
arquivos analíticos de políticas educativas

Revista acadêmica, avaliada por pares,
independente, de acesso aberto, e multilíngue



aape | epaa

Arizona State University

Volume 30 Número 149

4 de octubre de 2022

ISSN 1068-2341

Aplicação de Métodos Preditivos em Evasão no Ensino Superior: Uma Revisão Sistemática da Literatura

Marcelo Ferreira Tete

Universidade Federal de Goiás (UFG)

Marcos de Moraes Sousa

Instituto Federal Goiano (IF Goiano) – Campus Ceres

Thalia Santos de Santana

Instituto Federal Goiano (IF Goiano) – Campus Ceres



Salatyel Fellipe

Universidade Federal de Goiás (UFG)

Brasil

Citação: Tete, M. F., Sousa, M. M., Santana, T. S., & Fellipe, S. (2022). Aplicação de métodos preditivos em evasão no ensino superior: Uma revisão sistemática da literatura. *Arquivos Analíticos de Políticas Educativas*, 30(149). <https://doi.org/10.14507/epaa.30.6845>

Resumo: A evasão escolar é considerada um problema complexo que perpassa diversos níveis e dimensões de análise. O desenvolvimento de modelos preditivos tem sido uma resposta mais dinâmica e proativa ao enfrentamento desse problema. O objetivo desta pesquisa foi o de revisar sistematicamente a literatura sobre predição de evasão no ensino superior. O período de análise foi de 2010 a 2020, compreendendo seis bases de dados de artigos científicos e uma amostra de 48 estudos. Os resultados indicam as características metodológicas e contextuais do estado da arte da literatura em predição da evasão universitária, bem como permitem propor uma agenda de pesquisa para estudos posteriores. A amostra de estudos analisados evidenciou a carência de pesquisas que

Página web: <http://epaa.asu.edu/ojs/>

Facebook: /EPAAAAPE

Twitter: @epaa_aape

Artigo recebido: 1/5/2021

Revisões recebidas 1//2022

Accito: 8/4/2022

relatem ou proponham ações de gestão e de políticas educacionais para além de aplicação dos modelos preditivos de evasão.

Palavras-chave: evasão no ensino superior; estado da arte; modelos preditivos; gestão universitária

Predictive models for higher education dropout: A systematic literature review

Abstract: School dropout is considered a complex problem and one that cuts across several levels of analysis. The development of predictive models has been a more dynamic and proactive response to tackle this problem. This research offers a systematic literature review on dropout prediction in higher education. The analysis period was from 2010 to 2020, searching scientific studies in six databases and working with a sample of 48 studies. The results indicate methodological and contextual characteristics of the cutting-edge literature on dropout prediction and enable the proposition of a research agenda for future studies. The analysis revealed an absence of research reporting or proposing management actions and educational policies that go beyond applying dropout predictive models.

Keywords: higher education dropout; state of the art; predictive models; university management

Aplicación de métodos predictivos en el abandono de la educación superior: Una revisión sistemática de la literatura

Resumen: Se considera que el abandono escolar es un problema complejo y que atraviesa por varios niveles y dimensiones de análisis. El desarrollo de modelos de predicción ha sido una respuesta más dinámica y proactiva a este problema. La investigación tuvo el objetivo de revisar sistemáticamente la literatura sobre predicción de la deserción en la educación superior. El período de análisis fue de 2010 hacia 2020, comprendiendo seis bases de datos de artículos científicos, y una muestra de 48 estudios. Los resultados indican características metodológicas y contextuales del estado del arte de la literatura científica en predicción de deserción y permiten proponer una agenda para investigaciones futuras. La muestra de estudios analizados evidenció la ausencia de investigaciones que reporten o propongan acciones de gestión y políticas educativas para más allá de la aplicación de modelos predictivos de deserción.

Palavras-clave: evasión en la educación superior; estado del arte; modelos predictivos; gestión universitaria

Aplicação de Métodos Preditivos em Evasão no Ensino Superior: Uma Revisão Sistemática da Literatura

A educação superior no Brasil experimentou uma expressiva expansão a partir da década de 1990 que, embora tenha ampliado a demanda estudantil nesse nível de ensino, legou aos gestores educacionais diversos problemas, dentre os quais a permanência e a evasão universitária (Santos Junior et al., 2020). Evasão estudantil é um fenômeno recorrente na educação terciária e tem sido entendida como uma relevante pauta de política educacional (Bernárdez-Gómez & Belmonte, 2020). De acordo com o Mapa do Ensino Superior (SEMESP, 2019), a taxa de evasão nos cursos superiores presenciais brasileiros, em 2017, atingiu 18,6% nas instituições públicas e 25,8% nos estabelecimentos privados. Na modalidade a distância (EAD) as taxas de evasão dos cursos são mais elevadas, tendo alcançado 27,9% nas instituições públicas e 34,9% nas privadas, segundo o mesmo documento.

A evasão no ensino superior é considerada como um problema que pode ser classificado em diversos níveis de análise, desde a *nanoevasão* até a *macroevasão*¹, e permeia várias outras questões, como acesso, permanência, retenção e sucesso (Lima & Zago, 2017). Nesse contexto, a evasão de um discente pode significar inúmeros impactos nos âmbitos individual, institucional e estrutural.

Para o estudante, este processo transforma-se muitas vezes na perda da oportunidade de cursar uma graduação, com prejuízos desde financeiros até de tempo investido; para a universidade, a evasão estudantil acarreta prejuízos e ineficiências operacionais por conta do subaproveitamento dos recursos organizacionais como infraestrutura física, tecnologia e pessoal (Silva et al., 2020).

Apesar da multiplicidade de causas e efeitos, uma abordagem para enfrentar a evasão é tentar prevê-la de modo que se possa proativamente realizar intervenções que previnam a sua ocorrência. Com o crescimento do uso de mineração de dados educacionais a partir de 2014 (Agrusti et al., 2019), tem surgido diversas estratégias metodológicas para predição de evasão, tais como árvores de decisão (Hasbun et al., 2016), redes bayesianas (Sarra et al., 2019), regressão logística (Zhang & Rangwala, 2018) e redes neurais (Martinho et al., 2013).

Diante desse cenário, a presente pesquisa contribui no sentido de analisar sistematicamente a literatura sobre a aplicação de modelos preditivos ao problema da evasão escolar no ensino superior. Os resultados podem gerar conhecimento útil para que pesquisadores e gestores públicos desenvolvam suas próprias abordagens preditivas, bem como planejem e implementem programas e políticas públicas educacionais efetivas voltadas ao enfrentamento e gestão desse fenômeno nas Instituições de Ensino Superior (IES) brasileiras.

O estudo está assim organizado, a Seção 2 apresenta uma contextualização de usos e estudos relacionados à aplicação de modelos preditivos ao sucesso estudantil no ensino superior. A Seção 3 descreve os métodos empregados na realização da revisão sistemática da literatura (RSL). Já na Seção 4 são apresentados os resultados encontrados e suas principais discussões. Na Seção 5, são tecidas as considerações finais, e, por fim, são apresentadas as referências utilizadas.

Modelos Preditivos Aplicados ao Sucesso Estudantil no Ensino Superior

Entre as décadas de 1930 e 1960 surgem nos Estados Unidos os primeiros estudos abrangentes sobre permanência, evasão e êxito estudantil no ensino superior. Nesse período as pesquisas realizadas produziram muitos dados empíricos, mas é somente a partir dos anos 1970 que as primeiras teorias acerca das temáticas da permanência e da evasão de estudantes emergem, com os trabalhos seminais de William Spady e Vincent Tinto (Berger et al., 2012). Desde então, uma vasta literatura se desenvolveu nas duas décadas seguintes como resposta às transformações estruturais e demográficas ocorridas no sistema de ensino superior daquele país.

No final dos anos 1990 e início dos anos 2000, a revolução tecnológica impõe uma nova transformação ao sistema educacional estadunidense à medida que novas e sofisticadas tecnologias de software baseadas na Internet deram às universidades a possibilidade de empregá-las tanto para a gestão de grandes bases de dados acadêmicos e institucionais quanto para a exploração de um novo formato de ensino: os cursos *online* (Picciano, 2012). A expansão desses cursos ampliou as

¹ *Nanoevasão* refere-se a pequenas mudanças administrativas como, por exemplo, alterações de turno, modalidade de ensino (presencial ou EAD) ou formação (licenciatura ou bacharelado) que não implicam o abandono do curso pelo estudante. A *macroevasão*, por sua vez, corresponde à tripla evasão do estudante, isto é, de seu curso, da instituição de ensino e do sistema de educação superior. Há ainda os níveis intermediários: *microevasão*, o estudante evade do curso, mas ingressa em outro curso da mesma instituição; e *mesoevasão*, o estudante evade do curso e da instituição, mas ingressa em um curso de outra instituição, ou seja, permanece no sistema de ensino superior (Lima & Zago, 2017).

preocupações com o problema da evasão nesse período, haja vista que nessa modalidade de educação as taxas de desistência dos estudantes são maiores do que em cursos presenciais (Park & Choi, 2009). Isso levou à revisão dos modelos teóricos de evasão existentes de modo a adaptá-los para o contexto do ensino virtual (Rovai, 2003).

As tecnologias de informação e comunicação continuaram a avançar extraordinariamente nos Estados Unidos nas duas primeiras décadas do século XXI. Esse progresso facilitou o surgimento, no início dos anos 2000, de soluções computacionais de alerta preventivo de evasão que permitiram aos gestores educacionais das universidades compreender melhor o risco de evasão em termos demográficos e comportamentais, bem como a priorizar outro problema derivado de tal risco: o tempo de conclusão do curso (EAB, 2019). Os avanços continuaram e graças ao desenvolvimento de softwares capazes de gerar, armazenar e processar quantidades massivas de dados, técnicas estatísticas sofisticadas de análises preditivas começaram a ser testadas e reconhecidas por algumas universidades como ferramentas promissoras de classificação e identificação de padrões com vasto potencial de aplicação aos problemas relacionados à permanência e ao sucesso estudantil (Campbell et al., 2007). Algumas experiências nesse sentido podem ser destacadas.

A primeira delas é a iniciativa pioneira *Predictive Analytics Reporting Framework* (PAR), conduzida entre 2010 e 2014 com a participação de 20 instituições associadas à *Western Cooperative for Educational Technology* (WCET) e financiada pela *Bill & Melinda Gates Foundation*. O projeto PAR foi um esforço coletivo de utilização de análises preditivas para melhorar as taxas de permanência e de conclusão dos estudantes de cursos *online* (Wagner & Longanecker, 2016). Cada instituição contribuiu com dados acadêmicos anonimizados de seus alunos e cursos para a formação de um grande banco de dados mantido pela WCET, organização responsável pela construção dos modelos preditivos, gestão dos dados e operacionalização dos objetivos e atividades do projeto (Ice et al., 2012).

Em 2012 teve início outra iniciativa de uso de modelos preditivos por uma única instituição, a *Georgia State University* (GSU), que se tornou possivelmente um dos principais casos de sucesso nos Estados Unidos. Naquele ano a GSU decidiu transformar o seu serviço de aconselhamento e orientação estudantil, tornando-o proativo em vez de reativo. Para tanto, foram analisados dez anos de dados acadêmicos e criou-se o *Graduation and Progression System* (GPS), um software baseado em técnicas de predição que envia alertas preventivos aos funcionários do escritório de aconselhamento, indicando os estudantes com alto risco de evasão ou risco de não concluírem seus respectivos cursos no tempo ideal (GSU, 2017). O GPS contém mais de 800 indicadores de alerta de risco e disponibiliza, por meio de um sofisticado *dashboard*, análises do progresso acadêmico dos estudantes que permitem ao conselheiro identificar potenciais obstáculos ao sucesso do aluno e deflagrar intervenções antecipadas para superá-los (Kurzweil & Wu, 2015).

A GSU também faz uso de dados preditivos para apoiar estudantes calouros antes mesmo que as aulas do primeiro ano letivo, no outono, se iniciem. O modelo preditivo identifica os estudantes ingressantes que apresentam maior risco acadêmico e solicita a eles que frequentem o *Summer Success Academy* — programa de disciplinas de verão que confere 7 créditos ao aluno (Kurzweil & Wu, 2015). Ao longo desse período a instituição oferece tutoria, orientação acadêmica e programas de desenvolvimento acadêmico e capacitação financeira (GSU, 2021).

Em 2014 não só a GSU, mas a *Arizona State University* (ASU) e a *University of Texas at Austin* (UT Austin) já eram reconhecidas como líderes no uso de técnicas preditivas nos Estados Unidos

(Paterson, 2019). Essas três universidades naquele mesmo ano se uniram a outras oito instituições² para formar a *University Innovation Alliance* (UIA), um consórcio de universidades públicas de pesquisa criado com o compromisso de aumentar o número e a diversidade de estudantes graduados nos Estados Unidos, por meio da inovação compartilhada e aprendizado colaborativo (UIA, 2021a).

Em 2015, com base nas experiências bem sucedidas na GSU, ASU e UT Austin, e com financiamento da *Bill e Melinda Foundation*, foi lançada pela UIA uma iniciativa para promover o *scale-up* das ferramentas de predição para as demais instituições integrantes do consórcio (UIA, 2021b). O objetivo principal da proposta era aumentar as taxas de permanência e de graduação acadêmica — particularmente para estudantes de baixa renda, negros e de primeira geração³ — por meio da análise de dados institucionais a fim de identificar alertas preventivos e, com isso, gerar intervenções precoces (UIA, 2020). O projeto foi desenhado de modo que as três universidades mais experientes acima mencionadas assumissem a mentoria das demais (Paterson, 2019).

As três iniciativas apresentadas (PAR, *Georgia State University e University Innovation Alliance*), duas delas conduzidas de forma coletiva, evidenciam que a modelagem preditiva aplicada à gestão do sucesso do estudante — o que inclui gestão de evasão e permanência — já é uma realidade nos Estados Unidos há pelo menos 10 anos. Essa tecnologia tem sido cada vez mais adotada por instituições de ensino superior naquele país a ponto de ter chamado a atenção da *New America*, uma organização sem fins lucrativos sediada em Washington, DC, que tem publicado estudos que visam analisar o potencial das ferramentas de predição, bem como identificar as boas e más práticas em sua utilização. Três relatórios sobre o tema já foram publicados pela entidade nos anos de 2016, 2017 e 2018.

O primeiro foi dedicado a examinar como as instituições de ensino superior estavam utilizando modelos preditivos e delineou os desafios de fazê-lo de forma ética (Ekowo & Palmer, 2016). O uso ético dessa ferramenta foi novamente abordado no segundo relatório e cinco práticas orientadoras foram sugeridas aos gestores das instituições: 1) ter uma visão e um plano; 2) construir uma estrutura de apoio; 3) garantir o uso apropriado dos dados; 4) desenvolver modelos preditivos e algoritmos não enviesados; 5) atender aos objetivos institucionais e melhorar o desempenho dos estudantes, intervindo com cuidado (Ekowo & Palmer, 2017). O terceiro relatório teve um objetivo mais prático, ou seja, oferecer às instituições de ensino ferramentas para poderem decidir se é melhor desenvolver sistemas de predição internamente ou contratar fornecedores externos especializados — neste caso, sugerindo os passos para escolherem os melhores e mais éticos (Palmer, 2018).

Até aqui foram apresentados os desenvolvimentos e avanços no uso de ferramentas de predição ocorridos nos Estados Unidos, a fim de mostrar como instituições do sistema de ensino superior desse país estão testando e/ou utilizando, há pelo menos uma década, a abordagem da predição do sucesso dos estudantes em suas práticas de gestão. Isso não significa que não existam outras experiências internacionais em curso, mas aparentemente estas ainda se encontram concentradas em estudos experimentais de natureza acadêmica, não atingindo a escala e o grau de disseminação que pode ser verificada nos Estados Unidos. Mesmo se considerados apenas os estudos acadêmicos sobre técnicas preditivas na educação superior, a ampla hegemonia estadunidense em relação a outros países e continentes se comprova — pelo menos no que diz

² Iowa State University, Michigan State University, Ohio State University, Oregon State University, Purdue University, University of California – Riverside, University of Central Florida e University of Kansas. Atualmente, *University of Kansas* e *University of Texas – Austin* não fazem mais parte da UIA, em contrapartida, outras quatro universidades ingressaram no consórcio: North Carolina A&T State University, University of Illinois Chicago, UMBC, e Virginia Commonwealth University.

³ Estudantes que são os primeiros em suas respectivas famílias a ingressarem num curso superior.

respeito a preditores e sistemas de alerta precoce de evasão —, conforme demonstra o levantamento de literatura realizado por Liz-Domínguez et al. (2019).

No Brasil, a literatura em modelos preditivos aplicados ao ensino superior é ainda incipiente, majoritariamente publicada em congressos da área de informática e, em geral, está imersa no campo da mineração de dados, conforme demonstram os trabalhos de Colpo et al. (2020) e Maschio et al. (2018). Para que esse quadro se altere parece ser relevante a realização de mais estudos sobre a temática, seja para relatar e analisar estudos de casos ou realizar estudos sistemáticos como a presente proposta, a exemplo de outros autores que se debruçaram sobre o desenvolvimento de estudos dessa natureza com o propósito de identificar fatores que influenciam na evasão no ensino superior (Agrusti et al., 2019; Alturki et al., 2020; Proaño & Villamar, 2018).

Métodos

O presente estudo é classificado como uma revisão sistemática de literatura (RSL) ao objetivar uma discussão detalhada das evidências (Kitchenham et al., 2009) em relação à evasão no ensino superior com base em métodos preditivos. Para tanto, o estudo foi conduzido seguindo um rigoroso protocolo de estudos sistemáticos da literatura, em três fases principais: i) planejamento; ii) condução e iii) disseminação/publicação dos resultados.

Assim, com o intuito de alcançar o objetivo principal desta pesquisa, buscou-se responder a três questões de pesquisa (QP), elaboradas de acordo com o critério de estudos sistemáticos PICOC (Scannavino et al., 2017), sendo *P* – população (Artigos científicos completos avaliados por pares em bases de artigos científicos de alta reputação acadêmica sobre evasão em universidades); *I* – intervenção (Pesquisa bibliométrica); *C* – comparação (Aplicações no contexto brasileiro e estrangeiro, diferenças de variáveis e de métodos); *O* – resultados, do termo inglês *outcomes* (Estado da arte e construção de agenda de pesquisa para estudos em evasão escolar em nível superior); e *C* – contexto (Contexto de aplicação em nível de país).

As QPs formuladas são elencadas a seguir:

- QP1 – Quais são as características metodológicas e contextuais do estado da arte da predição da evasão escolar em nível superior?
- QP2 – Quais são as variáveis consideradas determinantes para explicar e prever a evasão escolar em nível superior?
- QP3 – De acordo com a literatura analisada, quais as ações as IES vêm desenvolvendo para lidar com a evasão na educação de nível superior, para além da aplicação de modelos preditivos?

Deste modo, para a *string* de busca, foram selecionados termos que correspondem ao escopo de pesquisa, todos em língua inglesa, visando aumentar o quantitativo de artigos retornados. Após a realização de uma busca-piloto por título, palavras-chave e resumo, juntamente à sucessivos refinamentos e levantamento de terminologias relevantes, foi obtida a seguinte *string* final:

(*dropout* OR "*drop out*" OR "*student desertion*" OR "*academic success*") AND ("*higher education*")
AND ("*predictor models*" OR "*dropout prediction*" OR "*educational data**")

Neste trabalho foi aplicada uma estratégia de busca automática ampla em seis bases e motores de busca em um filtro temporal de 10 anos, sendo escolhidas as seguintes: ACM *Digital*

Library, IEEE Xplore, Scielo, Science Direct, Scopus e Spell. Para a base de dados brasileira Scielo, executou-se a *string* também em língua portuguesa, contudo, o resultado foi insignificante visto que nenhum artigo foi retornado. A busca ocorreu em outubro de 2020, não sendo utilizadas demais estratégias combinadas, como o uso de busca manual.

Desta forma, em função da seleção dos estudos relevantes, foram definidos critérios de inclusão (CI) e de exclusão (CE) nesta RSL, conforme a seguir:

Inclusão

CI1: O estudo relata os determinantes da evasão escolar no ensino superior com emprego de técnicas e modelos preditivos aplicados à evasão escolar.

Exclusão

CE1: O artigo não foi publicado nos últimos 10 anos;

CE2: O artigo não trata de evasão estudantil na graduação;

CE3: O artigo não trata de modelos preditivos para a evasão;

CE4: O estudo é uma versão mais antiga de outro estudo já considerado;

CE5: O texto completo não se encontra disponível;

CE6: O trabalho não é um artigo completo de conferência ou periódico;

CE7: O trabalho não é um estudo primário.

Seguindo os critérios estabelecidos, o processo de seleção percorreu duas etapas: i) leitura dos metadados (título e resumo) e ii) leitura do texto completo. Para aceite ou exclusão, cada um dos artigos foi enquadrado em um dos CI ou CE elencados. Desta forma, os trabalhos aceitos em instância final enquadraram-se em um formulário de extração de dados objetivando responder QPs definidas no protocolo, conforme apresentado na Tabela 1.

Tabela 1

Formulário de Extração de Dados

Campos de Extração		
Veículo de publicação	Contexto	Amostragem
Ano de publicação	Modalidade	Principais variáveis determinantes da evasão
Instituição dos autores	Técnica/método	Principais resultados
Quantidade de autores	Natureza da pesquisa	Medidas de enfrentamento para além da predição
Tipo de publicação	Origem dos dados	Lacunias sugeridas

Nota. Elaborado pelos autores deste artigo.

Como forma de apoiar a realização desta RSL, utilizou-se da plataforma *online Parsif.al*⁴, considerada como *software* de apoio para a execução de estudos sistemáticos, desde a documentação

⁴ <https://parsif.al/>

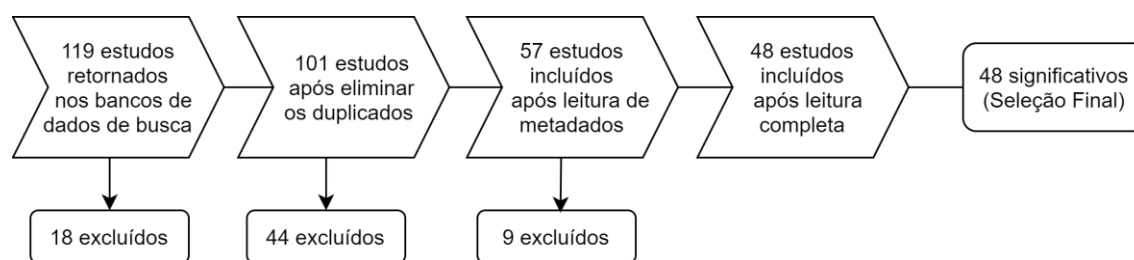
do protocolo até a extração dos dados de cada um dos artigos selecionados. Além disso, também foi utilizada a ferramenta Google Planilhas, para organização de dados e geração de gráficos dos resultados encontrados.

Resultados e Discussão

A aplicação da *string* nas bases de dados por meio de busca avançada resultou em 119 artigos retornados, sendo quatro da ACM Digital Library, 14 da IEEE Xplore, 34 da Scielo, cinco da Science Direct, 61 da Scopus e um da Spell. Do total de estudos, 18 foram duplicados, restando 101 deles para análise de metadados. Destes, 57 foram pré-selecionados após leitura de título e resumo e, ao final da leitura completa, 48 trabalhos representaram o conjunto de estudos significativos na RSL, como apresentado na Figura 1.

Figura 1

Fluxo de Informação do Processo de Seleção dos Artigos



Nota. Elaborado pelos autores deste artigo

A Tabela 2 relaciona os trabalhos eliminados de acordo com os critérios de exclusão aplicados, conforme definido pelo protocolo de pesquisa da RSL.

Tabela 2

Quantitativo de Artigos eliminados por Critério de Exclusão

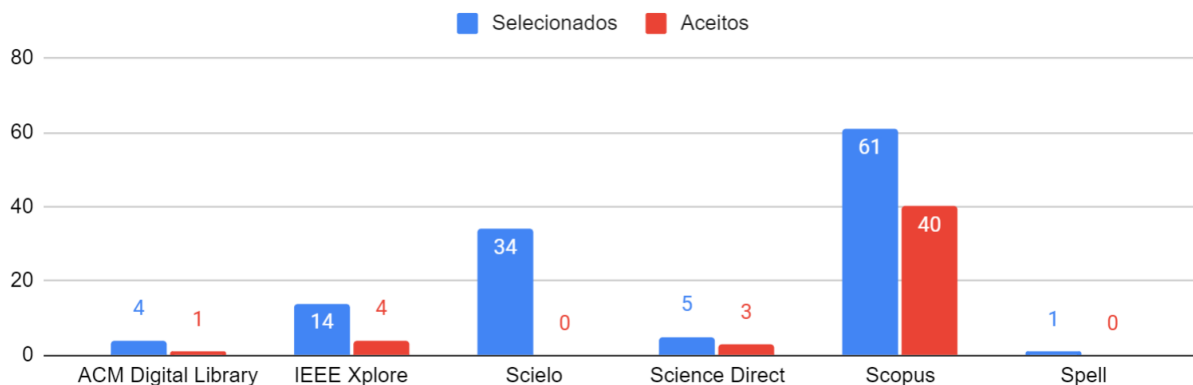
Critérios de Exclusão (CE)							
CE1	CE2	CE3	CE4	CE5	CE6	CE7	Total
10	16	14	1	6	6	0	53 excluídos

Nota. Elaborado pelos autores deste artigo.

Portanto, dos 48 trabalhos significativos, cerca de 83,3% foram derivados de um único mecanismo de busca, sendo a biblioteca digital Scopus considerada como a mais relevante para este estudo sistemático, como pode ser observado na Figura 2.

Figura 2

Estudos Aceitos por Bases de Dados



Nota. Elaborado pelos autores deste artigo.

Dentre os 48 artigos analisados, considerados entre os últimos 10 anos, foi possível verificar que os primeiros estudos aceitos datam apenas do ano de 2013. Também é perceptível uma tendência de aumento de publicações a partir de 2015 (Figura 3), o que ressalta que este campo de estudo vem se destacando nos últimos anos, conforme já observado anteriormente na literatura (Agrusti et al., 2019). Como a pesquisa foi realizada em outubro de 2020, muitos artigos ainda poderiam ser publicados e, portanto, não refletem a realidade da quantidade total do ano.

Figura 3

Estudos Aceitos por Ano de Publicação

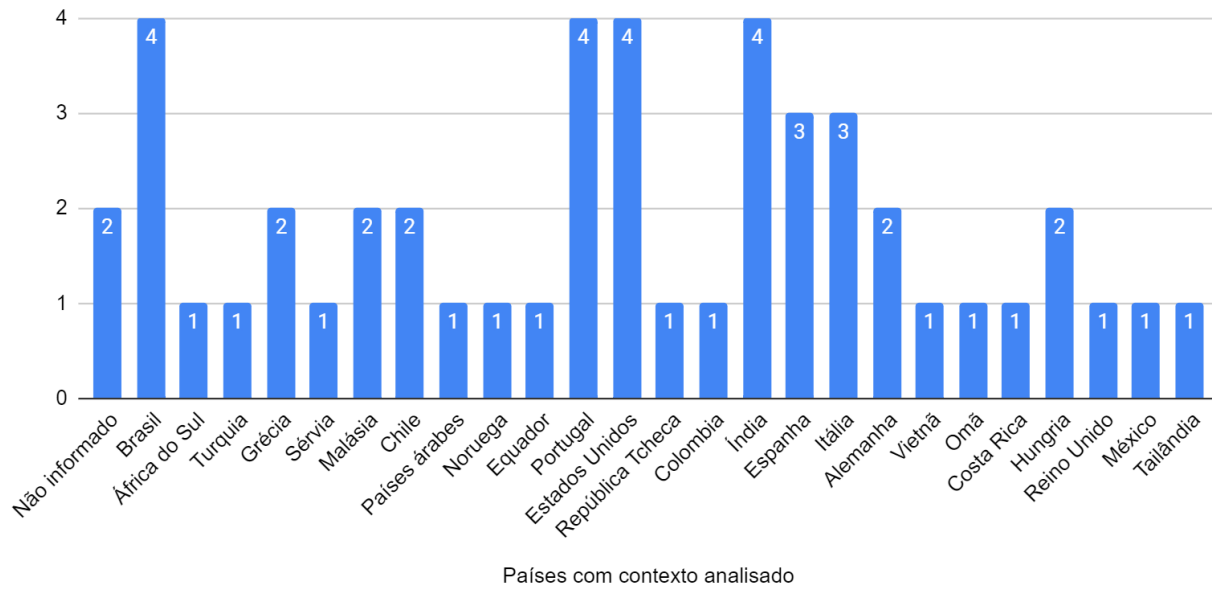


Nota. Elaborado pelos autores deste artigo.

Dessa amostra, o Brasil destaca-se entre os 25 países citados (Figura 4), juntamente com Estados Unidos, Portugal e Índia (todos com quatro artigos retornados).

Figura 4

Quantitativo de Artigos por Contexto Geográfico das Pesquisas Executadas

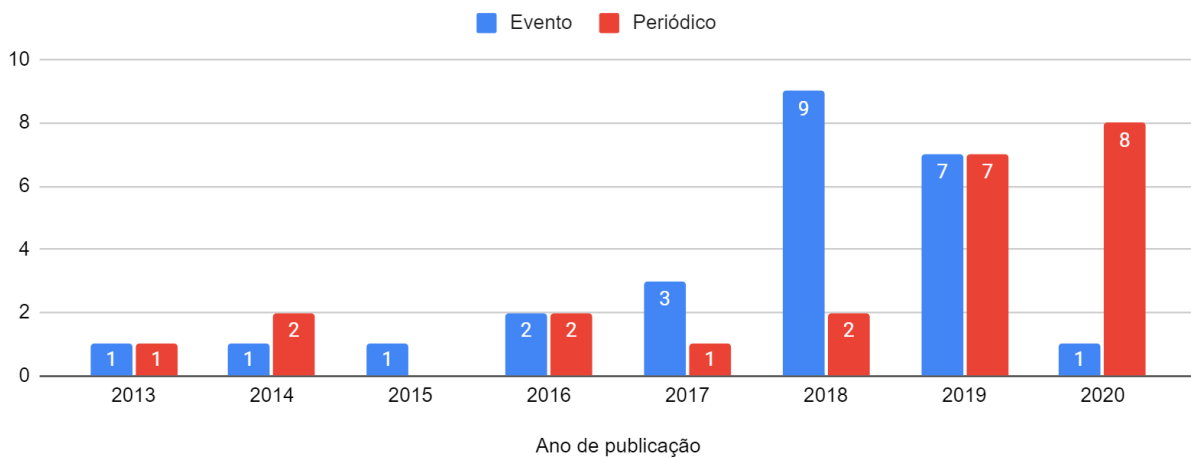


Nota. Elaborado pelos autores deste artigo.

Quando são verificados os tipos de publicações existentes, como artigos publicados em periódicos ou publicados em eventos, de forma geral os eventos retornaram maior número de pesquisas (52,1%). Entretanto, em uma evolução temporal, ao se considerar os estudos mais atuais, do ano de 2020, estes demonstraram grande impacto em veículos de publicação do tipo periódicos. Cabe destacar, porém, que em anos anteriores como os de 2019, 2016 e 2013, o mesmo quantitativo de trabalhos de ambos os tipos retornou do processo de busca, sendo que em 2018 os eventos destacaram-se, conforme demonstra a Figura 5.

Figura 5

Tipo de Veículo por Ano de Publicação



Nota. Elaborado pelos autores deste artigo.

Ressalta-se, ainda, que a área de computação no Brasil reconhece a publicação em eventos avaliados por pares como tão relevantes quanto a publicação em periódicos⁵, haja vista a existência de indicadores Qualis-Capes para eventos da área.

Dos métodos empregados em modelos preditivos de evasão, há um crescente interesse e aplicação de abordagens que envolvam técnicas de computação, a exemplo de mineração de dados educacionais⁶. Estas envolvem a descoberta de conhecimento em bases de dados (comumente denominado de KDD – *Knowledge Discovery in Database*), em benefício de informações significativas e relacionamentos entre variáveis que possam vir a contribuir na previsão da evasão escolar (Agrusti et al., 2019).

Desta forma, quanto às técnicas preditivas adotadas nos estudos analisados destacam-se os algoritmos de mineração de dados e de aprendizagem de máquina, sendo que os quatro mais mencionados nos artigos desta revisão foram árvores de decisão, redes bayesianas, regressão logística e redes neurais. As árvores de decisão, representam um modelo hierárquico em estrutura de árvore com raízes direcionadas (Agrusti et al., 2019), que podem ser visualizadas como um fluxograma com as regras obtidas, o que confere certa vantagem de facilidade na interpretação dos dados.

Em relação aos demais tipos citados, as redes bayesianas são modelos gráficos probabilísticos de classificação, que denotam uma representação da distribuição de probabilidades conjuntas das variáveis do modelo (Agrusti et al., 2019).

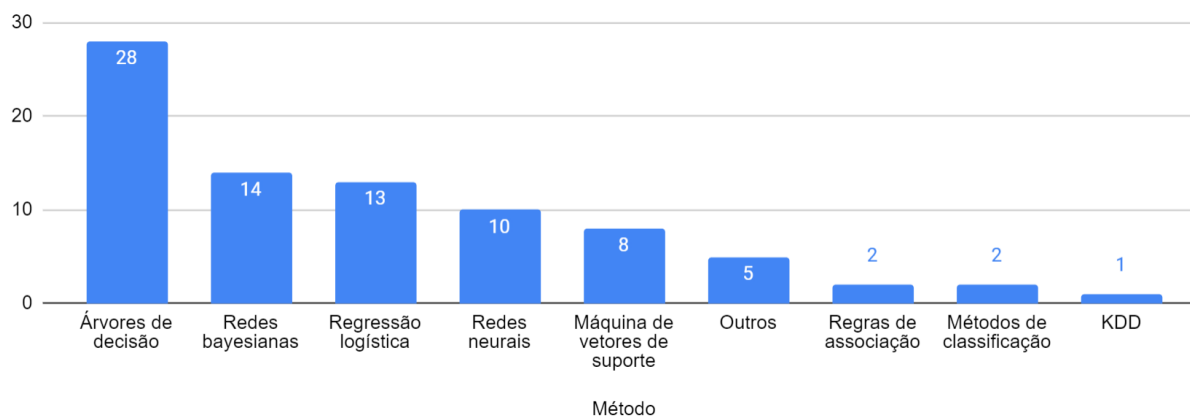
Quanto à regressão logística, esta diz respeito ao uso de um método estatístico capaz de encontrar padrões nos dados por meio do relacionamento entre dois conjuntos, como X e Y e, neste caso, permite o cálculo de probabilidades da ocorrência de determinado fenômeno analisado (Silva et al., 2020).

Já as redes neurais são modelos computacionais ou objetos matemáticos inspirados nos circuitos neurais do cérebro humano, cuja implementação, diferentemente do paradigma clássico da programação que requer do programador o conhecimento prévio do problema a ser resolvido, tem por premissa a abordagem da caixa preta (*black box*), ou seja, não é necessário que o programador entenda *a priori* o mecanismo subjacente ao fenômeno classificado para que crie o algoritmo correto (Agrusti et al., 2019).

As diversas técnicas identificadas na RSL foram agrupadas de forma geral considerando a divisão de Agrusti et al. (2019) e são apresentadas na Figura 6. Com maior adoção nas pesquisas, as árvores de decisão são citadas em 28 dos estudos, ao passo que redes bayesianas estão contidas em 14 dos trabalhos. Estes dois métodos de classificação vêm sendo considerados como os mais populares em estudos que abordam questões relacionadas ao desempenho do aluno (Khasanah & Harwati, 2017), principalmente devido à acurácia e precisão dos modelos em relação aos demais algoritmos.

⁵ <https://www.gov.br/capes/pt-br/centrais-de-conteudo/CienciaComputacao20132015.pdf>

⁶ A mineração de dados pode fazer uso de algoritmos de aprendizagem de máquina (*machine learning*, em inglês) com o intuito de que tal processo de obtenção de padrões ocorra de modo automatizado e considerando um elevado volume de dados. Tal área de estudo faz uso de elementos computacionais que permitem que os computadores “aprendam” e reconheçam padrões complexos baseados em dados, e que, dependendo do tipo de aprendizado, venham a tomar decisões em relação a eles (Han et al., 2011).

Figura 6*Frequência do Emprego das Técnicas de Classificação*

Nota. Elaborado pelos autores deste artigo.

Árvores de decisão são consideradas por Dias et al. (2008) vantajosas no contexto educacional, pois permitem chegar às decisões considerando os atributos mais representativos. As regras de decisão representadas na forma de uma árvore auxiliam no nível de compreensão pela grande maioria dos indivíduos, observando-se quais dos fatores possuem maior grau de influência — neste caso, para a evasão estudantil.

O segundo item mais abordado nos estudos analisados são as redes bayesianas, cuja adoção possibilita, por meio do estabelecimento de relações de causalidade, verificar por exemplo, um relacionamento de dependência entre as ações desenvolvidas no caráter de ensino-aprendizagem e o próprio desempenho alcançado pelos acadêmicos (Dias et al., 2008).

Apesar de ambas as técnicas preditivas possuírem alta velocidade de classificação, elas pertencem a perfis operacionais diferentes. Nesse sentido, as árvores de decisão realizam a predição com base em critérios de divisão dos dados de treinamento e, ao mesmo tempo, têm menor tolerância a dados ruidosos do que uma rede bayesiana. Assim, a seleção de uma ou outra deve considerar o cenário em que serão aplicadas (Sharma et al., 2013).

Khasanah e Harwati (2017) estabeleceram em sua pesquisa um comparativo entre as duas técnicas e sugerem que, dependendo dos fatores avaliados, pode-se elevar a taxa de precisão destes métodos, o que demonstra a importância da seleção dos atributos de maior impacto para melhorar a assertividade dos modelos preditivos. Dessa maneira, ressalta-se também que devido às vantagens e desvantagens de cada uma das abordagens, boa parte dos artigos identificados empregaram técnicas combinadas para previsão da evasão, a exemplo de árvores de decisão e regressão logística, ou mesmo, árvores de decisão e redes bayesianas.

No que concerne à modalidade de ensino a que os estudos levantados se referem, percebe-se a predominância da modalidade presencial (~81% da amostra), embora a educação à distância tenha atraído mais pesquisas a partir de 2018. Sobre a origem dos dados, apenas dois estudos tiveram origem em dados secundários, o que aparentemente indica a dificuldade de acesso a bases de dados públicas e consolidadas. Assim, considerando que poucos estudos têm amostras com mais de uma instituição de ensino, predominam os estudos em apenas uma organização, o que limita o poder preditivo, a acurácia e a generalização dos modelos.

Com relação aos fatores determinantes da evasão, de forma não exaustiva os estudos avaliam as seguintes dimensões de variáveis ou atributos:

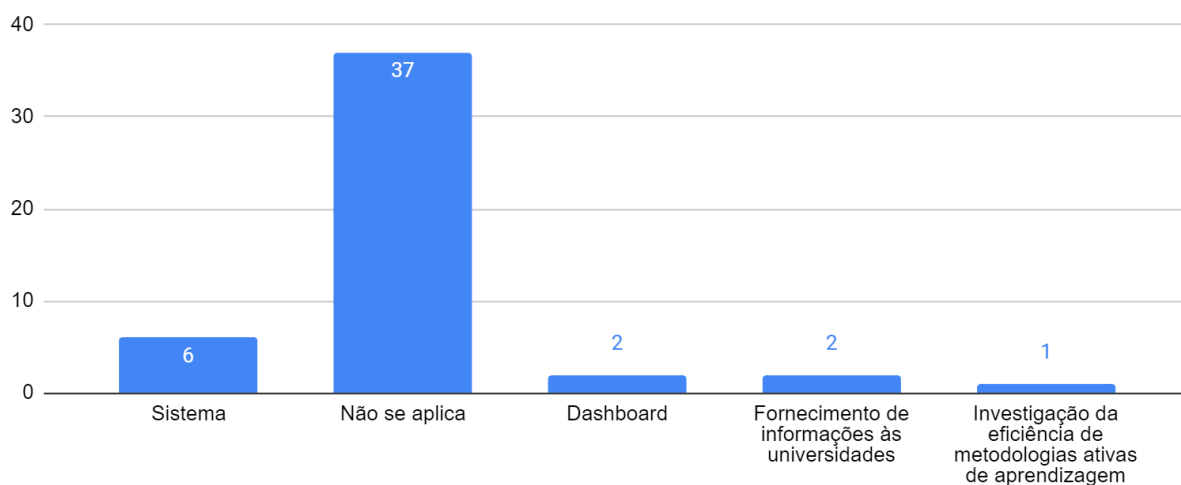
- **Socioeconômicas:** nessa dimensão são usadas variáveis, também denominados atributos, como gênero, etnia, idade, renda, estado civil, tipo de transporte, quantidade de filhos, se o aluno trabalha e qual o status do trabalho, se o aluno tem computador e qual o nível de conhecimento computacional, nível educacional e ocupação dos pais, distância da escola para a residência do aluno, tipo de escola secundária, uso de financiamento estudantil e de mecanismos de permanência, dentre outras;
- **Acadêmicas:** nessa dimensão são usadas variáveis/atributos como frequência, horas de estudo, notas parciais, nota média geral do curso, disciplinas aprovadas e reprovadas, consultas à biblioteca, dados do sucesso acadêmico do estudante no ensino médio, participação em tutorias e monitorias, outras atividades realizadas pelo aluno (extensão, línguas, esportes, culturais etc.);
- **Psicológicas e de Saúde:** mais recentemente também estão sendo incluídas variáveis/atributos relacionados com aspectos psicológicos e de saúde como dificuldade de aprendizagem e memória, resiliência acadêmica, satisfação com o método e com a aula, grau de satisfação com os espaços acadêmicos; grau de satisfação com tutorias e monitorias, bem-estar do aluno, problemas de saúde, mídia social, número e tempo gasto com os amigos, integração com a universidade, envolvimento com a comunidade, consumo de álcool e de cigarros, nível de motivação para o ingresso ao curso, condescendência com os maus hábitos;
- **Acesso:** os estudos que avaliam a educação à distância têm como principal base para o desenvolvimento de modelos preditivos variáveis/atributos relacionados com o acesso, como os *logs* realizados nos ambientes virtuais e ambientes de aprendizagem, acesso aos vídeos, blogs, questionários e diferentes recursos disponibilizados.

Uma lacuna importante verificada nessa amostra de estudos é que, em sua maioria, as pesquisas não realizam ou propõem ações de enfrentamento para além da predição de possível evasão dos estudantes (Figura 7). No Brasil, foi verificado apenas um trabalho que propôs esse tipo de ação. Trata-se de um estudo de pesquisadores do Instituto Federal do Ceará (Freitas et al., 2020), em que foi criada uma plataforma de previsão do abandono com base em informações socioeconômicas.

Dos 11 estudos que avançaram para além da predição de evasão, ressaltam-se o desenvolvimento de sistemas (reportado em seis trabalhos analisados), bem como aplicações com foco em *business intelligence*, *dashboards* e fornecimento de informações às universidades (sistemas de alerta). No contexto da instituição de ensino, em apenas um dos artigos foi detectada a investigação do efeito de metodologias ativas de aprendizagem no enfrentamento da evasão.

Figura 7

Ações de Enfrentamento da Evasão além de Modelos Preditivos



Medidas de enfrentamento para além da predição

Nota. Elaborado pelos autores deste artigo.

A Figura 8 resume de forma cruzada as dimensões dos atributos e variáveis, modalidade de ensino, país de contexto da pesquisa e estudo.

Figura 8

Dimensão, Modalidade, Contexto de Aplicação e Estudos Revisados

Dimensão	Modalidade	Contexto	Estudos
Socioeconômica	Presencial	Brasil	(Martinho et al., 2013) (Freitas et al., 2020)
		Sérvia	(Išljamović et al., 2016)
		Chile	(Hasbun et al., 2016) (Olaya et al., 2020)
		Equador	(Noboa et al., 2018)
		República Tcheca	(Vaclavek et al., 2018)
		Colômbia	(Hernandez et al., 2018)

Figura 8 (Cont.)*Dimensão, Modalidade, Contexto de Aplicação e Estudos Revisados*

Dimensão	Modalidade	Contexto	Estudos
Socioeconômica	Presencial	Índia	(Hussain et al., 2018) (Hegde & Prageeth, 2018) (Kamal & Ahuja, 2019) (Devasia et al., 2016)
		Estados Unidos	(Jayaraman et al., 2019)
		Itália	(Perchinunno et al., 2019) (Agrusti & Mezzini, 2020)
		Omã	(Krishnan et al., 2019)
		Costa Rica	(Fernández-Martín et al., 2018)
		Hungria	(Kiss et al., 2019) (Baranyi et al., 2020)
		México	(Urbina-Nájera et al., 2020)
		Portugal	(Miguéis et al., 2018) (Martins et al., 2020)
	EaD	Grécia	(Kostopoulos et al., 2017)
		Estados Unidos	(Kang & Wang, 2018) (Kashyap & Nayak, 2018)
Acadêmica	Presencial	Brasil	(Martinho et al., 2013) (Manrique et al., 2019) (Costa et al., 2019)
		África do Sul	(Mashiloane & Mchunu, 2013)
		Turquia	(Tekin, 2014)
		Sérvia	(Išljamović et al., 2016)
		Malásia	(Badr et al., 2016) (Yaacob et al., 2019)
		Chile	(Hasbun et al., 2016) (Olaya et al., 2020)
		Não Identificado	(Moscoso-Zea et al., 2017)

Figura 8 (Cont.)*Dimensão, Modalidade, Contexto de Aplicação e Estudos Revisados*

Dimensão	Modalidade	Contexto	Estudos
Acadêmica	Presencial	Bahrein, Egito, Jordânia, Kuwait, Líbano, Omã, Palestina, Arábia Saudita, Sudão	(Hussein & Khan, 2017)
		Equador	(Noboa et al., 2018)
		Portugal	(Lima et al., 2018) (Martins et al., 2019) (Miguéis et al., 2018) (Martins et al., 2020)
		Estados Unidos	(Zhang & Rangwala, 2018) (Jayaraman et al., 2019)
		República Tcheca	(Vaclavek et al., 2018)
		Colômbia	(Hernandez et al., 2018)
		Índia	(Hegde & Prageeth, 2018) (Kamal & Ahuja, 2019) (Devasia et al., 2016)
		Itália	(Perchinunno et al., 2019) (Sarra et al., 2019) (Agrusti & Mezzini, 2020)
		Alemanha	(Askinadze & Conrad, 2016) (Behr et al., 2020)
		Vietnã	(Mai et al., 2019)
		Omã	(Krishnan et al., 2019)
		Costa Rica	(Fernández-Martín et al., 2018)
		Hungria	(Kiss et al., 2019) (Baranyi et al., 2020)
		México	(Urbina-Nájera et al., 2020)
Tailândia	(Lam-On & Boongoen, 2014)		

Figura 8 (Cont.)*Dimensão, Modalidade, Contexto de Aplicação e Estudos Revisados*

Dimensão	Modalidade	Contexto	Estudos
Acadêmica	EaD	Grécia	(Kostopoulos et al., 2015) (Kostopoulos et al., 2017)
		Estados Unidos	(Kang & Wang, 2018) (Kashyap & Nayak, 2018)
		Espanha	(Figuroa-Cañas & Sancho-Vinuesa, 2020)
Psicológicas e de Saúde	Presencial	Noruega	(Giannakos et al., 2017)
		Índia	(Hegde & Prageeth, 2018) (Kamal & Ahuja, 2019) (Devasia et al., 2016)
		Itália	(Sarraf et al., 2019)
		México	(Urbina-Nájera et al. 2020)
Acesso	EaD	Não Identificado	(Mishra & Mishra, 2018)
		Grécia	(Kostopoulos et al., 2015)
		Espanha	(Figuroa-Cañas & Sancho-Vinuesa, 2019) (Lara et al., 2014) (Figuroa-Cañas & Sancho-Vinuesa, 2020)
		Reino Unido	(Mubarak et al., 2020)
		Estados Unidos	(Kashyap & Nayak, 2018)

Nota. Elaborado pelos autores deste artigo.

Considerações Finais

O presente estudo teve como objetivo realizar uma revisão sistemática da literatura (RSL) sobre predição de evasão no ensino superior no período de 2010 a 2020. Os resultados apresentados indicam a evolução recente do tema, o contexto geográfico de aplicação das técnicas de predição empregadas, as principais variáveis e dimensões classificadas por modalidade de ensino, bem como permitem identificar as principais lacunas dos estudos revisados. Além disso, também se destacam ações de impacto desenvolvidas no contexto da evasão, em especial nos Estados Unidos, constituindo um levantamento teórico-empírico de práticas exitosas que podem inspirar intervenções no ensino superior brasileiro.

Conforme descrito na seção metodológica, a RSL conduzida foi guiada por três questões de pesquisa (QP). A QP1 tem por intento identificar as características metodológicas e contextuais do estado da arte da predição da evasão escolar em nível superior. A QP2, por sua vez, indaga sobre as variáveis consideradas determinantes para explicar e prever a evasão escolar em nível superior.

Finalmente, a QP3 busca conhecer quais ações vêm sendo desenvolvidas para além da aplicação de modelos preditivos nas IES e nos respectivos estudos analisados.

Quanto à QP1, os artigos revisados evidenciam a utilização de modelos preditivos probabilísticos relacionados às técnicas de mineração de dados para antever a evasão, principalmente árvores de decisão, que são aplicadas inclusive de forma combinada com outros métodos. No que diz respeito à QP2, as variáveis socioeconômicas, acadêmicas, psicológicas e de saúde, bem como as relacionadas ao acesso, são as principais utilizadas nas pesquisas realizadas em diferentes países e que consideram o contexto de uma única IES. Por fim, em resposta à QP3, constatou-se que a grande maioria dos estudos analisados se limitam apenas à investigação da predição da evasão sem avançar nas ações que podem ser implementadas com os dados gerados pelos modelos preditivos com vistas à prevenção da evasão estudantil.

O resultado encontrado para a QP3, todavia, merece algumas considerações adicionais, na perspectiva de que a disponibilidade e o uso de modelos preditivos, por si, não são capazes de mitigar o problema da evasão. Para além da predição faz-se necessário desenvolver tanto ações de gestão dentro das IES como políticas públicas nos sistemas educacionais de nível superior com caráter preventivo e de suporte ao estudante.

No âmbito das IES, ações de permanência e sucesso acadêmico relacionadas à aplicação de modelos preditivos, a exemplo de centros de tutoria com enfoque no acolhimento e ensino-aprendizagem (Gutierrez et al., 2018), bem como a avaliação de outras ações de gestão educacional voltadas à melhoria do desempenho acadêmico podem ser desenvolvidas. Destaque-se também o papel do professor em motivar o aluno a superar dificuldades, priorizar a formação e progredir com qualidade nos estudos (Costa & Dias, 2015). Para além da detecção precoce do risco de evasão, as instituições de ensino podem contribuir para diminuir a assimetria de informações oferecendo orientação vocacional ao estudante antes da escolha de um curso (Bernárdez-Gómez & Belmonte, 2020; Cabral, 2018) por meio da disponibilização de informações claras sobre os cursos, sobre a profissão (Bernárdez-Gómez & Belmonte, 2020), e sobre o mercado de trabalho.

De modo complementar às ações internas de cada IES, as políticas públicas também cumprem um papel relevante no enfrentamento do fenômeno da evasão e isso fica bem evidente no caso brasileiro com o Plano Nacional de Assistência Estudantil (PNAES), criado em 2010 pelo Ministério da Educação (MEC). Este é considerado um marco em política de enfrentamento da evasão com foco na permanência e finalização do curso superior no âmbito da educação pública federal (Cabral, 2018; Santos Junior et al., 2020). Na esfera das instituições privadas de ensino superior, a política pública prioriza o financiamento estudantil por meio do Fundo de Financiamento Estudantil (FIES) e o Universidade para Todos (PROUNI), o que não deixa de ser uma forma de apoiar a permanência do estudante no ensino superior, embora isso não esteja explícito nesses instrumentos.

As políticas públicas existentes no Brasil de apoio à permanência ou de financiamento estudantil, cabe afirmar, endereçam apenas parte dos fatores causadores do fenômeno da evasão. Contudo, há questões estruturais que podem acentuar o quadro de exclusão de grupos minoritários e em vulnerabilidade, fatores políticos e econômicos como a “crescente concentração da renda e da propriedade; a extinção de postos de trabalho; o desemprego; a transformação de direitos universais em serviços oferecidos no mercado” (Azevedo et al., 2010, p. 3). Esses aspectos, associados com a pandemia causada pela COVID-19, podem aumentar ainda mais o processo de evasão e acentuar as desigualdades (Nunes, 2021).

Feitas essas considerações, acredita-se que este trabalho permitiu avaliar como o campo específico da predição de evasão tem tratado o tema do ponto de vista metodológico, das variáveis determinantes e do pós-predição, isto é, das ações de intervenção adotadas. Os resultados podem contribuir com a área de computação pela apresentação do estado da arte das técnicas preditivas

empregadas. Os resultados também dialogam com a gestão educacional, pela compreensão de como as pesquisas têm desdobrado o tema para as ações preventivas de intervenção visando à permanência e êxito dos estudantes. Ou seja, a detecção prévia do risco de evadir é fundamental para munir os gestores de informação, mas é somente uma das primeiras etapas. A partir da previsão, os gestores têm o desafio de definir ações institucionalizadas de enfrentamento da evasão, planejando e executando ações de acolhimento, intervenção pedagógica, dentre outras, a depender dos fatores preponderantes presentes no cálculo do risco de o estudante evadir.

Outra contribuição deste trabalho se dá na esfera das políticas públicas, pois pelo conhecimento mais acurado de quais fatores são determinantes para a evasão nas universidades do sistema nacional de ensino superior, políticas de permanência e êxito podem ser desenhadas de forma mais assertiva. Projetos-piloto abrangentes, envolvendo múltiplas instituições, que visem desenvolver soluções preditivas de risco de evasão e metodologias de intervenção e acompanhamento pedagógico podem fornecer evidências para a concepção dessas políticas.

É necessário ressaltar que o presente trabalho, ao utilizar a técnica de revisão sistemática de literatura (RSL), apresenta a limitação inerente à escolha das bases de dados e à definição das *strings* de busca. Embora tenham sido utilizadas bases de dados relevantes no campo da pesquisa em predição de evasão, escolhas e definições de *strings* diferentes poderiam implicar na recuperação de outros estudos, não contemplados.

Assim, para estudos futuros baseados em RSL, propõe-se que a cobertura de bases de dados científicas seja ampliada, inclusive aplicando outras estratégias como busca manual em eventos nacionais e internacionais de renome e que não tenham sido indexados pelas bases de dados utilizadas neste estudo. Adicionalmente, recomenda-se que pesquisas vindouras de revisão de literatura ou teórico-empíricas dediquem-se à seguinte agenda: (i) proposição e aplicação de ações de enfrentamento da evasão além da mera predição; (ii) inclusão de variáveis ainda pouco exploradas como dados do ensino médio, saúde e atividades extracurriculares dos alunos; (iii) criação de bases de dados públicas para ampliar a pesquisa, como já ocorre com a Plataforma Nilo Peçanha, criada pela Rede Federal de Educação Profissional, Científica e Tecnológica; e, (iv) avaliação dos impactos das ações de gestão educacional na evasão tanto por meio de abordagens qualitativas quanto quantitativas.

Referências

- Agrusti, F., Bonavolontà, G., & Mezzini, M. (2019). University dropout prediction through educational data mining techniques: A systematic review. *Journal of E-Learning and Knowledge Society*, 15(3), 161–182.
- Alturki, S., Hulpuş, I., & Stuckenschmidt, H. (2020). Predicting academic outcomes: A survey from 2007 till 2018. *Technology, Knowledge and Learning*. <https://doi.org/10.1007/s10758-020-09476-0>
- Azevedo, J. C., Rodrigues, C. C., & Curço, S. F. (2010). Escola cidadã: Políticas e práticas inclusivas. *Arquivos Analíticos de Políticas Educativas - Education Policy Analysis Archives*, 18(2), 1–23. <https://doi.org/10.14507/epaa.v18n2.2010>
- Berger, J. B., Ramírez, G. B., & Lyons, S. (2012). Past to present: A historical look at retention. In A. Seidman (Ed.), *College student retention: Formula for student success* (2nd ed., pp. 7–33). Rowman & Littlefield Publishers.
- Bernárdez-Gómez, A., & Belmonte, M. L. (2020). Evasão escolar, determinantes, políticas educacionais e itinerários subsequentes. *Research, Society and Development*, 9(10), 1–15. <https://doi.org/10.33448/rsd-v9i10.9234>

- Cabral, L. S. A. (2018). Políticas de ações afirmativas, pessoas com deficiência e o reconhecimento das identidades e diferenças no ensino superior brasileiro. *Arquivos Analíticos de Políticas Educativas*, 26(57), 1–33. <https://doi.org/10.14507/epaa.26.3364>
- Campbell, J., DeBlois, P., & Oblinger, D. (2007). Academic analytics: A new tool for a new era. *EDUCAUSE Review*, 42(4), 40–57.
- Colpo, M. P., Primo, T. T., Pernas, A. M., & Cechinel, C. (2020). Mineração de dados educacionais na previsão de evasão: uma RSL sob a perspectiva do congresso brasileiro de informática na educação. *Anais Do XXXI Simpósio Brasileiro de Informática Na Educação (SBIE 2020)*, 1102–1111. <https://doi.org/10.5753/cbie.sbie.2020.1102>
- Costa, S. L., & Dias, S. M. B. (2015). A permanência no ensino superior e as estratégias institucionais de enfrentamento da evasão. *Journal de Políticas Educacionais*, 9(17), 51–60. <https://doi.org/10.5380/jpe.v9i17/18.38650>
- Devasia, T., Vinushree T P., & Hegde, V. (2016). Prediction of students performance using educational data mining. *2016 International Conference on Data Mining and Advanced Computing (SAPIENCE)*, 91–95. <https://doi.org/10.1109/SAPIENCE.2016.7684167>
- Dias, M. M., Silva Filho, L. A., Lino, A. D. P., Favero, E. L., & Ramos, E. M. L. S. (2008). Aplicação de técnicas de mineração de dados no processo de aprendizagem na educação a distância. *XIX Simpósio Brasileiro de Informática Na Educação*, 105–114. <https://doi.org/10.5753/cbie.sbie.2008.105-114>
- EAB. (2019). *The evolution of student success and 360+ best practices to help you adapt*. eab.com/studentsuccess
- Ekowo, M., & Palmer, I. (2016). *The promise and peril of predictive analytics in higher education: A landscape analysis*. <https://files.eric.ed.gov/fulltext/ED570869.pdf>
- Ekowo, M., & Palmer, I. (2017). *Predictive analytics in higher education: Five guiding principles for ethical use*. <https://reportcenter.highered.texas.gov/sites/docid/cache/file/AC2E2644-9DF0-2DF2-467C1FD0F2571918.pdf>
- Freitas, Francisco, A., Vasconcelos, F. F., Peixoto, S. A., Hassan, M. M., Dewan, M. A., Albuquerque, V. H. C., & Rebouças Filho, P. P. (2020). IoT system for school dropout prediction using machine learning techniques based on socioeconomic data. *Electronics*, 9(10), 1612–1625.
- Georgia State University (GSU). (2017). *Case study: Georgia State University. Innovation in higher education*. https://agb.org/sites/default/files/u27335/casestudy_innovation_gsu.pdf
- Georgia State University (GSU). (2021). *Summer Success Academy*. <https://success.gsu.edu/initiatives/summer-success-academy/>
- Gutierrez, A. P., Oliveira, R. C., Guardiano, L. S., Carminatti, S. S. H., & Silva, F. C. (2018). Central de tutoria e monitoria: Uma experiência inovadora em EAD com excelência no atendimento ao aluno. *Série Educar - Tecnologia*, 10, 1–10.
- Han, J., Pei, J., & Kamber, M. (2011). *Data mining: Concepts and techniques*. Elsevier.
- Hasbun, T., Araya, A., & Villalon, J. (2016). Extracurricular activities as dropout prediction factors in higher education using decision trees. *2016 IEEE 16th International Conference on Advanced Learning Technologies (ICALT)*, 242–244. <https://doi.org/10.1109/ICALT.2016.66>
- Hegde, V., & Prageeth, P. P. (2018). Higher education student dropout prediction and analysis through educational data mining. *2018 2nd International Conference on Inventive Systems and Control (ICISC)*, 694–699. <https://doi.org/10.1109/ICISC.2018.8398887>
- Hernandez, E. J., Duque, N., Quintero, D. P., & Naranjo, J. C. E. (2018). Educational data mining for the analysis of student desertion. *Conference: LALA 2018*, 1–10.
- Hussain, S., Abdulaziz Dahan, N., Ba-Alwi, F. M., & Ribata, N. (2018). Educational data mining and analysis of students' academic performance using WEKA. *Indonesian Journal of Electrical*

- Engineering and Computer Science*, 9(2), 447–459. <https://doi.org/10.11591/ijeecs.v9.i2.pp447-459>
- Ice, P., Días, S., Swan, K., Burgess, M., Sharkey, M., Sherrill, J., Huston, D., & Okimoto, H. (2012). The PAR framework proof of concept: Initial findings from a multi-institutional analysis of federated postsecondary data. *Journal of Asynchronous Learning Networks*, 16(3), 63–86.
- Išljamović, S. Z., Jeremić, V., & Lalić, S. (2016). Indicators of study success related to impact of university students' enrollment status. *Croatian Journal of Education - Hrvatski Časopis Za Odgoj i Obrazovanje*, 18(2), 583–606. <https://doi.org/10.15516/cje.v18i2.1003>
- Kamal, P., & Ahuja, S. (2019). An ensemble-based model for prediction of academic performance of students in undergrad professional course. *Journal of Engineering, Design and Technology*, 17(4), 769–781. <https://doi.org/10.1108/JEDT-11-2018-0204>
- Khasanah, A. U., & Harwati. (2017). A comparative study to predict student's performance using educational data mining techniques. *IOP Conference Series: Materials Science and Engineering*, 215, 1–7. <https://doi.org/10.1088/1757-899X/215/1/012036>
- Kitchenham, B., Pearl Brereton, O., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2009). Systematic literature reviews in software engineering – A systematic literature review. *Information and Software Technology*, 51(1), 7–15. <https://doi.org/10.1016/j.infsof.2008.09.009>
- Kurzweil, M., & Wu, D. D. (2015). *Building a pathway to student success at Georgia State University*. https://scholarworks.gsu.edu/cgi/viewcontent.cgi?article=1141&context=ghpc_articles
- Lima, F. S., & Zago, N. (2017). Evasão no ensino superior: Desafios conceituais. *Congressos CLABES*.
- Liz-Domínguez, M., Caeiro-Rodríguez, M., Llamas-Nistal, M., & Mikic-Fonte, F. A. (2019). Systematic literature review of predictive analysis tools in higher education. *Applied Sciences*, 9(24), 5569. <https://doi.org/10.3390/app9245569>
- Martinho, V. R. D. C., Nunes, C., & Minussi, C. R. (2013). An intelligent system for prediction of school dropout risk group in higher education classroom based on artificial neural networks. *2013 IEEE 25th International Conference on Tools with Artificial Intelligence*, 159–166. <https://doi.org/10.1109/ICTAI.2013.33>
- Maschio, P., Vieira, M. A., Costa, N., Melo, S. De., & Júnior, C. P. (2018). Um panorama acerca da mineração de dados educacionais no brasil. *Anais Do SBIE 2018*, 1936. <https://doi.org/10.5753/cbie.sbie.2018.1936>
- Noboa, C., Ordóñez, M., & Magalhães, J. (2018). Statistical learning to detect potential dropouts in higher education: A public university case study. *Learning Analytics for Latin America 2018*, 12–21.
- Nunes, R. C. (2021). Um olhar sobre a evasão de estudantes universitários durante os estudos remotos provocados pela pandemia do COVID-19. *Research, Society and Development*, 10(3), 1–13. <https://doi.org/10.33448/rsd-v10i3.13022>
- Olaya, D., Vásquez, J., Maldonado, S., Miranda, J., & Verbeke, W. (2020). Uplift modeling for preventing student dropout in higher education. *Decision Support Systems*, 134, 113320. <https://doi.org/10.1016/j.dss.2020.113320>
- Palmer, I. (2018). *Choosing a predictive analytics vendor: A guide for colleges*. <https://www.newamerica.org/education-policy/reports/choosing-predictive-analytics-vendor-guide/>
- Park, J.-H., & Choi, H. J. (2009). Factors influencing adult learners' decision to drop out or persist in online learning. *Journal of Educational Technology & Society*, 12(4), 207–217.
- Paterson, J. (2019). *What a predictive analytics experiment taught 11 colleges about sharing data*. Deep Dive. <https://www.highereddive.com/news/what-a-predictive-analytics-experiment-taught-11-colleges-about-sharing-dat/552986/>

- Picciano, A. G. (2012). The Evolution of big data and learning analytics in American higher education. *Journal of Asynchronous Learning Networks*, 16(3), 9–20.
- Proaño, J. P. Z., & Villamar, V. C. P. (2018). Systematic mapping study of literature on educational data mining to determine factors that affect school performance. *2018 International Conference on Information Systems and Computer Science (INCISCOS)*, 239–245. <https://doi.org/10.1109/INCISCOS.2018.00042>
- Rovai, A. P. (2003). In search of higher persistence rates in distance education online programs. *The Internet and Higher Education*, 6(1), 1–16. [https://doi.org/10.1016/S1096-7516\(02\)00158-6](https://doi.org/10.1016/S1096-7516(02)00158-6)
- Santos Junior, J. da S., Magalhães, A. M. da S., & Real, G. C. M. (2020). A gestão da evasão nas políticas educacionais brasileiras: Da graduação à pós-graduação stricto sensu. *ETD-Educação Temática Digital*, 22(2), 460–478. <https://doi.org/10.20396/etd.v22i2.8654823>
- Sarra, A., Fontanella, L., & Di Zio, S. (2019). Identifying students at risk of academic failure within the educational data mining framework. *Social Indicators Research*, 146(1–2), 41–60.
- Scannavino, K. R. F., Nakagawa, E. Y., Fabbri, S. C. P. F., & Ferrari, Fa. C. (2017). *Revisão sistemática da literatura em engenharia de software: Teoria e prática*. LTC.
- SEMESP. (2019). *Mapa do ensino superior no Brasil*.
- Sharma, S., Agrawal, J., Agarwal, S., & Sharma, S. (2013). Machine learning techniques for data mining: A survey. *2013 IEEE International Conference on Computational Intelligence and Computing Research*, 1–6. <https://doi.org/10.1109/ICCIC.2013.6724149>
- Silva, F. C., Cabral, T. L. de O., & Pacheco, A. S. V. (2020). Evasão ou permanência? Modelos preditivos para a gestão do ensino superior. *Arquivos Analíticos de Políticas Educativas - Education Policy Analysis Archives*, 28(1), 136.
- University Innovation Alliance.(UIA). (2020). *Predictive analytics: Building the infrastructure for student success*. <https://theuia.org/sites/default/files/2020-08/UIA-Scale-Project-Predictive-Analytics.pdf>
- University Innovation Alliance.(UIA). (2021a). *Our commitment*. <https://theuia.org/>
- University Innovation Alliance. (UIA). (2021b). *Predictive analytics*. <https://theuia.org/project/predictive-analytics>
- Vaclavek, J., Kuzilek, J., Skocilas, J., Zdrahal, Z., & Fuglik, V. (2018). Learning analytics dashboard analysing first-year engineering students. *European Conference on Technology Enhanced Learning*, 575–578. https://doi.org/10.1007/978-3-319-98572-5_48
- Wagner, E., & Longanecker, D. (2016). Scaling student success with predictive analytics: Reflections after four years in the data trenches. *Change: The Magazine of Higher Learning*, 48(1), 52–59. <https://doi.org/10.1080/00091383.2016.1121087>
- Zhang, L., & Rangwala, H. (2018). Early identification of at-risk students using iterative logistic regression. In *Artificial intelligence in education* (pp. 613–626). Springer. https://doi.org/10.1007/978-3-319-93843-1_45

Sobre os Autores

Marcelo Ferreira Tete

Universidade Federal de Goiás (UFG), Goiânia, Goiás, Brasil

marcelo_tete@ufg.br

ORCID: <https://orcid.org/0000-0002-2780-7977>

Professor do curso de Administração da Faculdade de Administração, Ciências Contábeis e Economia (FACE) da Universidade Federal de Goiás (UFG). É doutor em Administração pela

Universidade de Brasília (UnB). Desenvolve pesquisas sobre Transições Sociotécnicas e Sistemas de Inovação em áreas como Educação Superior, Energia Renovável, Mobilidade Urbana e Cidades Inteligentes. É membro da Sustainability Transitions Research Network (STRN) e membro fundador da Sustainability Transitions Brazil (STB).

Marcos de Moraes Sousa

Instituto Federal Goiano (IF Goiano), Ceres, Goiás, Brasil
marcos.moraes@ifgoiano.edu.br

ORCID: <https://orcid.org/0000-0002-0901-0550>

Doutor em Administração pela Universidade de Brasília (UnB). Professor permanente no Programa de Pós-Graduação em Administração da Universidade Federal de Goiás (PPGADM – UFG). Professor permanente no Programa de Pós-Graduação em Educação Profissional e Tecnológica do Instituto Federal Goiano (ProfEPT – IF Goiano). Pesquisa temas relacionados à Administração da Justiça, Administração Pública, Permanência e Evasão de alunos, e, Inovação no Setor Público.

Thalia Santos de Santana

Instituto Federal Goiano (IF Goiano), Ceres, Goiás, Brasil
thalia.santana@ifgoiano.edu.br

ORCID: <https://orcid.org/0000-0003-1899-4819>

Professora dos cursos de Informática do Campus Ceres do Instituto Federal Goiano (IF Goiano). Possui graduação em Bacharelado em Sistemas de Informação e atualmente é mestranda no Programa de Pós-graduação em Ciência da Computação (PPGCC) do Instituto de Informática da Universidade Federal de Goiás (INF – UFG).

Salatyel Fellipe

Universidade Federal de Goiás (UFG), Goiânia, Goiás, Brasil
salatyel_fellipe@discente.ufg.br

ORCID: <https://orcid.org/0000-0003-2180-9258>

Especialista em Docência do Ensino Superior pela Faculdade Brasileira de Educação e Cultura (FABEC), e Gestão em Processos Gerenciais pela Faculdade Educacional da Lapa (FAEL). Possui graduação em Engenharia de Computação pela Universidade Federal de Goiás (UFG). É membro das redes de pesquisadores em Gestão Pública e Sustentabilidade da UFG e Sustainability Transitions Brazil (STB).

arquivos analíticos de políticas educativas

Volume 30 Número 149

4 de octubre de 2022

ISSN 1068-2341



Este artigo pode ser copiado, exibido, distribuído e adaptado, desde que a atribuição e os créditos sejam dados ao(s) autor(es) e **Arquivos Analíticos de Políticas Educativas**, as alterações sejam identificadas, e a mesma licença se aplique à obra derivada. Mais detalhes sobre a licença Creative Commons podem ser encontrados em <https://creativecommons.org/licenses/by-sa/4.0/>. Qualquer outro uso deve ser aprovado em conjunto com AAPE/EPAA ou o(s) autor(es). AAPE/EPAA é publicada pela *Mary Lou Fulton Teachers College* da *Arizona State University*. Os artigos que aparecem na AAPE/EPAA são indexados em CIRC (Clasificación Integrada de Revistas Científicas, Espanha), DIALNET (Espanha),

[Directory of Open Access Journals](#), EBSCO Education Research Complete, ERIC, Education Full Text (H.W. Wilson), PubMed, QUALIS A1 (Brasil), Redalyc, SCImago Journal Rank, SCOPUS, SOCOLAR (China).

Para notificar erros e fazer sugestões, entre em contato com fischman@asu.edu

Siga a comunidade da AAPE no Facebook em <https://www.facebook.com/EPAAAPE> e no **Twitter** @epaa_aape.