



# ANÁLISE DE EVASÃO NO ENSINO SUPERIOR A DISTÂNCIA COM MINERAÇÃO E INTEGRAÇÃO DE DADOS EDUCACIONAIS: UMA ABORDAGEM BASEADA EM GRAFOS PARA APOIO À GESTÃO ACADÊMICA

*Silvio Bortoleto*<sup>1</sup>, *Alexandre Rodízio Bento*<sup>2</sup>, *Matheus Mitsuo Yamafuku Benatti*<sup>3</sup>,  
*Luciano Franzoi Filho*<sup>4</sup>, *Leonardo de Lima Póss*<sup>5</sup>, *Rafaela Gonzatto da Silva*<sup>6</sup>

<sup>1</sup>Doutor, Docente do Curso de Engenharia de Software, Campus Curitiba-PR, Universidade Cesumar - UNICESUMAR.  
silvio.bortoleto@unicesumar.edu.br

<sup>2</sup>Mestre, Docente do Curso de Engenharia de Software, Campus Curitiba-PR, Universidade Cesumar - UNICESUMAR.  
alexandre.bento@unicesumar.edu.br

<sup>3</sup>Acadêmico do Curso de Engenharia de Software, Campus Curitiba-PR, Universidade Cesumar - UNICESUMAR.  
matheus.cwb2@gmail.com

<sup>4</sup>Acadêmico, do Curso de Engenharia de Software, Campus Curitiba-PR, Universidade Cesumar - UNICESUMAR.  
lucianofilhodomosco@gmail.com

<sup>5</sup>Acadêmico do Curso de Engenharia de Software, Campus Curitiba-PR, Universidade Cesumar - UNICESUMAR.  
leonardoposs@hotmail.com

<sup>6</sup>Acadêmico do Curso de Engenharia de Software, Campus Curitiba-PR, Universidade Cesumar - UNICESUMAR.  
rafaelagonzatto3@gmail.com

## RESUMO

A evasão no ensino superior a distância configura um desafio persistente no cenário educacional brasileiro, afetando diretamente o desempenho acadêmico institucional e as trajetórias de formação dos estudantes. Diante de um ambiente cada vez mais digitalizado e orientado por dados, compreender os fatores que influenciam a permanência discente torna-se essencial para o aprimoramento da gestão acadêmica e a formulação de estratégias de intervenção. Para investigar os elementos que caracterizam os estudantes com maior propensão à evasão, técnicas de mineração de dados supervisionada podem ser aplicadas de modo que gere conhecimento sobre informações acadêmicas, financeiras, comportamentais e demográficas. Além de extrair conhecimento, exibir de forma dinâmica e que permita ter uma visão macro e de rápida interpretação pode ser feita através de grafos, possibilitando a visualização de conexões e comunidades acadêmicas propensas à evasão. Os resultados apontaram a centralidade das variáveis comportamentais, especialmente relacionadas à frequência de acesso às plataformas de ensino, como principais indicadoras de risco, superando fatores acadêmicos e socioeconômicos. A modelagem em grafos evidenciou agrupamentos de estudantes em situações semelhantes, destacando também indivíduos potencialmente em risco futuro devido às conexões estabelecidas com grupos críticos. Essas evidências reforçam a importância de abordagens integradas, que unam atributos individuais às dinâmicas coletivas e institucionais, favorecendo ações de gestão mais estratégicas e alinhadas às especificidades do ensino superior a distância.

**PALAVRAS-CHAVE:** Evasão; Ensino Superior; Grafos; Gestão; Mineração de Dados.

## 1 INTRODUÇÃO

Nas últimas décadas, o acesso ao ensino superior no Brasil tem crescido significativamente, impulsionado por políticas públicas de inclusão e expansão da rede de instituições. No entanto, esse avanço tem sido acompanhado por um problema persistente: a alta taxa de evasão de estudantes, tanto em cursos presenciais de instituições públicas e privadas quanto em cursos à distância. A evasão representa não apenas desperdício de recursos, mas também a interrupção de trajetórias acadêmicas. Compreender os fatores associados à permanência e ao abandono no ensino superior é um tema que abrange âmbitos educacionais, sociais e econômicos do país (LOBO, 2012).

Paralelamente, o volume e a granularidade dos dados educacionais têm aumentado consideravelmente, resultado da intensificação dos processos de digitalização e armazenamento sistemático de informações no contexto acadêmico. Esses dados oferecem uma oportunidade valiosa para análises aprofundadas sobre o comportamento



acadêmico dos estudantes e o perfil das instituições de ensino. Em muitas situações estas informações podem estar armazenadas de maneiras não estruturadas, o que dificulta a consulta e análise. Apesar disso, mesmo dados não estruturados apresentam um grau mínimo de relação de informações. Nestas situações, a estrutura em formato de rede pode ser um artefato interessante para extração de informação e conhecimento (OLIVEIRA, 2021).

Este artigo explora a interseção entre mineração de dados educacionais, gestão acadêmica e tecnologias de representação baseadas em grafos, em um contexto de profundas transformações no ensino superior. Portanto, esse trabalho contribui para o debate sobre o uso ético e estratégico dos dados na educação e fomentar ações que permitam desenvolver soluções mais inteligentes, inclusivas e eficientes para enfrentar a evasão. O futuro da educação será cada vez mais orientado por dados, e cabe às instituições se prepararem para utilizar essas inovações de forma responsável, promovendo permanência, equidade e sucesso acadêmico.

## 2 MATERIAIS E MÉTODOS

Este estudo adotou uma abordagem descritiva, com o objetivo de identificar padrões, descrever características e estabelecer possíveis correlações entre variáveis relacionadas à evasão no ensino superior a distância. A pesquisa foi estruturada em cinco etapas complementares. Inicialmente, realizou-se uma revisão teórica sobre o fenômeno da evasão no ensino superior, com o levantamento de publicações acadêmicas que abordam os principais fatores associados à permanência e abandono estudantil. Em seguida, procedeu-se à coleta, extração e preparação dos dados, utilizando registros acadêmicos institucionais. Os dados passaram por processos de pré-processamento, envolvendo a limpeza de inconsistências, padronização de formatos e transformação de variáveis, a fim de garantir a qualidade e a integridade das informações para as análises subsequentes.

Na terceira etapa, aplicaram-se técnicas de Mineração de Dados utilizando a plataforma Weka, com o objetivo de identificar padrões comportamentais e acadêmicos dos estudantes, com foco na detecção de indicadores relevantes para a previsão da evasão. Posteriormente, foi realizada a modelagem e análise de grafos, empregando conceitos da Teoria dos Grafos para representar visualmente as relações entre os discentes, os níveis de risco de evasão e as variáveis de maior impacto identificadas nas análises supervisionadas. Por fim, os resultados obtidos nas etapas anteriores foram analisados de forma integrada, permitindo a identificação de fatores críticos associados à evasão e a proposição de possíveis estratégias de intervenção e monitoramento acadêmico voltadas à permanência estudantil.

A Mineração de dados foi utilizada para a análise de grandes volumes de dados educacionais, onde configura-se como uma estratégia essencial para enfrentar desafios recorrentes no ensino superior, a exemplo da evasão. A Mineração de Dados Educacionais (*Educational Data Mining* — EDM) constitui um campo interdisciplinar que combina técnicas de ciência de dados, estatística e inteligência artificial para identificar padrões ocultos e relações significativas em bases de dados produzidas por instituições de ensino (ROMERO; VENTURA, 2020). No âmbito da evasão, a EDM possibilita mapear fatores de risco e antecipar comportamentos de abandono com base em informações históricas e contextuais dos estudantes (OLIVEIRA, 2021).

Por meio de técnicas supervisionadas de mineração de dados, torna-se viável identificar os atributos mais relevantes para a previsão do risco de evasão no ensino

superior a distância. Para este estudo, utilizou-se uma base de dados contendo informações acadêmicas, financeiras, comportamentais e demográficas de estudantes



matriculados em cursos EAD. O conjunto de dados foi submetido a procedimentos de pré-processamento, incluindo normalização, limpeza e padronização, com o objetivo de assegurar sua consistência e integridade.

A fim de avaliar a relevância dos atributos no contexto da evasão, aplicou-se o método de Ganho de Informação, que mensura a contribuição individual de cada variável para a classificação da variável-alvo, definida como “Risco de Evasão”. Essa variável foi construída com base nos estados institucionais de acompanhamento acadêmico, organizados em categorias como “Acompanhado”, “Não Acessado”, “Crítico”, “Alerta”, “Alto Risco” e “Risco”. Valores mais elevados de ganho de informação indicam maior capacidade discriminativa da variável na previsão do risco de evasão, orientando a seleção de atributos para modelagens preditivas subsequentes.

**Tabela 1 – Peso Informacional das variáveis**

Peso Informacional	Variável
1.76783	QTD_DIAS_SEM_ACESSO_MODULO
1.66333	QTD_DIAS_DESDE_ULTIMO_ACESSO_OU_MATRICULA
0.44008	PCT_ACERTO_PRIMEIRA_AVALIACAO
0.17625	PCT_CONCLUSAO
0.16233	DISCIPLINAS_FEITAS
0.12967	PARCELAS EM ABERTO
0.12077	VALOR DA INADIMPLENCIA
0.10791	SITUACAO_ACADEMICA
0.09288	STATUS_ACADEMICO
0.09159	SITUACAO_FINANCEIRA
0.05092	DISCIPLINAS_GRADE
0.05024	CURSO
0.04053	FAIXA_CONCLUSAO
0.04013	FAIXA_LIGACOES
0.02773	QUANT DE DISCIPLINAS MATRICULADAS
0.02343	Município
0.01663	FILIAL
0.00956	UF
0.00761	TIPO DE POLO QUE OFERTA SAUDE/ENGENHARIA
0.00491	IDADE
0.00454	SEXO
0.00204	REGIONAL

Fonte: (autores, 2025)

Os resultados, apresentados na Tabela 1, mostraram que os atributos mais relevantes para prever o risco de evasão estão relacionados ao comportamento de acesso dos estudantes à plataforma educacional. As variáveis “Quantidade de dias sem acesso a partir do início do módulo de ingresso” e “Quantidade de dias desde o último acesso” destacaram-se como os principais indicadores de risco de evasão, corroborando estudos que apontam o engajamento contínuo como crucial para a permanência no EAD.

Em contrapartida, variáveis demográficas e institucionais, como “Idade”, “Sexo”, “Tipo de polo” e “Regional”, apresentaram baixo ganho de informação, sugerindo que essas características isoladas não são expressivas para discriminar estudantes em relação ao risco de evasão. Embora importantes para caracterizar o perfil discente, essas variáveis não demonstram relevância significativa para prever o desfecho em questão.

Esses achados não apenas permitem a seleção de atributos relevantes para modelagens preditivas, mas também fundamentam a identificação de perfis de estudantes em risco. Além disso, ressaltam a importância de indicadores comportamentais, especialmente aqueles relacionados à interação com o ambiente virtual de aprendizagem, como insumo estratégico para o monitoramento acadêmico.

Apesar dos avanços recentes, as abordagens tradicionais de Mineração de Dados Educacionais ainda tratam os dados de forma tabular, desconsiderando as relações estruturais presentes no ecossistema acadêmico. Surge a necessidade de reinterpretação de dados educacionais com grafos, pois a modelagem em grafos se sobressai nesse contexto por representar os dados de maneira relacional e dinâmica. Composta por nós, que representam estudantes, disciplinas, professores e demais elementos institucionais, e por



arestas, que descrevem as interações entre esses elementos, essa estrutura possibilita mapear redes e explorar relações implícitas, trajetórias de evasão e padrões não captados por técnicas convencionais (BORBA, 2013).

Essa abordagem viabiliza algoritmos capazes de identificar alunos em risco devido ao isolamento acadêmico, cursos com índices elevados de evasão e componentes curriculares críticos (PAPAMITSIOU; ECONOMIDES, 2014). Além disso, visualizações interativas em grafos têm se mostrado valiosas para a gestão acadêmica, revelando agrupamentos de evasão, trilhas problemáticas e conexões entre fatores contextuais e institucionais (PENÃ-AYALA, 2014). Ao evidenciar não apenas atributos individuais, mas também as inter-relações entre estudantes e o ambiente institucional, essa modelagem expõe padrões muitas vezes invisíveis em análises convencionais.

Os grafos permitem, ainda, identificar grupos com comportamentos semelhantes de acesso às plataformas, evidenciando que alunos em situação crítica tendem a se concentrar em determinados cursos ou períodos de inatividade, indicando vulnerabilidades associadas a componentes curriculares específicos. Também possibilitam detectar estudantes de baixo risco conectados a grupos mais vulneráveis, sinalizando propensões futuras não capturadas por modelos supervisionados, e destacando a importância das conexões acadêmicas para ações preventivas mais precisas.

### 3 RESULTADOS E DISCUSSÕES

Este estudo partiu da hipótese de que a aplicação integrada de técnicas de mineração de dados educacionais e modelagem em grafos possibilitaria identificar os principais fatores associados à evasão no ensino superior a distância, além de revelar padrões coletivos e trajetórias críticas entre estudantes em situação de risco. Presumia-se que variáveis demográficas, especialmente aquelas relacionadas à localização regional e à condição financeira, se destacariam como preditores relevantes, superando atributos comportamentais e institucionais.

Os resultados revelaram um cenário distinto do previsto: as variáveis com maior poder preditivo para o risco de evasão são essencialmente comportamentais. O critério de ganho de informação, que mede a redução da incerteza sobre a variável-alvo ao se conhecer uma variável preditora, indicou alto impacto para atributos relacionados ao uso da plataforma. Em contextos educacionais, valores acima de 1,0 são considerados relevantes e acima de 1,5, altamente significativos. Nesse contexto, a quantidade de dias sem acesso ao módulo (1,76783) e os dias desde o último acesso ou matrícula (1,66333) destacaram-se como os principais indicadores de risco, evidenciando que a frequência de interação com o ambiente virtual é decisiva no contexto analisado.

Em contrapartida, variáveis demográficas e institucionais, como idade (0,00491), sexo (0,00454), tipo de polo (0,00761), UF (0,00956), filial (0,01663) e regional (0,00204), apresentaram ganhos informacionais baixos, evidenciando sua limitada relevância isolada para a previsão da evasão. Embora úteis para caracterizar o perfil discente, esses atributos não demonstraram capacidade significativa para discriminar estudantes quanto ao abandono acadêmico.

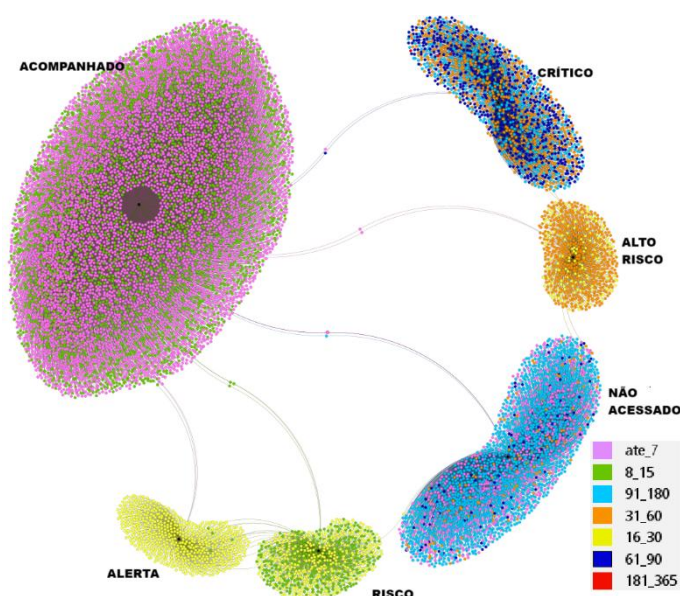
O ganho informacional destacou variáveis acadêmicas intermediárias, como percentual de acerto na primeira avaliação (0,44008) e de conclusão de disciplinas (0,17625), que, embora menos expressivas que os indicadores comportamentais, ajudam a identificar padrões de risco. Fatores financeiros, como parcelas em aberto (0,12967) e valor da inadimplência (0,12077), ocuparam posições medianas, sugerindo influência moderada sobre o desfecho acadêmico.

Esses achados não apenas direcionam a seleção de variáveis mais relevantes para futuras modelagens preditivas, como também reforçam a centralidade dos indicadores



comportamentais na detecção de risco de evasão. Ao evidenciar a importância da frequência de acesso e do comportamento de uso da plataforma, os resultados reiteram a necessidade de monitoramento contínuo das interações acadêmicas como insumo estratégico para a gestão e para a implementação de políticas de permanência mais eficazes.

A modelagem em grafos corroborou esses resultados, evidenciando a formação de agrupamentos bem definidos de estudantes com padrões semelhantes de comportamento e risco. Na construção dos grafos, os nós foram classificados conforme o risco de evasão, com cores distintas por nível, enquanto as arestas conectaram estudantes com padrões próximos de comportamento e inatividade. As visualizações revelaram comunidades acadêmicas com alta concentração de estudantes em situação crítica, além de conexões entre indivíduos de diferentes níveis de risco, indicando que a evasão é influenciada por fatores coletivos e contextuais.



**Figura 1** – Grafo de faixa de dias sem acessar um módulo de estudo por risco de evasão  
Fonte: Adaptado de Gephi (os autores, 2025)

A Figura 1 ilustra essa estrutura relacional, onde cada nó representa um estudante, colorido de acordo com o tempo de inatividade na plataforma, e agrupado conforme as classificações institucionais de risco: Acompanhado, Alerta, Risco, Alto Risco, Não Acessado e Crítico. Observou-se que estudantes com maior tempo de inatividade (azul escuro, ciano e vermelho) predominam no cluster “Crítico”, o que reforça a correlação entre ausência de acesso e propensão à evasão. Além disso, a sobreposição entre os dados de risco e comportamento de acesso reafirma a importância dos indicadores comportamentais na caracterização da evasão e evidencia a relevância das conexões acadêmicas e padrões coletivos para o planejamento de estratégias preventivas mais eficazes.

#### 4 CONSIDERAÇÕES FINAIS

A evasão no ensino superior a distância permanece como um desafio relevante para as instituições educacionais, exigindo abordagens integradas e metodologias inovadoras para seu enfrentamento. Este estudo demonstrou que a combinação de técnicas de Mineração de Dados Educacionais e modelagem em grafos constitui uma estratégia eficaz para identificar padrões de risco, possibilitando análises individuais e coletivas que influenciam a permanência acadêmica.



Os resultados revelaram que variáveis comportamentais, sobretudo aquelas relacionadas à frequência de acesso às plataformas virtuais, são os principais preditores de evasão, superando atributos demográficos e institucionais. A representação em grafos permitiu visualizar comunidades acadêmicas vulneráveis e conexões entre estudantes em diferentes níveis de risco, oferecendo subsídios estratégicos para a gestão acadêmica.

Com trabalhos futuros, a continuidade do trabalho será feita com a inclusão de dados longitudinais para acompanhar a evolução dos comportamentos ao longo do tempo, bem como a incorporação de informações qualitativas, como *feedbacks* e avaliações institucionais. Sugere-se, ainda, o desenvolvimento de sistemas de alerta precoce baseados em inteligência artificial, capazes de orientar intervenções personalizadas, contribuindo para ambientes educacionais responsivos e eficazes.

## REFERÊNCIAS

BORBA, Elizandro Max. **Medidas de centralidade em grafos e aplicações em redes de dados**. 2013. 116 f. Dissertação (Mestrado em Matemática Aplicada) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2013. Disponível em: <https://lume.ufrgs.br/handle/10183/86094>. Acesso em: 22 de jun. 2025.

ENRICONE, Délcia. Estudante da Educação Superior. In: MOROSINI, M. C. (Org.) **Enciclopédia de Pedagogia Universitária**, Glossário. v. 2. Brasília: INEP, 2006. p. 407-437.

GEPHI. **The Open Graph Viz Platform** [online]. 2022. Disponível em: <https://gephi.org/>. Acesso em: 12 jul. 2025.

INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. **Divulgação dos resultados: Censo da Educação Superior 2023**. 2023. Disponível em: [https://download.inep.gov.br/educacao\\_superior/censo\\_superior/documentos/2023/apresentacao\\_censo\\_da\\_educacao\\_superior\\_2023.pdf](https://download.inep.gov.br/educacao_superior/censo_superior/documentos/2023/apresentacao_censo_da_educacao_superior_2023.pdf). Acesso em: 10 jun. 2025.

KUHN, Ighor. **Identificação de Trajetórias de Aprendizagem com o uso de Grafos Direcionados e Técnicas de Mineração de Dados Visando a Detecção de Evasão em Cursos EAD**. 2023. Tese (Doutorado em Informática na Educação) - Universidade Federal do Rio Grande do Sul, Centro de Estudos Interdisciplinares em Novas Tecnologias na Educação, Porto Alegre, 2023. Disponível em: <https://lume.ufrgs.br/bitstream/handle/10183/279018/001208936.pdf?sequence=1&isAllowed=y>. Acesso em: 13 jun. 2025.

LOBO, Maria Beatriz de Carvalho Melo. Panorama da evasão no ensino superior brasileiro: Aspectos gerais das causas e soluções. In: **ABMES Cadernos nº 25**. 2012. P. 09-58. Disponível em: <https://www.calameo.com/read/0044830657857c7b29821>. Acesso em: 10 jun. 2025.

MACEDO, Juliano. de. **EVASÃO NO ENSINO SUPERIOR: UMA REVISÃO DA LITERATURA SOBRE CONCEITOS E CLASSIFICAÇÕES**. **Cenas Educacionais**, [S. l.], v. 7, p. e18997, 2024. DOI: 10.5281/zenodo.13773961. Disponível em: <https://www.revistas.uneb.br/index.php/cenaseducacionais/article/view/18997>. Acesso em: 15 jun. 2025.



NIEROTKA, Rosileia Lucia, SALATA, Andre; MARTINS, Melina Klitzke. (2023). Fatores associados à evasão no ensino superior: Um estudo longitudinal. **Cadernos de Pesquisa**, 53, e09961. Disponível em: <https://doi.org/10.1590/198053149961>. Acesso em: 07 jun. 2025.

OLIVEIRA, João Lucas dos Santos. **Mineração de dados educacionais baseada em grafos**: uma análise em cursos de computação com alto índice de retenção. 2021. 91 f. Dissertação (Mestrado em Ciência da Computação) – Universidade Federal de Goiás, Instituto de Informática, Goiânia, 2021. Disponível em: [https://files.cercomp.ufg.br/weby/up/1289/o/dissertacao\\_JoaoLucasDosSantosOliveira\\_PP GCC.pdf](https://files.cercomp.ufg.br/weby/up/1289/o/dissertacao_JoaoLucasDosSantosOliveira_PP GCC.pdf). Acesso em: 12 de jun 2025.

PAPAMITSIOU, Zacharoula.; ECONOMIDES, Anastasios. A. Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. **Educational Technology & Society**, Athabasca, v. 17, n. 4, p. 49–64, 2014.

PAULA, Gustavo Bruno de; PICANÇO, Felícia. Desigualdades após o acesso: Origem social e evasão do sistema de ensino superior. **Educação & Sociedade**, Volume: 45. 2024. Disponível em: <https://doi.org/10.1590/ES.281915>. Acesso em: 07 jun. 2025.

PEÑA-AYALA, Alejandro. Educational data mining: A survey and a data mining-based analysis of recent works. **Expert Systems with Applications**, Oxford, v. 41, n. 4, p. 1432–1462, 2014. Disponível em: <https://romisatriawahono.net/lecture/rm/survey/machine%20learning/Ayala%20-%20Educational%20Data%20Mining%20-%202014.pdf>. Acesso em: 20 jun. 2025.

ROMERO, Cristobal.; VENTURA, Sebastian. Educational data mining and learning analytics: An updated survey. **Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery**, Hoboken, v. 10, n. 3, e1355, 2020. Disponível em: <https://doi.org/10.1002/widm.1355>. Acesso em: 17 jun. 2025.