

Mineração de dados auxiliando na descoberta das causas da evasão escolar: Um Mapeamento Sistemático da Literatura

Leonardo Torres Marques, UFERSA/UERN, leonardo.torresmarques@gmail.com

Angélica Félix De Castro, UFERSA, ang_lica@ufersa.edu.br

Bruno Torres Marques, UFC, brunotorgesepj@gmail.com

Jesaias Carvalho Pereira Silva, UFERSA/UERN, jesayassilva@gmail.com

Paulo Gabriel Gadelha Queiroz, UFERSA, paulpggq@gmail.com

Resumo: Este trabalho apresenta um Mapeamento Sistemático da Literatura sobre evasão escolar, em que se buscou identificar tecnologias de mineração de dados e fatores indutores para evasão escolar, que vem sendo exploradas para desvendar as possíveis causas da evasão escolar. As buscas foram realizadas em quatro bases de dados científicas, com o objetivo de responder a seguinte questão de pesquisa: “Quais ferramentas, técnicas e fatores indutores vem sendo utilizados para desvendar possíveis causas da evasão escolar?”. Observou-se que a ferramenta *Weka* é a mais utilizada para auxiliar a desvendar as causas da evasão escolar. Entre as técnicas, destaca-se a utilização da classificação. Por fim, o mapeamento identificou que os principais trabalhos da área se concentram em estudar fatores relacionados às características individuais do aluno.

Palavras-Chave: mapeamento sistemático da literatura, evasão escolar, mineração de dados, fatores indutores.

Data Mining Helping Discover Causes of Dropout: A Systematic Mapping of Literature

Abstract. This work presents a Systematic Literature Mapping about school dropout, in which it was tried to identify data mining technologies and inducing factors for school evasion, that has been explored to uncover the possible causes of school dropout. Searches were performed in four scientific databases, with the purpose of answering the following research question: "What tools, techniques and inducing factors have been used to unravel possible causes of school dropout?". It was observed that the Weka tool is the most used to help unravel the causes of dropout. Among the techniques, the use of classification stands out. Finally, the mapping identified that the main works in the area focus on studying factors related to individual student characteristics.

Keywords: systematic literature mapping, school dropout, data mining, inductive factors.

1. Introdução

Nos últimos anos tem crescido o interesse e a preocupação em muitos países, sobretudo no Brasil, com o problema da evasão escolar e a determinação de seus principais fatores indutores relacionados. A evasão escolar é um caso social complexo, definido como o encerramento do ciclo de estudos (Gaioso, 2005). Esse problema aflige as instituições de ensino em geral, sejam públicas ou privadas. Observa-se que a saída dos alunos ocasiona externalidades negativas em âmbitos sociais, acadêmicos, econômicos e ambientais, repercutindo nacionalmente de maneira negativa nas políticas públicas de desenvolvimento educacional, pela falta de retorno dos recursos investidos.

Um levantamento realizado pelo Inep (2015) revela um acréscimo desordenado na taxa de evasão escolar no Brasil. Os dados são referentes a uma avaliação da trajetória dos alunos entre 2010 e 2014. Em 2010, 11,4% dos alunos abandonaram o curso para o qual

foram admitidos. Em 2014, esse número chegou a 49%. A pesquisa apresenta um panorama preocupante, pois mostra uma grande ociosidade no sistema. Esse levantamento aponta que das 6,1 milhões de novas vagas em instituições públicas e privadas de ensino superior, somente 42,1% estão preenchidas e 13,5% das vagas remanescentes foram ocupadas.

O estudo das causas da evasão escolar e a tomada de medidas preventivas estão fortemente ligados ao contexto de cada instituição de ensino. A identificação dos fatores que influenciam a evasão escolar e atribuição de uma ordem de importância para estes fatores é um trabalho complexo, que está diretamente ligado à análise do conjunto de alunos (Manhães et al., 2011). Vale destacar que cada instituição deve identificar as causas do fenômeno em seu ambiente educacional (Nagai; Cardoso, 2017; Paz; Cazella, 2018).

Uma solução propícia para o estudo das causas da evasão escolar é o uso da descoberta de conhecimento, por meio de técnicas de bancos de dados ou Mineração de Dados (MD), denominado de *Educational Data Mining* (EDM) (Júnior et al. 2019). Essa área de pesquisa é responsável pelo desenvolvimento de métodos para entender melhor os alunos e o contexto em que eles aprendem (Romero; Ventura, 2010). É possível aplicar técnicas de EDM para criar modelos que preveem especificamente o abandono escolar e a falha do aluno.

Sendo assim se faz necessário desvendar as principais ferramentas, técnicas e algoritmos para utilização adequada da EDM nesse contexto, evasão escolar. Uma possibilidade é a utilização do mapeamento sistemático. De acordo com Petersen et al. (2008), um Mapeamento Sistemático de Literatura (*Systematic Literature Mapping* – SLM, em inglês) é uma pesquisa episódica que usa uma metodologia adequada e confiável para identificar, avaliar e interpretar toda a pesquisa disponível relevante para uma questão de pesquisa em particular, ou área de tópico, ou fenômeno de interesse de uma maneira repetível e imparcial.

Diante desse contexto, os objetivos deste Mapeamento Sistemático de Literatura são: (i) conhecer as melhores ferramentas de MD utilizadas para a descoberta das causas da evasão escolar; (ii) identificar quais técnicas de MD têm sido utilizadas para a descoberta das causas da evasão escolar; (iii) desvendar quais os tipos de fatores indutores têm sido investigados para a descoberta das causas da evasão escolar.

O restante deste trabalho está organizado da seguinte forma: na Seção 2, apresentam-se os Procedimentos e Métodos; na Seção 3, detalha-se a Condução do Mapeamento Sistemático da Literatura; na Seção 4, são analisados e discutidos os Resultados do Mapeamento Sistemático da Literatura e Discussões; e, na Seção 5, são discutidas as Conclusões e Considerações Finais deste trabalho.

2. Procedimentos e Métodos

Nesta seção são apresentados os detalhes do processo utilizado neste SLM. Salienta-se, que este SLM se baseia no protocolo definido por Petersen et al. (2008), em que são aludidas as questões de pesquisa, são definidas as palavras-chave e *string* de busca, delimita-se o intervalo de tempo para pesquisa, são definidas as bases de dados de pesquisa, os critérios de inclusão e exclusão e, os procedimentos para seleção dos estudos.

2.1 Questões de Pesquisa

De acordo com os objetivos apresentados na Seção 1, foram definidas as seguintes questões de pesquisa que norteiam este trabalho:

- **Questão primária (QP1):** Quais das melhores ferramentas de MD utilizadas para a descoberta das causas da evasão escolar?
- **Questão primária (QP2):** Quais das técnicas de MD têm sido utilizadas para a descoberta das causas da evasão escolar?
- **Questão primária (QP3):** Que tipos de fatores indutores têm sido investigados para a descoberta das causas da evasão escolar?

2.2 Definição das Palavras-chave e *String* de Busca

A determinação das palavras-chave em SLMs é estabelecida a partir das questões de pesquisa sob investigação. Além das palavras-chave, para a definição de uma *string* de busca abrangente, é importante encontrar sinônimos para cada das palavras-chave, conforme apresentado na Tabela 1.

Tabela 1 – Palavras-chave e sinônimos.

Palavra-Chave	Sinônimos em Inglês
<i>Data Mining</i>	<i>Knowledge Discovery, Data Extraction, Predicting, Prediction, Detection e Detecting</i>
<i>School Dropout</i>	<i>School Drop-Out, School Evasion, School Retention, School Truancy and School Failure</i>

Após se efetuar as combinações das palavras-chave e sinônimos, foi delineada a seguinte *string* de busca:

(“Data Mining” OR “Knowledge discovery” OR “data extraction” OR predicting OR prediction OR detection OR detecting) AND (“school dropout” OR “school drop-out” OR “school evasion” OR “school retention” OR “school truancy” OR “school failure”)

Observa-se que a *string* de busca foi validada por um especialista e por meio de uma busca piloto realizada em uma das bases de dados, conforme se apresenta na Subseção 2.4.

2.3 Intervalo de Tempo para Pesquisa

Desde o início da década de 1970, já existia preocupação com o problema da evasão escolar, diante da complexidade e a necessidade de encontrar soluções plausíveis. Neste período, Tinto (1975) apresentou o primeiro modelo de fatores contribuintes para o processo de evasão escolar. Entretanto, considerou-se estudos realizados no período dos últimos 10 anos, compreendido entre 2008 e (dezembro de) 2018, por se considerar que, as constantes evoluções, no que se trata de ferramentas, técnicas, algoritmos e processos, tem aperfeiçoado a busca das possíveis causas da evasão escolar.

2.4 Bases de Dados de Pesquisa

Para o SLM em questão, as bases escolhidas foram as seguintes: *IEEEExplore*, *ACM Digital Library*, *Science Direct* e *Scopus*. O critério de escolha dessas bases de dados se deve ao fato de serem as bases mais difundidas no campo da tecnologia. Essas bases de dados foram desenvolvidas com o intuito de propiciar, em uma única plataforma, centenas

V. 17 Nº 3, dezembro, 2019 _____ RENEOTE
DOI:

de revistas científicas e seus respectivos trabalhos, assentes por qualidade e originalidade (Dantas et al., 2018).

2.5 Critérios de Inclusão e Exclusão

Os critérios de inclusão e exclusão são definidos para auxiliar na condução de um SLM, com o intuito de apoiar a classificação de relevância dos estudos (Fuzeto e Braga, 2016). Os critérios de inclusão estão intimamente relacionados às questões de pesquisa em análise. Vale ressaltar que para um trabalho ser selecionado é suficiente atender a pelo menos um critério de inclusão. De maneira análoga, para excluir um trabalho é suficiente atender a um critério de exclusão. Dessa forma, foram definidos como critérios de exclusão (CE):

- **CE1:** estudos em andamento, que não tenham sido concluídos, ou que o estudo final não esteja disponível;
- **CE2:** todos os estudos que não tratem da utilização de MD na descoberta das causas da evasão escolar, serão excluídos;
- **CE3:** serão excluídos artigos em outros idiomas que não o inglês, devido à baixa ascendência de estudos sobre o tema em outros idiomas;
- **CE4:** caso existam estudos duplicados, apenas um deles será aceito (o mais recente/desenvolvido), excluindo-se todos os demais.

2.6 Procedimentos para Seleção dos Estudos

Foram seguidas 3 (três) fases de filtragem de trabalhos, conforme se apresenta a seguir.

- **Fase 1:** exclusão de estudos em andamento, que não tenham sido concluídos, ou que o estudo final não esteja disponível;
- **Fase 2:** filtragem de trabalhos por meio da análise de títulos e resumos das publicações coletadas e eleger aquelas que satisfazem aos critérios de inclusão e excluir os que atendem aos critérios de exclusão;
- **Fase 3:** filtragem de trabalhos por meio da leitura completa dos trabalhos selecionados na etapa anterior.

Observa-se que todas as etapas foram validadas e discutidas com dois (2) pesquisadores experientes que guiaram o processo do mapeamento.

3. Condução do SLM

Nesta seção, detalha-se o procedimento de busca e seleção de trabalhos. As buscas nas bases de dados foram realizadas por meio de *Advanced Search*, sendo adaptadas as particularidades de cada base.

Tabela 2 – Resultados das buscas.

Base	Fase 1	Fase 2	Fase 3
<i>IEEEExplore</i>	8	2	2
<i>ACM Digital Library</i>	4	1	1
<i>Science Direct</i>	38	4	3
<i>Scopus</i>	100	9	8
Total	150	16	14

Pode-se observar na Tabela 2 os quantitativos dos trabalhos analisados em cada fase. Na Fase 1 foram encontrados cento e cinquenta (150) trabalhos, entre os quais cinco (5) eram trabalhos repetidos. Na Fase 2, foram selecionados dezesseis (16) trabalhos após a leitura dos títulos e resumos. Na fase 3, após a leitura completa dos trabalhos e aplicando-se os critérios (inclusão/exclusão), foram selecionados quatorze (14) trabalhos.

Na Tabela 3 são apresentados os trabalhos selecionados, com os respectivos anos de publicação e o continente em que a pesquisa foi explorada. Constata-se que apesar das oscilações nos últimos dez (10) anos, tem se mantido o número de publicações e práticas referentes a busca pelas possíveis causas da evasão escolar e por soluções para o problema.

Tabela 3 – Relação dos trabalhos selecionados.

Índice	Trabalhos	Ano Publicação	Continentes
#1	Archambault et al. (2009)	2009	América do Norte
#2	Bowers (2010)	2010	América do Norte
#3	Fall and Roberts (2012)	2012	América do Norte
#4	Sabates et al. (2013)	2013	Ásia
#5	Márquez-Vera et al. (2013)	2013	América do Norte
#6	Martinho et al. (2013)	2013	América do Sul
#7	Eicher et al. (2014)	2014	Europa
#8	Jiménez-Gómez et al. (2015)	2015	Europa
#9	Parr and Bonitz (2015)	2015	América do Norte
#10	Márquez-Vera et al. (2016)	2016	América do Norte
#11	Weybright et al. (2017)	2017	África
#12	Rovira et al. (2017)	2017	Europa
#13	Wood et al. (2017)	2017	América do Norte
#14	Adelman et al. (2018)	2018	América Central

Em relação a distribuição por continentes, é possível observar que o interesse e pesquisas em relação a evasão escolar, tem sido grande em diversas partes do planeta. Destaca-se o quantitativo na América do Norte, de onde foram retornados metade dos trabalhos. Vale salientar, que os Estados Unidos tem sido o país que mais investiga sobre o problema.

4. Resultados e Discussões

Nesta seção são apresentadas as respostas para as questões de pesquisa, a partir dos trabalhos que foram selecionados, com potencial de responder aos questionamentos que nortearam esta pesquisa. Na Figura 1, apresenta-se a distribuição percentual em que ferramentas e técnicas de MD foram encontradas nos trabalhos. Observa-se que a análise desses dados é apresentada a seguir.

4.1 Quais das melhores ferramentas de MD utilizadas para a descoberta das causas da evasão escolar?

Na Figura 1(a) é apresentado o quantitativo de uso das ferramentas de MD nos trabalhos selecionados. É possível verificar que as ferramentas *Weka* e *Mplus* destacam-se, pois foram utilizadas em três (3) trabalhos. As demais ferramentas, *Kaplan-Meier*, *HLM*, *Python*, *SPSS* e *Software R* foram utilizadas em apenas 1 (um) trabalho.

A *Weka* foi utilizada nos trabalhos de Márquez-Vera et al. (2016), Márquez-Vera et al. (2013) e Jiménez-Gómez et al. (2015). A ferramenta é uma coleção de algoritmos de

aprendizado de máquina para tarefas de MD, oferece aparatos para preparação de dados, classificação, regressão, agrupamento, mineração de regras de associação e visualização.

O *Mplus* foi utilizado nos trabalhos de Parr e Bonitz (2015), Archambault et al. (2009) e Fall e Roberts (2012). A ferramenta é um programa de modelagem estatística para análises de dados, que oferece uma ampla variedade de modelos, estimadores, algoritmos e exibições gráficas de dados e resultados de análises. Essa ferramenta também permite análise de dados transversais e longitudinais, dados de nível único e multinível, dados provenientes de diferentes populações com heterogeneidade observada ou não observada e dados que contêm valores omissos.

O *SAS-STAT* empregado no trabalho de Weybright et al. (2017), permite realizar análises estatísticas, incluindo análise de variância, regressão, análise categórica de dados, análise multivariada, análise de sobrevivência, análise psicométrica, análise de *cluster*, análise não paramétrica, análise de modelos mistos e análise de dados de pesquisa com numerosos exemplos, além de informações sobre sintaxe e uso.

O *HLM* utilizado no trabalho de Wood et al. (2017), é um *software* estatístico para modelagem linear hierárquica. O *Python* foi utilizado no trabalho de Rovira et al. (2017) com as bibliotecas *pandas* e *scikit-learn*. Observa-se que o *Python* é uma linguagem de programação de uso geral que pode ser usada para diferentes fins: coleta de dados, engenharia de dados, análise, *Web Scraping*, construção de aplicativos *web* e etc. O *Software R* utilizado no trabalho de Eicher et al. (2014). Salienta-se que ele utilizado para dar suporte em análise de dados e estatística. O *SPSS* manuseado no trabalho de Bowers (2010), é um pacote de *software* usado para análise estatística interativa ou em lote.

Tabela 4 – Avaliação das ferramentas.

Ferramentas	Licença	Documentação	Open Source
<i>Weka</i>	Livre	Sim	Sim
<i>Mplus</i>	Paga	Sim	Não
<i>SAS-STAT</i>	Paga	Sim	Não
<i>HLM</i>	Paga	Sim	Não
<i>Python</i>	Livre	Sim	Não
<i>Software R</i>	Livre	Sim	Sim
<i>SPSS</i>	Paga	Sim	Não

Foram definidos três (3) critérios para definir avaliar as ferramentas, os critérios foram: se a licença para utilização da ferramenta é paga ou livre, se oferece documentação e se é *open source*. Na Tabela 4 é possível observar a avaliação para cada ferramenta. Constatou-se que a *weka* e o *software R*, foram as ferramentas que obtiveram melhores avaliações, pois, ambas têm licença livre, oferecem documentação e são *open sources*.

4.2 Quais das técnicas de MD têm sido utilizadas para a descoberta das causas da evasão escolar?

Na Figura 1(b) são exibidas as técnicas de MD que foram exploradas nos trabalhos encontrados. Destacam-se as técnicas de Classificação que foram utilizadas em quatro (4) trabalhos, em seguida as Equações Estruturais empregadas em três (3) trabalhos.

A técnica de Classificação foi explorada nos trabalhos de Rovira et al. (2017), Márquez-Vera et al. (2016), Jiménez-Gómez et al. (2015) e Sabates et al. (2013). Essa técnica é uma tarefa da MD que associa ou classifica objetos a determinadas classes, ela busca prever uma classe de um novo dado automaticamente. As Equações Estruturais utilizadas nos trabalhos de Parr e Bonitz (2015), Archambault et al. (2009) e Fall e

Roberts (2012), são utilizadas para analisar múltiplas variáveis, que normalmente estão associadas a um determinado indivíduo, por exemplo.

A técnica de Regressão abordada nos trabalhos de Adelman et al. (2018) e Wood et al. (2017), é utilizada para se prever um valor real, por exemplo, em que período do curso um aluno irá evadir. A técnica de Análise de Sobrevida foi explorada nos trabalhos de Weybright et al. (2017) e Bowers (2010), e é utilizada para se estabelecer o tempo até a ocorrência de determinado evento.

A Análise Fatorial, empregada nos trabalhos de Archambault et al. (2009) e Fall e Roberts (2012), é uma técnica estatística exploratória, que tem como objetivo reduzir grandes conjuntos variáveis em grupos menores e altamente correlacionados. A técnica de Rede Neural foi utilizada no trabalho de Martinho et al. (2013), essa técnica é inspirada no sistema nervoso central de um animal, para realização de aprendizado de máquina e reconhecimento de padrões. A Análise Multinível utilizada no trabalho de Eicher et al. (2014), é uma técnica que permite a regressão diferenciada de dados longitudinais, ou seja, coletados ao longo do tempo.

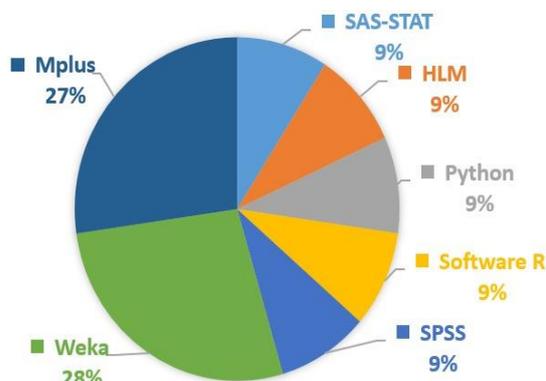


Figura 1(a) - Ferramentas de MD.

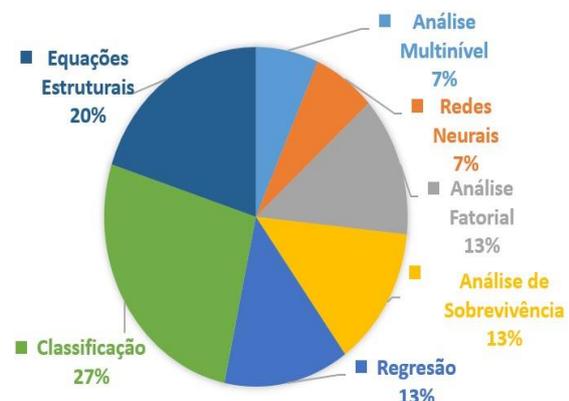


Figura 1(b) - Técnicas de MD.

Figura 1 - Ferramentas e técnicas de MD.

4.3 Que tipos de fatores indutores têm sido investigados para a descoberta das causas da evasão escolar?

Na Figura 2, são exibidos os fatores envolvidos no risco de evasão escolar, com os índices dos respectivos trabalhos retornados no SLM. A relação de índices e trabalhos é apresentada na Tabela 2. Salienta-se que em todos os trabalhos, foram explorados dados inerentes às características individuais do aluno. Destaca-se que apenas nos trabalhos (#3, #11 e #14) foram explorados dados inerentes às características individuais do aluno, fatores internos às instituições de ensino e externos às instituições de ensino.

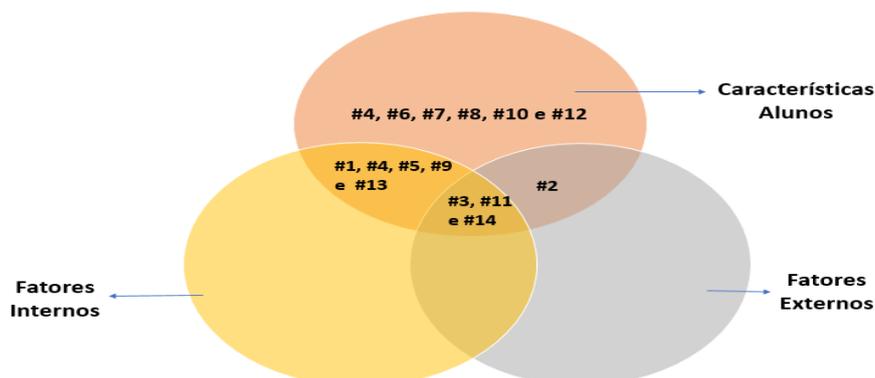


Figura 2 - Fatores indutores para à evasão escolar.

Os fatores relacionados às características individuais do aluno são pertinentes a desinteresse e insatisfação com a escola ou universidade, descoberta de novos interesses, fraco desempenho, dificuldades de aprendizagem, constantes reprovações, baixa frequência nas aulas, falta de habilidade de estudo, entre outros (Tinto, 1975).

Os fatores associados às instituições são relativos a currículos ultrapassados e distendidos, estrutura débeis de apoio ao ensino (laboratórios; equipamentos de informática; bibliotecas; etc), didático-pedagógicas (métodos inadequados da avaliação do desempenho do aluno, deficiência de formação pedagógica ou desinteresse do professor, desprovimento ou número reduzido de programas institucionais para o estudante), entre outros (Tinto, 1975).

Os fatores externos às instituições são fatores concernentes ao meio em que o aluno está inserido. Exemplos desse tipo de fatores são: falta de reconhecimento social com educação, falta de perspectivas empregatícias, falta de políticas públicas plausíveis e continuadas voltadas para educação, entre outras (Tinto, 1975).

4 Conclusões e Considerações Finais

Neste trabalho foi apresentado o planejamento, a condução e os resultados de um SLM sobre o problema de evasão escolar, em que se buscou elencar ferramentas, técnicas e fatores indutores utilizados para prever as causas do problema nos últimos 10 (dez) anos. A busca pelos trabalhos resultou na pré-seleção de dezesseis (16) trabalhos, dentre os quais quatorze (14) foram incluídos para a extração de dados.

A partir dos resultados, evidenciou-se que a ferramenta *Weka* se consolida como uma das ferramentas de MD mais utilizadas para auxiliar a desvendar causas da evasão escolar, por meio da aplicação de técnicas e algoritmos sobre conjuntos de dados. Com essa ferramenta é possível acompanhar processos de MD de pequenas bases de dados, realizando a avaliação dos resultados obtidos e a comparação de algoritmos. Ademais, é possível executar tarefas relacionadas ao pré-processamento de dados como, por exemplo, a seleção e a transformação de atributos.

Foi possível verificar que a técnica de classificação, vem sendo bastante utilizada e tem retornado alta precisão nas previsões de tendência de evasão escolar. Vale realçar que em 75% das vezes que essa técnica foi utilizada, utilizou-se a ferramenta *Weka*. Os algoritmos de classificação mais utilizados nos trabalhos foram: *naive bayes* (NB), *support-vector machine* (SMO), *network of radial basis function* (RBFNetwork), *multilayer perceptron* (MLP), *k-nearest neigh-bours* (IBk), *Jrip*, *OneR*, *J48*, *PART* e *AdaBoostM1*.

Em relação aos fatores indutores para a evasão escolar, constatou-se que em todos os trabalhos são investigadas possíveis causas relacionadas às características individuais do aluno, analisando dados armazenados nas instituições e/ou levantando dados por meio de questionários, da perspectiva do aluno. Vale ressaltar, que possíveis fatores relacionados à instituição e externos à instituição, ainda são pouco explorados.

Por fim, é importante enfatizar que qualquer pesquisa desta natureza possui riscos à validade dos seus resultados. Para este trabalho em particular, detectam-se os seguintes pontos de fragilidade: a pré-seleção dos artigos, devido ao conjunto limitado de bases consultadas, ao número de artigos encontrados, e pelo fato de alguns títulos e resumos não terem refletido de forma adequada o conteúdo de seus textos; as palavras-chave de busca que, mesmo se manuseando alguns sinônimos, podem não contemplar todos os existentes para os termos da pesquisa, podendo ser insuficientes para capturar todos os trabalhos relevantes; e, o dilatado tempo necessário para a efetuação de um SLM, podendo conduzir a uma diferente captura de resultados nas bases consultadas, em diferentes datas.

Concluindo-se, esta pesquisa pode ser utilizada para direcionar trabalhos futuros, auxiliando como base para novas revisões e para um trabalho, já em andamento, que utiliza tecnologias de MD para desenvolver uma abordagem de descoberta de conhecimento em bases de dados, que visa desvendar possíveis causas de evasão escolar ou mesmo, identificar alunos com potencial de evasão e tomar medidas para mitigar esse potencial. Além disso, pretende-se dar continuidade a este SLM na busca de novos trabalhos, em novas bases de pesquisa, na tentativa de reduzir as ameaças à sua validade.

Agradecimentos

Este estudo foi financiado em parte pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - código financeiro 001. Os autores agradecem à CAPES e a Pós-Graduação.

Referências Bibliográficas

1. ADELMAN, M., HAIMOVICH, F., HAM, A., and VAZQUEZ, E. (2018). Predicting school dropout with administrative data: new evidence from guatemala and honduras. **Education Economics**, 26(4):356–372.
2. ARCHAMBAULT, I., JANOSZ, M., FALLU, J.-S., and PAGANI, L. S. (2009). Student engagement and its relationship with early high school dropout. **Journal of adolescence**, 32(3):651–670.
3. BOWERS, A. J. (2010). Grades and graduation: A longitudinal risk perspective to identify student dropouts. **The Journal of Educational Research**, 103(3):191–207.
4. Dantas, Á. M. C., Viana, H., Abijaude, J., and Sobreira, P. (2018). Internet das coisas e aprendizagem colaborativa: Revisão sistemática da literatura. In **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**, volume 29, page 278.
5. EICHER, V., STAERKLÉ, C., and CLÉMENCE, A. (2014). I want to quit education: A longitudinal study of stress and optimism as predictors of school dropout intention. **Journal of adolescence**, 37(7):1021–1030.
6. FALL, A.-M. and ROBERTS, G. (2012). High school dropouts: Interactions between social context, self-perceptions, school engagement, and student dropout. **Journal of adolescence**, 35(4):787–798.
7. FUZETO, R. and BRAGA, R. (2016). Um mapeamento sistemático em progresso sobre internet das coisas e educação à distância. In **Anais dos Workshops do Congresso Brasileiro de Informática na Educação**, volume 5, page 1334.
8. GAIOSO, N. d. L. (2005). O fenômeno da evasão escolar na educação superior no brasil. **Brasília, DF: Universidade Católica de Brasília**.
9. INEP, MEC. "Censo da educação superior 2013: resumo técnico.". **INEP**. (2015).
10. JIMÉNEZ-GÓMEZ, M.Á., LUNA, J. M., ROMERO, C., and VENTURA, S. (2015). Discovering clues to avoid middle school failure at early stages. In **Proceedings of the Fifth International Conference on Learning Analytics and Knowledge**, pages 300–304. ACM.
11. JÚNIOR, O. d. G. F., RODRIGUES, W. R. M., BARBIRATO, J. C. C., and de BARROS COSTA, E. (2019). Melhoria da gestão escolar através do uso de técnicas de mineração de dados educacionais: um estudo de caso em escolas municipais de maceió. **RENOTE**, 17(1):296–305.
12. MANHÃES, L. M. B., Da CRUZ, S. M. S., COSTA, R. J. M., ZAVALETA, J., and ZIMBRÃO, G. (2011). Previsão de estudantes com risco de evasão utilizando técnicas de mineração de dados. In **Brazilian Symposium on Computers in**

- Education (Simpósio Brasileiro de Informática na Educação-SBIE)**, volume 1.
13. MÁRQUEZ-VERA, C., CANO, A., ROMERO, C., NOAMAN, A. Y. M., MOUSA FARDOUN, H., and VENTURA, S. (2016). Early dropout prediction using data mining: a case study with high school students. **Expert Systems**, 33(1):107–124.
 14. MÁRQUEZ-VERA, C., MORALES, C. R., and SOTO, S. V. (2013). Predicting school failure and dropout by using data mining techniques. **IEEE Revista Iberoamericana de Tecnologías del Aprendizaje**, 8(1):7–14.
 15. MARTINHO, V. R. D. C., NUNES, C., and MINUSSI, C. R. (2013). An intelligent system for prediction of school dropout risk group in higher education classroom based on artificial neural networks. In **2013 IEEE 25th International Conference on Tools with Artificial Intelligence**, pages 159–166. IEEE.
 16. NAGAI, N. P. and CARDOSO, A. L. J. (2017). A evasão universitária: Uma análise além dos números. **Revista Estudo & Debate**, 24(1).
 17. PARR, A. K. and BONITZ, V. S. (2015). Role of family background, student behaviors, and school-related beliefs in predicting high school dropout. **The Journal of Educational Research**, 108(6):504–514.
 18. PAZ, F. J. and CAZELLA, S. C. (2018). Integrando sistemas de recomendação com mineração de dados educacionais e learning analytics: Uma revisão sistemática da literatura. **RENOTE**, 16(1).
 19. PETERSEN, K., FELDT, R., MUJTABA, S., and MATTSSON, M. (2008). Systematic mapping studies in software engineering. In **Ease**, volume 8, pages 68–77.
 20. ROMERO, C. and VENTURA, S. (2010). Educational data mining: a review of the state of the art. **IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)**, 40(6):601–618.
 21. ROVIRA, S., PUERTAS, E., and IGUAL, L. (2017). Data-driven system to predict academic grades and dropout. **PLoS one**, 12(2): e0171207.
 22. SABATES, R., HOSSAIN, A., and LEWIN, K. M. (2013). School drop out in bangladesh: Insights using panel data. **International Journal of Educational Development**, 33(3):225–232.
 23. TINTO, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. **Review of educational research**, 45(1):89–125.
 24. WEYBRIGHT, E. H., CALDWELL, L. L., XIE, H., WEGNER, L., and SMITH, E. A. (2017). Predicting secondary school dropout among south african adolescents: A survival analysis approach. **South African journal of education**, 37(2).
 25. WOOD, L., KIPERMAN, S., ESCH, R. C., LEROUX, A. J., and TRUSCOTT, S. D. (2017). Predicting dropout using student-and school-level factors: An ecological perspective. **School Psychology Quarterly**, 32(1):35.