. 460227, 2025. DOI: 10.34044/j.kjss.2025.46.2.27.
- VALERO CAJAHUANCA, J. E.; NAVARRO RAYMUNDO, Á. F.; LARIOS FRANCO, A. C.; JULCA FLORES, J. D. Deserción universitaria: evaluación de diferentes algoritmos de machine learning para su predicción. Revista de Ciencias Sociales, v. 28, n. 3, p. 362-375, 2022.
- FERNÁNDEZ-MARTÍN, T.; SOLÍS-SALAZAR, M.; HERNÁNDEZ-JIMÉNEZ, M. T.; MOREIRA-MORA, T. E. Un análisis multinomial y predictivo de los factores asociados a la deserción universitaria. Revista Electrónica Educare, v. 23, n. 1, p. 1-25, 2019. DOI: 10.15359/ree.23-1.5.
- MARTÍN-CARO ÁLAMO, E. Análisis de estrategias innovadoras para retención estudiantil con inteligencia artificial: una perspectiva multidisciplinaria. European Public & Social Innovation Review, v. 9, p. 1-20, 2024. DOI: 10.31637/epsir-2024-440.
- MUTROFIN, S.; GINARDI, R. V. H.; FATICHAH, C.; KURNIAWARDHANI, A. A critical assessment of balanced class distribution problems: the case of predict student dropout. Test Engineering and Management, v. 81, p. 1764-1770, 2019.
- NGUYEN, H. T.; LE, P.; NGUYEN, T. T. T.; SU, A. K. A multivariate analysis of the early dropout using classical machine learning and local interpretable model-agnostic explanations. CTU Journal of Innovation and Sustainable Development, v. 16, n. especial, p. 98-106, 2024. DOI: 10.22144/ctujoisd.2024.327.
- QUINTANA QUISPE, J. O.; CUENTAS TOLEDO, O.; CUENTAS TOLEDO, M.; CALIZAYA LLATASI, E. E.; RAMOS SAIRA, E. M. Early prediction of university student dropout using machine learning models. Nanotechnology Perceptions, v. 20, n. S5, p. 659-669, 2024.
- RABELO, A. M.; ZÁRATE, L. E. A model for predicting dropout of higher education students. Data Science and Management, v. 8, n. 1, p. 72-85, 2025. DOI: 10.1016/j.dsm.2024.07.001.
- ROSLAN, N.; JAMIL, J. M.; SHAHARANEE, I. N. M.; ALAWI, S. J. S. Prediction of student dropout in Malaysian's private higher education institute using data mining application. Journal of Advanced Research in Applied Sciences and Engineering Technology, v. 45, n. 2, p. 168-176, 2025. DOI: 10.37934/araset.45.2.168176.
- ROVIRA, S.; PUERTAS, E.; IGUAL, L. Data-driven system to predict academic grades and dropout. PLoS ONE, v. 12, n. 2, e0171207, 2017. DOI: 10.1371/journal.pone.0171207.
- SALLAN, G.; BEHAL, S. Prediction of student dropout using enhanced machine learning algorithm. Advances in Mathematics: Scientific Journal, v. 9, n. 6, p. 3821-3826, 2020. DOI: 10.37418/amsj.9.6.61.
- SIHARE, S. R. Student dropout analysis in higher education and retention by artificial intelligence and machine learning. SN Computer Science, v. 5, art. 202, 2024. DOI: 10.1007/s42979-023-02458-w.

- SIVAKUMAR, S.; VENKATARAMAN, S.; SELVARAJ, R. Predictive modeling of student dropout indicators in educational data mining using improved decision tree. Indian Journal of Science and Technology, v. 9, n. 4, p. 1-5, 2016. DOI: 10.17485/ijst/2016/v9i4/87032.
- TARMIZI, S. S. A.; MUTALIB, S.; ABDUL HAMID, N. H.; ABDUL RAHMAN, S. A review on student attrition in higher education using big data analytics and data mining techniques. International Journal of Modern Education and Computer Science, v. 11, n. 8, p. 1-14, 2019. DOI: 10.5815/ijmecs.2019.08.01.
- TETE, M. F.; SOUSA, M. M.; SANTANA, T. S.; FELLIPE, S. Aplicação de métodos preditivos em evasão no ensino superior: uma revisão sistemática da literatura. Arquivos Analíticos de Políticas Educativas, v. 30, n. 149, 2022. DOI: 10.14507/epaa.30.6845.
- ACO TITO, A. E.; HANCCO CONDORI, B. O.; PÉREZ VERA, Y. Análisis comparativo de técnicas de machine learning para la predicción de casos de deserción universitaria. RISTI - Revista Ibérica de Sistemas e Tecnologias de Informação, n. 51, p. 84-98, 2023. DOI: 10.17013/risti.51.84-98.
- VEGA, H.; SANEZ, E.; DE LA CRUZ, P.; MOQUILLAZA, S.; PRETELL, J. Intelligent system to predict university students dropout. International Journal of Online and Biomedical Engineering, v. 18, n. 7, p. 27-43, 2022. DOI: 10.3991/ijoe.v18i07.30195.
- VILLAR, A.; ANDRADE, C. R. V. de. Supervised machine learning algorithms for predicting student dropout and academic success: a comparative study. Discover Artificial Intelligence, v. 4, art. 2, 2024. DOI: 10.1007/s44163-023-00079-z.
- ZÁRATE-VALDERRAMA, J.; BEDREGAL-ALPACA, N.; CORNEJO-APARICIO, V. Modelos de clasificación para reconocer patrones de deserción en estudiantes universitarios. Ingeniare. Revista Chilena de Ingeniería, v. 29, n. 1, p. 168-177, 2021.