

Evasão em cursos superiores na área de Computação: Um Mapeamento Sistemático da Literatura

Pedro C. C. Mundim¹, Thais Regina de M. B. Silva¹,
Gláucia Braga e Silva¹, Daniel M. Barbosa¹

¹Instituto de Ciências Exatas e Tecnológicas – Universidade Federal de Viçosa (UFV)
Rodovia LMG 818, km 06 – 35690-000 – Florestal – MG – Brazil

{pedro.mundim, thais.braga, glaucia, danielmendes}@ufv.br

Abstract. *Student dropout in higher education computer courses affects both students and institutions, impacting the university structure, academic trajectories, and the job market. This paper presents a Systematic Literature Mapping (SLM) on this issue, based on 28 studies found in the SOL, ACM, Scopus, and IEEE databases. The results indicate that the field of computer science is well represented in the analyzed studies (18/28), and that exploratory analysis and the use of machine learning (ML) algorithms were the predominant methodologies in the primary studies. The main attributes related to dropout were identified, along with a trend toward developing models to predict dropout rates.*

Resumo. *A evasão estudantil nos cursos superiores de Computação afeta tanto os estudantes quanto as instituições, impactando a estrutura universitária, a trajetória acadêmica e o mercado de trabalho. Este artigo apresenta um Mapeamento Sistemático da Literatura (MSL) sobre essa questão, com base em 28 trabalhos encontrados nas bases SOL, ACM, Scopus e IEEE. Os resultados indicam que a área de Computação é bem representativa nos estudos analisados (18/28) e que a análise exploratória e o uso de algoritmos de aprendizado de máquina (ML) foram as metodologias predominantes nos estudos primários. Também foram identificados e categorizados os principais atributos relacionados à evasão e a tendência de desenvolver modelos para prever taxas de evasão.*

1. Introdução

A evasão no ensino superior impacta alunos e instituições em todo o mundo, gerando consequências sociais e financeiras que afetam a estrutura universitária e o percurso acadêmico dos discentes [Da Cruz et al. 2023]. Considerando a área de Computação, [Duran et al. 2023] destacam que a evasão está relacionada a questões estruturais no sistema educacional, influenciadas por fatores acadêmicos, socioeconômicos, de etnia/raça, de gênero e emocionais, internos e externos às instituições. Analisar a evasão nesse contexto é essencial, dado seu histórico elevado e a crescente demanda por profissionais em um mercado em expansão [Menolli et al. 2020].

Este trabalho apresenta um Mapeamento Sistemático da Literatura (MSL), que tem como objetivo identificar e analisar as pesquisas mais recentes sobre a evasão no ensino superior, com foco em cursos da área de Computação, destacando as abordagens metodológicas utilizadas e como foram aplicadas. A necessidade deste MSL está em,

portanto, consolidar o conjunto de publicações encontrados na literatura, fornecendo às instituições e pesquisadores insights que contribuam para mitigar o problema.

Como principais resultados deste MSL, observou-se que as abordagens metodológicas utilizadas pelos trabalhos primários considerados neste contexto são a análise exploratória e a modelagem de algoritmos de aprendizado de máquina para se tentar prever a evasão. A partir desta análise, foi possível ainda construir uma categorização dos atributos utilizados, composta por quatro dimensões: Demográfico Socioeconômico e Pessoal, Conhecimento Prévio, Institucional e Desempenho Acadêmico.

O restante do artigo está organizado de forma que: a Seção 2 traz os trabalhos relacionados encontrados na literatura; a Seção 3 descreve o protocolo para a condução do MSL; a Seção 4 apresenta os resultados; e, a Seção 5 traz as considerações finais.

2. Trabalhos Relacionados

Para identificar os trabalhos relacionados, foram utilizadas as bases SOL, ACM, Scopus e IEEE, usando os termos da Tabela 2. A seleção incluiu artigos de 2018 a 2023, em inglês e português, que tratassesem da evasão no ensino superior e fossem Revisões Sistemáticas da Literatura (RSL) ou MSL, visando consolidar estudos sobre o tema.

Em [Silva et al. 2023] foi realizada uma RSL com objetivo de identificar os principais fatores associados à predição de abandono e retenção no ensino superior utilizados por estudos primários. As bases de dados utilizadas pelos autores foram SCOPUS e IEEE (sem o uso de bases nacionais), tendo sido selecionados pela string de busca, 52 trabalhos para serem analisados. A string foi composta pelos termos: *student, undergraduate, predict, dropout, retention, attrition, metric, measurement, indicate, higher education e bachelor degree*. Os resultados revelaram 29 fatores utilizados pelos trabalhos, em que alguns dos mais citados foram média geral, gênero, curso, idade, etnia, horário de curso e tipo de ingresso na universidade. Esses fatores foram classificados pelos autores em três grandes grupos: demográficos, acadêmicos e de aprendizagem. Gênero, média geral e regularidade de estudo foram os fatores mais recorrentes em cada grupo, respectivamente.

Os autores em [da Silva et al. 2019] apresentam um MSL que investiga a área de Mineração de Dados Educacionais (EDM), com foco na identificação de tendências e contribuições para estudos futuros. Os 357 trabalhos primários foram selecionados a partir das bases da ACM, IEEE e por consultas ao Google Acadêmico (sem o uso de bases nacionais), por meio de uma string de busca que possuía os termos *Mineração de dados Educacionais e Educational Data Mining*. Eles foram categorizados em três critérios: foco de pesquisa, tipo de contribuição e tipo de pesquisa. Os resultados revelaram que a maioria dos estudos está centrada no desempenho acadêmico dos alunos e na identificação de fatores relacionados à evasão escolar, com foco na predição e relação entre eles.

O trabalho de [de Oliveira and Barbosa 2023] descreve uma RSL sobre o uso de modelos de aprendizagem de máquina multiníveis para análise e previsão de evasão escolar. O objetivo dos autores foi oferecer uma visão abrangente sobre os fatores associados à evasão em diferentes níveis educacionais, além de investigar o uso de técnicas de inteligência artificial no aprimoramento e interpretação desses modelos. Foram utilizadas as bases de pesquisa ACM, IEEE, SCOPUS, Elsevier, Engineering Village, Web of Science (sem o uso de bases nacionais) e, no total, 36 estudos primários foram selecionados. Alguns termos de busca foram: *multilevel, hierarchical, nested, mixed, random coefficients,*

analysis, model, academic, student, dropout e evasion. Os achados da pesquisa destacaram que o ambiente educacional mais pesquisado é o ensino médio, seguido pelos níveis secundário e superior. Os fatores de impacto para a evasão, na modelagem dos trabalhos primários, foram divididos em dois níveis: estudantil e escolar. No primeiro, prevaleceram aspectos demográficos e socioeconômicos; no segundo, questões estruturais. Entre os métodos mais usados, sobressaíram a regressão logística e linear.

Em [Colpo et al. 2020], os autores realizaram uma RSL sobre a previsão da evasão estudantil com técnicas de EDM, analisando trabalhos do Congresso Brasileiro de Informática na Educação. Foram encontrados 104 trabalhos no período de 2006 - 2019, dos quais 23 foram para análise. Alguns termos utilizados para a seleção foram *evasão e dropout*. Os resultados indicaram que a maioria das pesquisas utiliza a classificação, destacando os algoritmos de árvores de decisão implementados no Weka.

Já [dos Santos et al. 2021b] analisaram 50 trabalhos nacionais e internacionais sobre EDM aplicados à evasão escolar, examinando ferramentas, algoritmos e bases de dados. Os termos da busca foram *prediction of students, school dropout, school retention, school failure, educational data mining, knowledge discovery, machine learning, institution e university*. Os resultados indicam a predominância do Weka e do Scikit-learn e, entre os algoritmos mais recorrentes, destacam-se Árvore de Decisão, Naive Bayes, Random Forest e SVM, com os dois últimos apresentando melhor desempenho. A maioria das bases de dados analisadas possui entre 501 e 5.000 registros e contém majoritariamente atributos acadêmicos, seguidos por fatores socioeconômicos.

A Tabela 1 compara os trabalhos discutidos com este MSL. Diferente dos estudos anteriores, que abrangem diversos níveis de ensino ou áreas não especificadas, este MSL foca exclusivamente em cursos da área de Computação no ensino superior. Além disso, este trabalho considera um intervalo temporal mais recente, de 2018 a 2023, o que permite analisar tendências e padrões atuais, diferentemente, por exemplo, de [Silva et al. 2023] e [de Oliveira and Barbosa 2023], que analisaram períodos iniciados nas décadas de 1990. Por fim, os trabalhos de [da Silva et al. 2019], [de Oliveira and Barbosa 2023], [Colpo et al. 2020] e [dos Santos et al. 2021b] estão restritos a um recorte de pesquisas primárias mais específicas, sendo, trabalhos sobre EDM e modelos de ML.

Tabela 1. Comparativo entre trabalhos relacionados.

Trabalho	Tipo de revisão	Nível de ensino	Curso(s)	Intervalo Temporal	Quantidade de Trabalhos
[Silva et al. 2023]	RSL	Superior	Não especificado	1993-2021	52
[da Silva et al. 2019]	MSL	Técnico e Superior	Não especificado	2014-2019	357
[de Oliveira and Barbosa 2023]	RSL	Fundamental, Médio, e Superior	Não especificado	1997-2022	36
[Colpo et al. 2020]	RSL	Fundamental, Médio, Técnico e Superior	Não especificado	2006-2019	23
[dos Santos et al. 2021b]	RSL	Fundamental, Médio, Técnico e Superior	Não especificado	2008-2020	50
Este trabalho	MSL	Superior	Área de Computação	2018-2023	28

3. Mapeamento Sistemático da Literatura

A estrutura deste MSL foi definida em três fases – planejamento, condução e documentação – seguindo as diretrizes estabelecidas para estudos dessa natureza [Silva et al. 2023]. Esse delineamento permite organizar o estudo de forma sistemática, assegurando que cada etapa contribua para a construção de um panorama consistente e embasado sobre o tema investigado.

Na fase de planejamento, as seguintes questões de pesquisa foram definidas: **QP1** - Qual é o panorama geral das pesquisas sobre evasão no ensino superior e, em particular,

como essas pesquisas se relacionam com a área de Computação?; **QP2** - Como as abordagens metodológicas encontradas têm sido aplicadas para se analisar a evasão no ensino superior em cursos da área de Computação?

Em seguida, foi definida a string de busca (Tabela 2) a ser utilizada nas buscas avançadas das bases de pesquisa. A string abrange duas partes: termos relacionados à evasão e retenção; e termos referentes a tipos de Instituições de Ensino Superior (IESs).

Tabela 2. String de Busca para o MSL.

Palavras-Chave	
Evasão/Retenção	(dropout, attrition, abandonment, retention, failure, desertion, transfer, change of course, persistence)
Tipos de IESs	(public university, federal university, federal institute, higher education institution)
String de Busca	("dropout" OR "attrition" OR "abandonment" OR "retention" OR "failure" OR "desertion" OR "transfer" OR "change of course" OR "persistence") AND ("public university" OR "federal university" OR "federal institute" OR "higher education institution")

Para execução da string de busca, foram consideradas quatro bases científicas: SOL¹, ACM Digital Library², Scopus³ e IEEE Xplore⁴, por serem as principais bases de estudos científicos da área de Computação. Por fim, para seleção dos estudos, a Tabela 3 apresenta os Critérios de Inclusão (CI) e de Exclusão (CE) definidos de modo a selecionar os trabalhos de maior qualidade e relevância para este estudo.

Tabela 3. Critérios de Inclusão e de Exclusão.

Critérios de Inclusão
CI1 - Trabalhos que abordem evasão no ensino superior presencial, incluindo os que englobam cursos da área de Computação.
CI2 - Trabalhos primários.
Critérios de Exclusão
CE1 - Trabalhos duplicados.
CE2 - Trabalhos que não estejam em inglês ou português.
CE3 - Trabalhos publicados até 2022 que não foram citados nenhuma vez.
CE4 - Trabalhos publicados em veículos sem Qualis ou com Qualis C.

A fase de condução abrangeu a execução da string de busca em cada uma das bases citadas e a seleção dos trabalhos, a partir da aplicação dos critérios de inclusão e exclusão definidos, conforme ilustra a Figura 1. Visando a seleção de trabalhos mais recentes, este MSL foi realizado nas bases de dados com a filtragem no intervalo de 2018 a 2023. Com a execução da string de busca nas bases, foram retornados 6158 trabalhos.

A partir dos 6158 trabalhos retornados, inicialmente, aplicaram-se os critérios de inclusão para identificar quais estudos estavam alinhados com o tema da evasão no ensino superior presencial, incluindo aqueles relacionados a cursos na área de Computação, e verificar se eram trabalhos primários. Os critérios foram aplicados aos títulos e às palavras-chave dos trabalhos e, em seguida aos resumos, resultando em 83 trabalhos selecionados.

¹SOL: <https://sol.sbc.org.br/index.php/indice>

²ACM: <https://dl.acm.org/>

³Scopus: <https://www.scopus.com/>

⁴IEEE Xplore: <https://ieeexplore.ieee.org/Xplore/home.jsp>

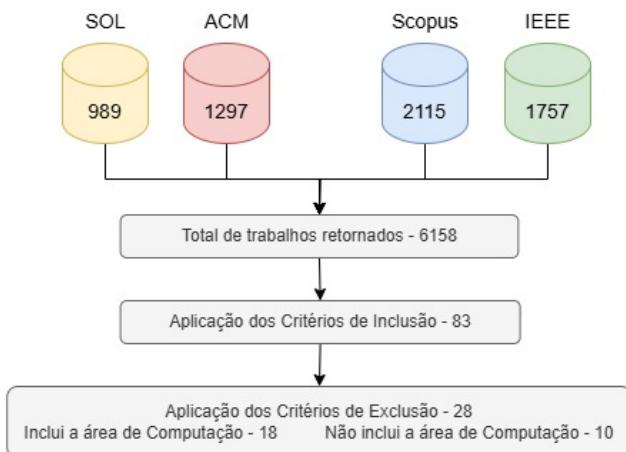


Figura 1. Processo de Seleção dos Trabalhos.

Sobre os 83 trabalhos previamente selecionados, foram aplicados os critérios de exclusão, com os seguintes resultados: o CE1 eliminou 10 trabalhos que eram duplicados; com o CE2 foram excluídos 7 trabalhos que não estavam redigidos em inglês ou português; o CE3 descartou 10 trabalhos que não haviam sido citados nenhuma vez, em que esta verificação foi feita através do Google Acadêmico, verificando um trabalho de cada vez; e por fim, o CE4 eliminou 28 trabalhos publicados em veículos sem classificação Qualis ou com classificação Qualis C. Dessa forma, após a aplicação desses quatro critérios de exclusão, restaram 28 estudos, que foram considerados relevantes para análise. Vale destacar ainda que não houve nenhuma restrição em relação às localidades das instituições que foram estudadas nos trabalhos selecionados.

Por fim, na fase de documentação, os trabalhos selecionados foram sintetizados na Tabela 4, que apresenta cada um dos 28 estudos, com um identificador, sua citação bibliográfica e uma breve descrição. Os trabalhos foram organizados de forma agrupada, sendo que os de identificadores de TR1 a TR18 abordam a evasão em cursos da área de Computação, enquanto os de TR19 a TR28 tratam a evasão em outras áreas.

4. Resultados e Discussões

4.1. QP1 - Qual é o panorama geral das pesquisas sobre evasão no ensino superior e, em particular, como essas pesquisas se relacionam com a área de Computação?

Para responder à primeira pergunta de pesquisa foram analisados os resultados referentes a todos os 28 trabalhos selecionados para as análises.

Na Figura 2a tem-se a distribuição dos anos de publicação dos 28 trabalhos, oferecendo um panorama da evolução das pesquisas sobre evasão no ensino superior entre 2018 e 2023. O maior número de publicações ocorreu em 2021, com 9 trabalhos, possivelmente devido a fatores como a pandemia de COVID-19, que trouxe novos desafios à permanência estudantil. Nos anos seguintes, a distribuição foi mais equilibrada, com uma queda em 2022 e um aumento em 2023, indicando que o tema continua relevante.

A Tabela 5 apresenta a distribuição dos cursos analisados nos trabalhos. Observa-

Tabela 4. Descrição dos 28 Trabalhos selecionados para Análise.

ID	Referência	Descrição
TR1	[Duran et al. 2023]	Investigar fatores relacionados à intenção de abandono em programas de computação.
TR2	[Santos et al. 2019]	Avaliar modelos de predição para identificar o perfil dos estudantes em risco de evasão.
TR3	[Hoed et al. 2018]	Analizar a evasão dos cursos de graduação em Computação no Brasil.
TR4	[Oliveira et al. 2020]	Analizar a evasão em três cursos de computação da UFG e seus fatores contribuintes.
TR5	[Rocha et al. 2021]	Examinar os impactos sociais e econômicos da evasão e seus fatores contribuintes.
TR6	[Costa et al. 2021]	Predizer alunos em risco de evasão com base nas notas dos 3 primeiros semestres.
TR7	[Menolli et al. 2020]	Estudar a evasão nos cursos de sistema de informação com o uso de metodologia BI.
TR8	[Da Cruz et al. 2023]	Propor um método para analisar a evasão em Sistemas de Informação e cursos relacionados.
TR9	[Opazo et al. 2021]	Comparar modelos de ML na previsão da evasão de alunos do primeiro ano de engenharia.
TR10	[Marques et al. 2020]	Identificar os motivos da evasão no curso de Ciência da Computação da UFERSA.
TR11	[Santos et al. 2021]	Utilizar técnicas de mineração de dados para prever evasão e formação de alunos.
TR12	[dos Santos et al. 2021a]	Propor um modelo de Cadeia de Markov para analisar a probabilidade de evasão e conclusão.
TR13	[do Carmo et al. 2022]	Investigar a evasão dos estudantes de um curso de Ciência da Computação da UDESC.
TR14	[Viana et al. 2022]	Descrever um processo de MDE da UFPI para predizer entre alunos evadidos e graduados.
TR15	[Fukao et al. 2023]	Analizar estatísticas de evasão, motivações e avaliar o impacto da pandemia na evasão.
TR16	[Vasconcelos and Andrade 2018]	Investigar as razões da evasão no curso de Licenciatura em Computação da UFRPE.
TR17	[Carvalho et al. 2019]	Identificar os fatores relacionados à evasão dos calouros dos cursos de computação da UFAM.
TR18	[Silva et al. 2021]	Identificar os motivos de evasão dos alunos de Ciência da Computação da UFC.
TR19	[Llauro et al. 2021]	Identificar estudantes em risco de evasão usando dados, questionários e percepção de tutores.
TR20	[Santos et al. 2020]	Propor um modelo que integra árvores de decisão e algoritmos genéticos para classificar a evasão.
TR21	[Nájera and Ortega 2022]	Desenvolver um modelo de ML para antecipar e reduzir a evasão no ensino superior.
TR22	[Da Silva et al. 2019]	Propor três modelos de regressão para prever a evasão em Instituições de Ensino Superior.
TR23	[Pachas et al. 2021]	Aplicar ML para identificar fatores da evasão estudantil e analisar o tempo até sua ocorrência.
TR24	[de Brito et al. 2020]	Identificar atributos relevantes na evasão escolar utilizando técnicas de MDE e ML.
TR25	[Colpo et al. 2021]	Identificar o perfil de evasão em cursos de graduação, utilizando técnicas de MDE e ML.
TR26	[Kantorski et al. 2023]	Apresentar um processo de análise e previsão da evasão utilizando métodos de MDE e ML.
TR27	[Carvalho et al. 2023]	Investigar modelos de MDE, focando no problema da evasão de estudantes no ensino superior.
TR28	[Falcão et al. 2023]	Construir e avaliar um modelo de classificação para prever a evasão de alunos ingressantes.

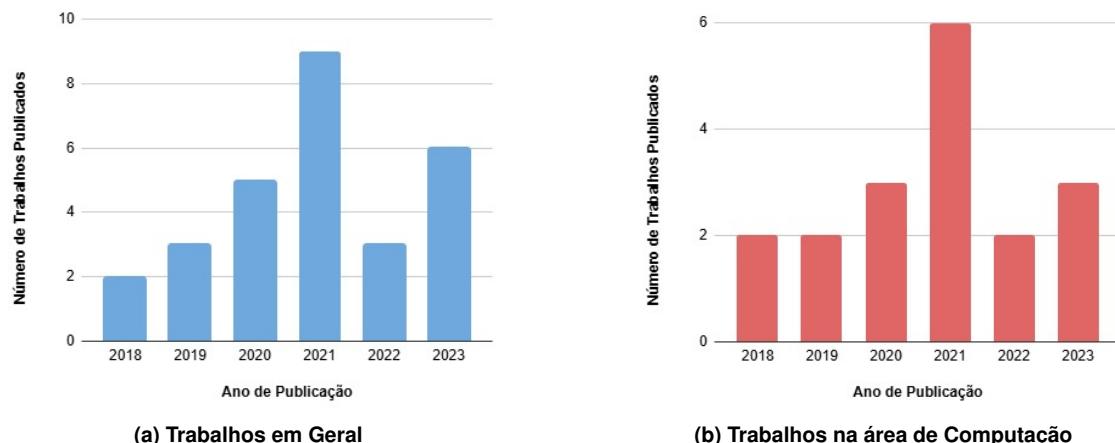


Figura 2. Distribuição dos Anos de Publicação por Trabalhos.

se uma predominância de estudos voltados para a área de Computação, com destaque para Ciência da Computação, que lidera com 13 trabalhos, seguida por Sistemas de Informação com 6. Esse padrão reflete a crescente preocupação com a evasão nos cursos de Computação, que, apesar da alta demanda no mercado, enfrentam altas taxas de evasão devido à carga curricular rigorosa e à complexidade técnica. Em contraste, áreas como Artes, Arquitetura e Ciências Sociais são abordadas por apenas um trabalho cada, evidenciando a lacuna nas investigações sobre evasão em contextos acadêmicos menos explorados. Isso destaca a necessidade de expandir as pesquisas para outras áreas.

Já a Tabela 6 mostra a distribuição dos trabalhos conforme as abordagens metodológicas utilizadas na análise dos cursos. A Análise Exploratória se destaca como a

Tabela 5. Distribuição dos Cursos por Trabalhos.

Curso	Trabalhos	Total
Ciência da Computação	TR1, TR2, TR3, TR4, TR5, TR6, TR10, TR11, TR13, TR14, TR15, TR17, TR18	13
Sistemas de Informação	TR1, TR2, TR3, TR4, TR7, TR8	6
Engenharia da Computação	TR1, TR2, TR3, TR17	4
Engenharia de Software	TR1, TR3, TR4, TR17	4
Licenciatura em Computação	TR3, TR16	2
Engenharias não especificadas	TR19, TR21	2
Engenharia de Informática	TR9	1
Informática	TR15	1
Engenharias Bioinformática	TR9	1
Engenharia Civil	TR9	1
Engenharia Industrial	TR9	1
Engenharia Mecatrônica	TR9	1
Engenharia de Minas	TR9	1
Engenharia Mecânica	TR9	1
Artes	TR19	1
Negócio	TR19	1
Arquitetura	TR19	1
Ciências Sociais	TR21	1
Ciências Administrativas	TR21	1
Química	TR28	1
Física	TR28	1
Ciências Biológicas	TR28	1
Nutrição	TR28	1
Letras (Português Espanhol)	TR28	1

metodologia predominante, empregada em todos os 28 trabalhos revisados, sendo amplamente adotada como uma abordagem inicial para compreender os dados e identificar padrões de evasão no ensino superior. Dentre as técnicas de Análise Exploratória, a EDM foi a mais utilizada. Já o Aprendizado de Máquina (ML) aparece em 17 trabalhos, refletindo o crescente interesse por métodos de análise e predição da evasão.

Tabela 6. Distribuição dos Trabalhos por Abordagem Metodológica.

Abordagem Metodológica	Trabalhos	Total
Análise Exploratória	TR1 - TR28	28
Aprendizado de Máquina	TR2, TR6, TR8, TR9, TR10, TR11, TR14, TR17, TR20, TR21, TR22, TR23, TR24, TR25, TR26, TR27, TR28	17

Para concluir a QP1, foi feito um recorte focado nos 18 trabalhos sobre evasão na área de Computação. A Figura 2b mostra a distribuição dos anos de publicação desses trabalhos. O maior número de publicações foi em 2021, com 6 trabalhos, e, nos demais anos, a distribuição foi mais equilibrada. Os 18 trabalhos que analisam cursos de Computação representam cerca de 64% do total, e ao comparar os gráficos das Figuras 2b e 2a, observa-se que, os dois trabalhos de 2018 abordaram a evasão em cursos de Computação. Já de 2019 a 2021, esses trabalhos representaram 66% da amostra, e, apesar da queda em 2022, o domínio dos estudos sobre evasão em Computação se manteve.

A Tabela 7 apresenta a distribuição dos cursos da área de Computação analisados nos 18 trabalhos. Como comentado, destacam-se os cursos de Ciência da Computação e Sistemas de Informação, que apareceram em 13 e 6 trabalhos respectivamente, seguidos por Engenharia da Computação e Engenharia de Software, com 4 trabalhos cada, e Licenciatura em Computação, com 2 trabalhos. A predominância de Ciência da Computação e Sistemas de Informação reflete o maior número de instâncias desses cursos na plataforma do MEC. Além disso, Ciência da Computação possui um ciclo básico com disciplinas

como Cálculo, Física e Programação, conhecidas pelos altos índices de reprovação, um fator que pode influenciar diretamente a evasão.

Tabela 7. Distribuição de Trabalhos por Cursos: área de Computação.

Curso	Trabalhos	Total
Ciência da Computação	TR1, TR2, TR3, TR4, TR5, TR6, TR10, TR11, TR13, TR14, TR15, TR17, TR18	13
Sistemas de Informação	TR1, TR2, TR3, TR4, TR7, TR8	6
Engenharia da Computação	TR1, TR2, TR3, TR17	4
Engenharia de Software	TR1, TR3, TR4, TR17	4
Licenciatura em Computação	TR3, TR16	2
Engenharia de Informática	TR9	1

No que se refere à abordagem metodológica de análise dos dados, os trabalhos que abordam evasão em cursos da área de Computação seguem o mesmo padrão dos demais, conforme a Tabela 8. Todos os 18 trabalhos utilizaram Análise Exploratória como abordagem inicial para compreensão dos dados e identificação de padrões de evasão, sendo que 9 deles aplicaram EDM. Em contraste, o uso de métodos de análise e predição de evasão por meio de ML foi explorado em cerca de 44% dos trabalhos da área de Computação (8/18), enquanto nos trabalhos de outras áreas esse uso foi de 90% (9/10).

Tabela 8. Distribuição dos Trabalhos por Abordagem Metodológica: área de Computação.

Abordagem Metodológica	Trabalhos	Total
Análise Exploratória	TR1 - TR18	18
Aprendizado de Máquina	TR2, TR6, TR8, TR9, TR10, TR11, TR14, TR17	8

4.2. QP2 - Como as abordagens metodológicas encontradas têm sido aplicadas para se analisar a evasão no ensino superior em cursos da área de Computação?

Para responder à QP2, foram analisadas as abordagens metodológicas e consolidados os atributos extraídos das bases de dados dos trabalhos selecionados da área de Computação.

A Tabela 9 apresenta os 62 atributos identificados nos trabalhos da área de Computação, que utilizam análises exploratórias ou modelos de previsão de evasão no ensino superior por meio de algoritmos de aprendizado de máquina. Esses atributos estão organizados em quatro dimensões principais: *Demográfico, Socioeconômico e Pessoal; Conhecimento Prévio; Institucional; e Desempenho Acadêmico*. Essa categorização é uma das principais contribuições deste trabalho, pois organiza de maneira sistemática os elementos usados para analisar a evasão nos cursos da área de Computação. Ao facilitar a identificação de padrões e a compreensão das variáveis mais relevantes, ela oferece uma base para estudos futuros e também proporciona uma visão geral dos atributos em cada dimensão, que podem ser combinados e explorados em investigações subsequentes.

Ao comparar a categorização deste MSL com os trabalhos relacionados, observa-se que tanto [Silva et al. 2023] quanto [de Oliveira and Barbosa 2023] utilizaram dimensões semelhantes. O primeiro estudo agrupou 29 atributos nas categorias demográficas, acadêmicas e de aprendizagem. Já o segundo identificou 15 atributos, distribuídos em três dimensões: nível estudantil, escolar e outros. No entanto, como este MSL se concentra em trabalhos da área de Computação, uma área de grande relevância atualmente, a análise resultou em um número bem maior: 62 atributos. Essa variedade justificou o uso das quatro dimensões mencionadas para organizá-los de forma eficaz.

Essa categorização revela a diversidade de atributos analisados, abrangendo desde características pessoais e sociais dos alunos até fatores institucionais e de desempenho acadêmico. Também é possível observar quais atributos estão mais relacionados em cada dimensão. Na categoria *Demográfico Socioeconômico e Pessoal*, os mais frequentes foram gênero e idade, em que apareceram em 10 e 6 trabalhos respectivamente, seguidos por cidade/local de residência em 5 trabalhos. Na categoria *Conhecimento Prévio*, o conhecimento em matemática e informática se destacou em 6 trabalhos. Em *Institucional*, os atributos mais recorrentes foram o curso, a estrutura do curso e a forma de ingresso, em 7, 6 e 5 trabalhos, respectivamente. Por fim, em *Desempenho Acadêmico*, destacaram-se a situação de evasão, ano de admissão e número de reprovações, em 9, 5 e 5 trabalhos.

Outra consolidação de dados que foi produzida neste trabalho foi a Tabela 10 que apresenta os algoritmos de ML utilizados nos trabalhos analisados. Ao comparar com o estudo de [de Oliveira and Barbosa 2023], observam-se algoritmos em comum, como SVM, Random Forest, árvores de decisão e regressão logística. No entanto, este MSL não encontrou a regressão linear, a qual foi utilizada no trabalho citado, mas apresentou redes neurais, KNN e Naive Bayes e outros. A ausência de regressão linear pode refletir uma escolha por algoritmos mais adaptados a relações não lineares e dados complexos.

Tabela 10. Técnicas de ML dos Trabalhos que envolvem a área de Computação.

Algoritmo	Trabalhos	Total
Random Forest	TR2, TR6, TR9, TR11, TR14, TR17	6
Árvores de Decisão	TR2, TR6, TR9, TR11, TR14	5
Support Vector Machine (SVM)	TR2, TR9, TR14, TR17	4
K-Nearest Neighbors (KNN)	TR2, TR8, TR9, TR14	4
Regressão Logística	TR6, TR9, TR17	3
Naive Bayes	TR2, TR9, TR14	3
Redes Neurais	TR2, TR9	2
K-Means	TR10	1
Extra Trees (ET)	TR14	1
Multiayer Perception (MLP)	TR14	1

5. Considerações Finais

Este artigo apresentou um Mapeamento Sistemático da Literatura (MSL) sobre evasão no ensino superior, com foco em cursos da área de Computação. Os resultados indicaram que as principais abordagens metodológicas envolvem análise exploratória com uso de EDM e a aplicação de algoritmos de aprendizado de máquina. Além disso, 18 dos 28 trabalhos analisados focaram em cursos da área de Computação. Por fim, foi possível identificar e categorizar os principais atributos relacionados à evasão, além de observar a tendência crescente de desenvolvimento de modelos preditivos para taxas de evasão.

Como contribuições deste estudo, destacam-se: o panorama geral das pesquisas sobre evasão no ensino superior, com recorte para a área de Computação, destacando as abordagens metodológicas aplicadas; e a categorização dos atributos analisados em: Demográfico Socioeconômico e Pessoal, Conhecimento Prévio, Institucional e Desempenho Acadêmico, que fornece uma base para estudos futuros relacionados ao desenvolvimento de modelos de predição de evasão. Além disso, pesquisas futuras podem incluir levantamentos junto às instituições de ensino superior sobre estratégias para mitigar a evasão, permitindo a avaliação de seus impactos e a identificação de novos atributos relevantes.

Tabela 9. Dimensões dos atributos e identificadores de trabalhos.

Demográfico, Socioeconômico e Pessoal		
Atributo	Trabalhos	Total
Gênero	TR1, TR4, TR6, TR7, TR8, TR9, TR14, TR15, TR17, TR18	10
Idade	TR1, TR6, TR8, TR9, TR14, TR18	6
Cidade / Local de residência	TR5, TR8, TR9, TR10, TR14	5
Raça / Etnia	TR1, TR7, TR10, TR14	4
Região / Região de Origem (do país)	TR7, TR9, TR10, TR14	4
Apoio Estudantil	TR5, TR18	2
Estado de residência	TR7, TR14	2
Estado Civil	TR14, TR17	2
Deslocamento / Tipo de Transporte	TR17, TR18	2
Estatuto Socioeconômico	TR1, TR17	2
Naturalidade	TR6	1
Nacionalidade	TR7	1
Relação Conjugal dos pais	TR10	1
Tipo de moradia (alugada / própria)	TR17	1
Quantidade de Filhos	TR17	1
Acesso à internet / computador	TR17	1
Impossibilidade de Trabalhar e Estudar	TR18	1
Problemas de Saúde Mental	TR18	1
Dificuldade de Foco e de Organização	TR18	1
Falta de Motivação	TR18	1
Conhecimento Prévio		
Atributo	Trabalhos	Total
Conhecimento Matemático / Conhecimento em Informática	TR3, TR9, TR10, TR16, TR17, TR18	6
Escola / Grau Acadêmico (Público / Privado)	TR8, TR9, TR14	3
Preferência / Escolha do curso	TR9, TR17	2
Graduação Prévia	TR14, TR17	2
Pontuação no ensino médio	TR9	1
Pontuação nos testes de Admissão Universitária	TR9	1
Posição em escala de notas no ensino médio	TR9	1
Pontuação em Linguagem	TR9	1
Pontuação em Teste Nacional	TR9	1
Institucional		
Atributo	Trabalhos	Total
Curso	TR1, TR4, TR5, TR11, TR15, TR17, TR18	7
Estrutura do curso / Disciplinas	TR4, TR5, TR11, TR13, TR16, TR18	6
Forma de Ingresso	TR10, TR14, TR15, TR17, TR18	5
Tipo de instituição (pública ou privada)	TR6, TR7	2
Corpo docente (nome ou ID)	TR4, TR18	2
Grade Curricular Desatualizada	TR5, TR18	2
Mercado de trabalho	TR5, TR18	2
Estrutura Física da Instituição	TR5, TR18	2
Recursos pedagógicos	TR5	1
Estrutura Administrativa da Instituição	TR5	1
Acompanhamento discente	TR5	1
Falta de interdisciplinaridade	TR5	1
Modalidade (integral, diurno, noturno)	TR7	1
Método de Ensino (Presencial / EAD)	TR7	1
Trabalho Teórico (em horas)	TR8	1
Trabalho Prático (em horas)	TR8	1
Universidade (nome ou ID)	TR9	1
ID / Matrícula	TR9	1
Carga horária obrigatória	TR18	1
Atuação da Coordenação do curso	TR18	1
Desempenho Acadêmico		
Atributo	Trabalhos	Total
Situação de Evasão	TR6, TR7, TR8, TR9, TR11, TR12, TR13, TR15, TR17	9
Ano de Admissão	TR4, TR9, TR14, TR15, TR17	5
Quantidade de Reprovações	TR4, TR8, TR13, TR14, TR18	5
Quantidade de Aprovações	TR4, TR8	2
Período de Admissão	TR11, TR14	2
Problemas Vocacionais / Falta de afinidade com o Curso	TR3, TR18	2
Integração social / Engajamento	TR5, TR18	2
Período do curso	TR7, TR8	2
Frequência	TR13, TR14	2
Período de Saída	TR11	1
Facilidade com Programação	TR3	1
Média nos três primeiros períodos	TR6	1
Primeiros Meses	TR8	1
Disciplina (Concluída / Trancada / Cancelada)	TR14	1
Média da nota dos Aprovados	TR14	1
Média da nota dos Reprovados	TR14	1

Agradecimentos

Agradecemos à CAPES pela concessão da bolsa de estudos e ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Viçosa (PPGCC/UFV) pelo apoio institucional.

Referências

- Carvalho, C. S., Mattos, J. C., and Aguiar, M. S. (2023). Avaliação da interpretabilidade de modelos por meio da clusterização de explicações no contexto da predição de evasão no ensino superior. In *Simpósio Brasileiro de Informática na Educação (SBIE)*, pages 1191–1201. SBC.
- Carvalho, L., Santos, A., Nakamura, F., and Oliveira, E. (2019). Detecção precoce de evasão em cursos de graduação presencial em computação: um estudo preliminar. In *Workshop Sobre Educação Em Computação (WEI)*, pages 233–243. SBC.
- Colpo, M. P., Primo, T. T., and de Aguiar, M. S. (2021). Predição da evasão estudantil: uma análise comparativa de diferentes representações de treino na aprendizagem de modelos genéricos. In *Simpósio Brasileiro de Informática na Educação (SBIE)*, pages 873–884. SBC.
- Colpo, M. P., Primo, T. T., Pernas, A. M., and Cechinel, C. (2020). Mineração de dados educacionais na previsão de evasão: uma rsl sob a perspectiva do congresso brasileiro de informática na educação. In *Simpósio Brasileiro de Informática na Educação (SBIE)*, pages 1102–1111. SBC.
- Costa, A. G., Mattos, J. C., Primo, T. T., Cechinel, C., and Muñoz, R. (2021). Model for prediction of student dropout in a computer science course. In *2021 XVI Latin American Conference on Learning Technologies (LACLO)*, pages 137–143. IEEE.
- Da Cruz, R. C., Juliano, R. C., Monteiro Souza, F. C., and Correa Souza, A. C. (2023). A score approach to identify the risk of students dropout: an experiment with information systems course. In *Proceedings of the XIX Brazilian Symposium on Information Systems*, pages 120–127.
- Da Silva, P. M., Lima, M. N., Soares, W. L., Silva, I. R., Fagundes, R. A. d. A., and De Souza, F. F. (2019). Ensemble regression models applied to dropout in higher education. In *2019 8th Brazilian Conference on Intelligent Systems (BRACIS)*, pages 120–125. IEEE.
- da Silva, V. R. L. B., de Albuquerque Silva, F., and Burégio, V. (2019). Characterizing educational data mining. In *2019 14th Iberian Conference on Information Systems and Technologies (CISTI)*, pages 1–5. IEEE.
- de Brito, B. C. P., de Mello, R. F. L., and Alves, G. (2020). Identificação de atributos relevantes na evasão no ensino superior público brasileiro. In *Simpósio Brasileiro de Informática na Educação (SBIE)*, pages 1032–1041. SBC.
- de Oliveira, M. M. and Barbosa, E. F. (2023). Multilevel modeling for the analysis and prediction of school dropout: a systematic review. In *2023 IEEE 47th Annual Computers, Software, and Applications Conference (COMPSAC)*, pages 103–112. IEEE.

- do Carmo, É. P., Gasparini, I., and de Oliveira, E. H. T. (2022). Identificação de trajetórias de aprendizagem em um curso de graduação e sua relação com a evasão escolar. In *Simpósio Brasileiro de Informática na Educação (SBIE)*, pages 323–333. SBC.
- dos Santos, J. F., de Sousa, J. D. A., Mello, R. F., Cristina, C. T., and Alves, G. (2021a). Um modelo para análise do impacto da retenção e evasão no ensino superior utilizando cadeias de markov absorventes. In *Simpósio Brasileiro de Informática na Educação (SBIE)*, pages 813–823. SBC.
- dos Santos, V. H. B., Saraiva, D. V., and de Oliveira, C. T. (2021b). Uma análise de trabalhos de mineração de dados educacionais no contexto da evasão escolar. In *Simpósio Brasileiro de Informática na Educação (SBIE)*, pages 1196–1210. SBC.
- Duran, R., Bim, S. A., Gimenes, I., Ribeiro, L., and Correia, R. C. M. (2023). Potential factors for retention and intent to drop-out in brazilian computing programs. *ACM Transactions on Computing Education*, 23(3):1–33.
- Falcao, A. P., Villwock, R., and Miloca, S. A. (2023). Análise de dados pré-universidade para predizer a evasão de alunos ingressantes em uma instituição de ensino superior. In *Simpósio Brasileiro de Informática na Educação (SBIE)*, pages 1293–1304. SBC.
- Fukao, A. T., Colanzi, T. E., Martimiano, L. A., and Feltrim, V. D. (2023). Estudo sobre evasão nos cursos de computação da universidade estadual de maringá. In *Simpósio Brasileiro de Educação em Computação (EduComp)*, pages 86–96. SBC.
- Hoed, R. M., Ladeira, M., and Leite, L. L. (2018). Influence of algorithmic abstraction and mathematical knowledge on rates of dropout from computing degree courses. *Journal of the Brazilian Computer Society*, 24(1):1–16.
- Kantorski, G. Z., Martins, R. Z., Balejo, A., and Frick, M. (2023). Mineração de dados educacionais para predição da evasão em cursos de graduação presenciais no ensino superior. In *Simpósio Brasileiro de Informática na Educação (SBIE)*, pages 1133–1142. SBC.
- Llauró, A., Fonseca, D., Villegas, E., Aláez, M., and Romero, S. (2021). Educational data mining application for improving the academic tutorial sessions, and the reduction of early dropout in undergraduate students. In *Ninth International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM'21)*, pages 212–218.
- Marques, L. T., Marques, B. T., Rocha, R. S., Chaves, L., de Castro, A. F., Queiroz, P. G. G., et al. (2020). Evasão acadêmica e suas causas em cursos de bacharelado em ciência da computação: Um estudo de caso na ufersa. In *Simpósio Brasileiro de Informática na Educação (SBIE)*, pages 1042–1051. SBC.
- Menolli, A., Horita, F., Dias, J. J. L., and Coelho, R. (2020). Bi-based methodology for analyzing higher education: A case study of dropout phenomenon in information systems courses. In *XVI Brazilian Symposium on Information Systems*, pages 1–8.
- Nájera, A. B. U. and Ortega, L. A. M. (2022). Predictive model for taking decision to prevent university dropout. *IJIMAI*, 7(4):205–213.
- Oliveira, J. L., Ambrósio, A. P., Silva, U., Brancher, J., and Franco, J. J. (2020). Undergraduate students' effectiveness in an institution with high dropout index. In *2020 IEEE Frontiers in Education Conference (FIE)*, pages 1–7. IEEE.

- Opazo, D., Moreno, S., Álvarez-Miranda, E., and Pereira, J. (2021). Analysis of first-year university student dropout through machine learning models: A comparison between universities. *Mathematics*, 9(20):2599.
- Pachas, D. A. G., Garcia-Zanabria, G., Cuadros-Vargas, A. J., Camara-Chavez, G., Poco, J., and Gomez-Nieto, E. (2021). A comparative study of who and when prediction approaches for early identification of university students at dropout risk. In *2021 XLVII Latin American Computing Conference (CLEI)*, pages 1–10. IEEE.
- Rocha, R., Vanderlei, I., Araujo, J., et al. (2021). A model of actions to reduce evasion in higher technology courses: An experience report. In *2021 16th Iberian Conference on Information Systems and Technologies (CISTI)*, pages 1–6. IEEE.
- Santos, C. H. D., de Lima Martins, S., and Plastino, A. (2021). É possível prever evasão com base apenas no desempenho acadêmico? In *Simpósio Brasileiro de Informática na Educação (SBIE)*, pages 792–802. SBC.
- Santos, G., Belloze, K. T., Tarrataca, L., Haddad, D. B., Bordignon, A. L., and Brandão, D. N. (2020). Evolvedtree: Analyzing student dropout in universities. In *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, pages 173–178. IEEE.
- Santos, K. J. d. O., Menezes, A. G., de Carvalho, A. B., and Montesco, C. A. (2019). Supervised learning in the context of educational data mining to avoid university students dropout. In *2019 IEEE 19th International Conference on Advanced Learning Technologies (ICALT)*, volume 2161, pages 207–208. IEEE.
- Silva, E. C., Freitas, S., Soares Ramos, C., Muniz De Menezes, A. E., and Rodrigues De Araujo, L. K. S. (2023). A systematic review of the factors that impact the prediction of retention and dropout in higher education. In *2023 Hawaii International Conference on System Sciences (HICSS-56)*, pages 1271–1280.
- Silva, R. A. d. S., AF, B. B., Maria de Fátima, P. F., de Sousa Santos, I., and Andrade, R. M. (2021). Evasão em computação na ufc sob a perspectiva dos alunos. In *Workshop sobre Educação em Computação (WEI)*, pages 338–347. SBC.
- Vasconcelos, V. and Andrade, E. (2018). Análise da evasão de alunos na licenciatura em computação. In *Workshop sobre Educação em Computação (WEI)*. SBC.
- Viana, F. S., Santana, A. M., and Rabêlo, R. d. A. L. (2022). Avaliação de classificadores para predição de evasão no ensino superior utilizando janela semestral. In *Simpósio Brasileiro de Informática na Educação (SBIE)*, pages 908–919. SBC.