

Mineração de Dados Educacionais na Previsão de Evasão: uma RSL sob a Perspectiva do Congresso Brasileiro de Informática na Educação

Miriam Pizzato Colpo^{1,2}, Tiago Thompsen Primo¹, Ana Marilza Pernas¹,
Cristian Cechinel^{1,3}

¹Programa de Pós-Graduação em Computação
Universidade Federal de Pelotas (UFPEL) – Pelotas, RS, Brasil

²Instituto Federal de Educação, Ciência e Tecnologia Farroupilha (IFFar)
Santa Maria, RS, Brasil

³Centro de Ciências, Tecnologias e Saúde (CTS)
Universidade Federal de Santa Catarina (UFSC) – Araranguá, SC, Brasil

{miriam.colpo, tiago.primo, marilza}@inf.ufpel.edu.br,
contato@crisiancechinel.pro.br

Abstract. *This work presents a systematic review about studies that use educational data mining techniques in the context of student dropout prediction and that were published in the Brazilian Congress on Computers in Education, which is the main academic event of the topic on the country. Contextual (educational systems, modalities and levels), techniques (tasks, categories of algorithms and tools) and data (types, coverage and volume) characteristics were extracted and analyzed in order to identify how this area is being explored in the Brazilian research scenario. As a result, lacks in the exploration of the basic and technical education levels and in the availability of mechanisms to support academic managers in decision-making were evidenced.*

Resumo. *Este trabalho apresenta uma revisão sistemática de estudos que utilizam técnicas de mineração de dados educacionais no contexto da previsão de evasão estudantil e que foram publicados no Congresso Brasileiro de Informática na Educação, principal evento da área no país. Características contextuais (redes, modalidades e níveis de ensino), técnicas (tarefas, categorias de algoritmos e ferramentas) e de dados (tipos, abrangência e volume) foram analisadas com o intuito de identificar como esse tema está sendo abordado no cenário de pesquisa brasileiro. Como resultado, foram evidenciadas carências na exploração dos níveis da educação básica e técnica e na disponibilização, aos gestores acadêmicos, de mecanismos de suporte à tomada de decisões.*

1. Introdução

A evasão, definida pelo Ministério da Educação (MEC) como “a saída definitiva do estudante do curso de origem, sem concluí-lo” [Brasil 1996, p.15], é um problema a ser combatido pelo sistema educacional brasileiro. Os últimos indicadores de fluxo da educação superior, por exemplo, disponibilizados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), indicam que a média das taxas

de desistência acumuladas em 2016 ultrapassou 50%, considerando alunos ingressantes em 2010 [INEP 2020]. Além de prejuízos acadêmicos e sociais, a evasão reflete negativamente no desenvolvimento econômico do país, dada a relação entre o nível de escolarização e os ganhos salariais da população [Pontili et al. 2018]. Para a rede privada de ensino, evasão significa perda de receitas, enquanto, para a rede pública, são recursos do Estado investidos sem o devido retorno. Em ambas, causa subaproveitamento de recursos humanos e de infraestrutura [Silva Filho et al. 2007].

Conhecer o perfil de estudantes evadidos pode auxiliar instituições de ensino na identificação de alunos em risco e, conseqüentemente, no desenvolvimento de estratégias e políticas institucionais de caráter preventivo [Oliveira Júnior et al. 2017]. Com essa intenção, muitos estudos analisam, de forma manual, dados desses alunos, geralmente obtidos a partir da aplicação de entrevistas e questionários. Porém, além da impossibilidade de generalização, os resultados dessas pesquisas são suscetíveis a viesamentos, considerando a dificuldade de contato com parcela significativa dessa população [Silva 2013]. Adicionalmente, embora armazenem informações de alto potencial estratégico, sistemas de gestão acadêmica e ambientes virtuais de aprendizagem produzem grandes volumes de dados educacionais, que inviabilizam explorações manuais [Romero and Ventura 2020].

Conseqüentemente, com o intuito de automatizar a análise da evasão, diversas pesquisas têm utilizado técnicas de Mineração de Dados Educacionais (MDE), área que visa o desenvolvimento de métodos capazes de explorar grandes volumes de dados educacionais e, assim, compreender de forma mais eficaz o comportamento de alunos e outros fatores relacionados à aprendizagem [Baker et al. 2011]. A classificação, que, de acordo com [Han et al. 2012], tem o objetivo de encontrar um modelo que descreva e distinga classes de dados ou conceitos, é a tarefa de MDE que costuma ser empregada no contexto da previsão de evasão, seja para a descoberta automática de padrões e atributos relacionados a esse fenômeno, o que permite o melhor entendimento de suas causas e, dessa forma, pode auxiliar profissionais a prever, mediante acompanhamento pedagógico (previsão "manual"), alunos em risco; seja para a identificação automática de alunos que apresentam padrão de evasão (previsão automática).

[Maschio et al. 2018] apresentam um mapeamento sistemático sobre o uso de MDE no Brasil, considerando suas diversas aplicações. Com escopos mais específicos e próximos deste trabalho, [Marques et al. 2019] apresentam um mapeamento acerca das tecnologias de mineração que têm sido exploradas na identificação das causas da evasão escolar e [Mduma et al. 2019] realizam uma pesquisa sobre as técnicas de aprendizagem de máquina que são utilizadas na previsão de evasão. Diferentemente de [Marques et al. 2019], neste trabalho são analisadas pesquisas de MDE no contexto da previsão de evasão, que é mais amplo e, embora inclua, não se restringe à identificação das causas de evasão, por meio de uma Revisão Sistemática da Literatura (RSL), que não é realizada em [Mduma et al. 2019]. Além disso, diferentemente de ambos, esta revisão foca no cenário de pesquisa brasileiro e agrega aspectos contextuais, técnicos e de dados que careciam de exploração.

Mais especificamente, a RSL realizada neste trabalho tem o objetivo de identificar características contextuais (escopos de evasão, redes, modalidades e níveis de ensino), técnicas (tarefas, categorias de algoritmos e ferramentas de MDE) e de dados (tipos, abrangência e volume) que têm sido exploradas pelas pesquisas que utilizam técnicas de

MDE no contexto da previsão de evasão estudantil, considerando a perspectiva do Congresso Brasileiro de Informática na Educação (CBIE). A metodologia e o planejamento desta RSL, incluindo a definição das questões de pesquisa, das estratégias de busca e dos critérios de seleção e qualidade, são apresentados na Seção 2, enquanto detalhes de sua condução são descritos na Seção 3. Na Seção 4 são sintetizados e discutidos os resultados obtidos, a fim de responder as questões de pesquisa previamente estabelecidas, e, por fim, conclusões e considerações finais são apresentadas na Seção 5.

2. Metodologia e Planejamento da Revisão

De acordo com [Kitchenham and Charters 2007], uma RSL tem o objetivo de identificar e avaliar pesquisas existentes em uma área ou tema de interesse. Esses autores propuseram a metodologia de RSL considerada neste trabalho, que objetiva conhecer o cenário de pesquisa brasileiro acerca do uso de MDE no âmbito da previsão de evasão estudantil, considerando características de contextos, dados e técnicas das publicações do CBIE. Para isso, foram definidas as seguintes questões de pesquisa:

- Q1. Qual o escopo de evasão e as unidades federativas envolvidas nas pesquisas?
- Q2. Quais níveis, modalidades e redes de ensino são considerados pelos estudos?
- Q3. Qual a natureza, abrangência e volume dos dados utilizados?
- Q4. Quais tarefas, técnicas e ferramentas de MDE estão sendo utilizadas?

As buscas desta RSL foram realizadas no Portal de Publicações da Comissão Especial de Informática na Educação (CEIE)¹ da Sociedade Brasileira de Computação, considerando os anais dos eventos vinculados ao CBIE: Simpósio Brasileiro de Informática na Educação (SBIE), Workshop de Informática na Escola (WIE) e Workshops do CBIE (WCBIE). Embora também estivessem disponíveis para consulta, por abordarem trabalhos de caráter formativo, os anais da Jornada de Atualização em Informática na Educação (JAIE) não foram considerados. Em relação à expressão de busca, após um processo de refinamento, que compreendeu a execução e a avaliação de consultas preliminares, decidiu-se utilizar uma única palavra-chave, relacionada ao contexto de pesquisa, junto da sua tradução em Inglês: “evasão OR dropout”.

Como os eventos considerados já restringem o formato (artigo científico) e os idiomas (Português e Inglês) das publicações, foram estabelecidos apenas três critérios de seleção, com o intuito de garantir que os trabalhos: (i) descrevam de forma detalhada a aplicação de uma solução específica, excluindo-se, dessa forma, revisões da literatura; (ii) adotem técnicas de MDE em soluções que tenham a intenção de contribuir para a previsão da evasão estudantil; e (iii) detalhem o desenvolvimento e os resultados da pesquisa. Além disso, para refinar o processo de seleção/exclusão dos estudos, foram observados dois critérios de qualidade: (i) os trabalhos devem ter mais de cinco páginas, fornecendo um detalhamento mínimo da solução; e (ii) em suas validações, as pesquisas devem considerar um conjunto de dados de, no mínimo, 100 instâncias/alunos.

3. Condução da Revisão

Após planejada, em março de 2020, a RSL teve sua execução iniciada, com o auxílio da ferramenta StArt², desenvolvida pela Universidade Federal de São Carlos. A condução das buscas e do processo de seleção dos artigos compreendeu três fases:

¹<https://www.br-ie.org/pub/>

²http://lapes.dc.ufscar.br/tools/start_tool

- Fase 1: foram realizadas as buscas e a coleta dos artigos. Após, foram analisadas as publicações que possuíam os mesmos autores e, para os casos que apresentavam diferentes estágios da mesma pesquisa, apenas a mais recente foi mantida.
- Fase 2: foram analisados os títulos e resumos, tendo sido excluídos os trabalhos que demonstravam não atender a algum dos critérios de seleção e qualidade.
- Fase 3: foram analisados os textos completos e excluídos os estudos que não satisfaziam os critérios estabelecidos no planejamento.

Embora não tenham sido estabelecidas restrições de ano/período nas buscas da Fase 1, ao se executar as consultas novamente, em junho de 2020, foi identificado um aumento no número de artigos retornados, decorrente da aparente inclusão de algumas edições antigas do CBIE no repositório. Dessa forma, embora esses estudos adicionais tenham sido analisados e incorporados a essa revisão, **faz-se importante mencionar que o Portal da CEIE ainda não contempla todas as edições dos eventos do CBIE.**

Como resultado da fase inicial de seleção, obteve-se 104 estudos, publicados no intervalo de 2006 a 2019. O quantitativo de artigos resultante de cada fase de seleção, considerando os eventos que compõem o CBIE separadamente, é apresentado na Figura 1(a). Pode-se observar que o SBIE detém o maior número de estudos em todas as fases, seguido pelo WCBIE. O WIE teve a totalidade de suas publicações eliminadas já na Fase 2, sendo a maioria relacionada a relatos de experiências envolvendo o uso de tecnologias digitais em ambientes educacionais, o que vai ao encontro da natureza aplicada do evento.

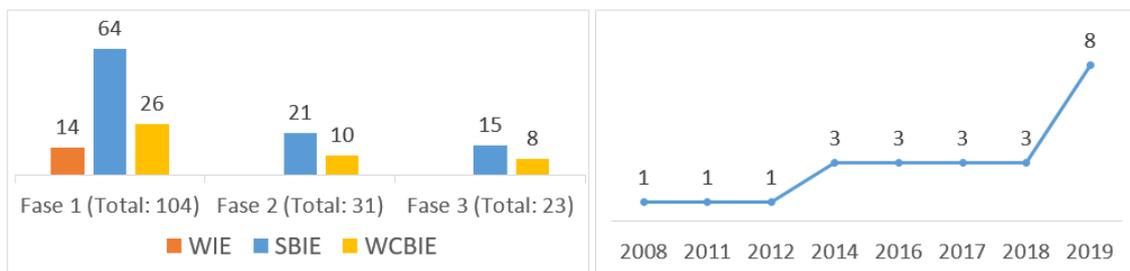


Figura 1(a). Artigos por evento e fase de seleção. Figura 1(b). Artigos selecionados por ano.

Figura 1. Evolução da seleção (a) e distribuição temporal (b) dos artigos.

Na Figura 1(b) é apresentada a distribuição dos 23 artigos selecionados³, que correspondem a 22% dos estudos coletados na Fase 1, por ano de publicação. É possível observar um aumento significativo no quantitativo de publicações na última década, principalmente no último ano, que superou o dobro do número de artigos dos anos anteriores. Essa evolução, além de indicar um aumento no interesse da aplicação de MDE no contexto da evasão estudantil, confirma o potencial de exploração dessa linha de pesquisa.

4. Resultados e Discussões

Nesta seção serão sintetizados e discutidos os dados extraídos dos artigos selecionados, com o intuito de responder as questões de pesquisa, estabelecidas na Seção 2. Para facilitar a compreensão, essas informações serão organizadas em subseções: a Subseção 4.1 descreverá características contextuais, relacionadas a Q1 e Q2; enquanto as Subseções 4.2 e 4.3 apresentarão resultados sob as perspectivas dos dados e técnicas utilizadas nos estudos, respondendo a Q3 e Q4, respectivamente.

³Lista dos artigos selecionados: <http://bit.ly/artigosCBIE>

4.1. Contexto

Estendendo a todos os níveis de ensino a categorização proposta por [Brasil 1996, p.16], pode-se distinguir a evasão entre os escopos de **curso**, **instituição** e **sistema**, que ocorrem, respectivamente, quando um aluno desliga-se do curso, da instituição ou do nível de ensino em que estava matriculado. Adicionalmente, com base nos trabalhos selecionados, nesta RSL também foi considerado o escopo de **disciplina**, que, embora mais restritivo, também é prejudicial e pode anteceder a evasão nas demais categorias. Dessa forma, a fim de começar a responder a primeira questão de pesquisa, na Figura 2(a), são apresentados os quantitativos de artigos em relação aos escopos de evasão. Embora a maioria dos estudos considere o abandono de curso, faz-se importante mencionar que, dentre os de escopo institucional, a maioria investiga dados da rede privada. Talvez porque esse escopo sempre encerra o vínculo cliente-serviço (ou seja, a evasão de curso, que inclui trocas de curso, pode não ser tão crítica quanto à institucional para a rede privada). Além disso, a totalidade focada na evasão de sistema considera a educação básica, o que se justifica a medida que esses trabalhos, em geral, utilizam dados do Censo Escolar, que consideram as matrículas de todas as instituições do país.

Relacionada ao segundo item da primeira questão de pesquisa, a Figura 2(b) apresenta um mapeamento dos estados envolvidos nos estudos, considerando as afiliações de seus autores (artigos com múltiplas afiliações de um mesmo estado são contabilizados uma única vez para essa unidade federativa). Fica evidenciada uma maior concentração das pesquisas nas regiões Sul (11), Nordeste (11) e Sudeste (6) do Brasil, podendo-se destacar a participação dos estados do Rio Grande do Sul (RS) e do Rio de Janeiro (RJ).

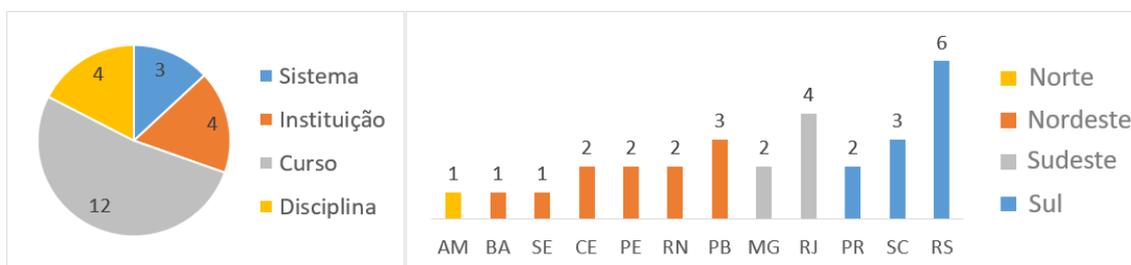


Figura 2(a). Escopos de evasão.

Figura 2(b). Estados dos autores.

Figura 2. Artigos selecionados por escopos de evasão (a) e estados (b).

Também relacionada ao contexto e respondendo à segunda questão de pesquisa, a Figura 3(a) apresenta o quantitativo de artigos por nível e modalidade de ensino considerados. Pode-se perceber que os níveis da educação básica e técnica são pouco explorados, sendo a grande maioria dos trabalhos (18) destinada ao ensino superior, o que acompanha os resultados obtidos por [Maschio et al. 2018], que apontam foco dos estudos brasileiros de MDE no ensino superior e na Educação a Distância (EAD). Porém, em relação à modalidade, os resultados são divergentes e indicam predominância do ensino presencial nas pesquisas que utilizam MDE no contexto da evasão. Essa dominância da modalidade presencial pode, inclusive, ter influência nos resultados relacionados aos escopos de evasão (Figura 2(a)), já que estudos relativos à disciplinas parecem ser uma tendência na modalidade EAD, dada a maior disponibilidade de dados interacionais dos alunos em ambientes virtuais de aprendizagem. A distribuição dos artigos em relação às redes e aos níveis de ensino é apresentada na Figura 3(b). Pode-se notar que a maioria dos trabalhos considera

a evasão em instituições públicas, embora exista uma parcela significativa de pesquisas voltadas ao ensino superior da rede privada. Faz-se importante mencionar que um único estudo é contabilizado seis vezes nesse gráfico por utilizar dados do Censo Escolar, disponibilizados pelo INEP, e sinalizar apenas a filtragem por regiões, sugerindo considerar os níveis Fundamental, Médio e Técnico de ambas as redes de ensino.

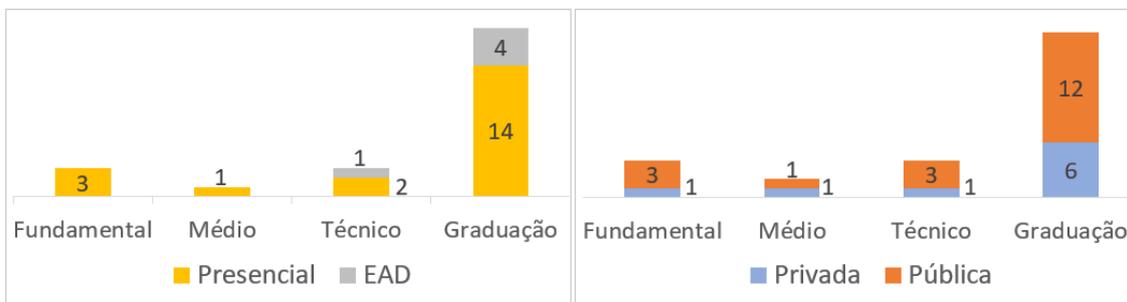


Figura 3(a). Níveis e modalidades de ensino.

Figura 3(b). Níveis e redes de ensino.

Figura 3. Mapeamento por modalidades (a), redes (b) e níveis de ensino.

4.2. Dados

Respondendo à terceira questão de pesquisa, a Figura 4(a) apresenta a distribuição dos artigos selecionados em relação aos tipos (natureza) de dados e aos níveis de ensino considerados. É possível perceber que o único nível presente em todos os tipos é o de Graduação, o que é compreensível por ser o nível mais pesquisado. Observa-se também que os dados acadêmicos e sociais são amplamente utilizados e concentram a totalidade dos estudos vinculados à educação básica, o que se justifica a medida que esses trabalhos, em geral, se baseiam em dados do Censo Escolar. Seguindo a ordem, em proporção de adoção, estão os dados econômicos, interacionais, de análise de sentimentos e temporais. Cabe mencionar que essa última categoria se relaciona apenas ao uso de séries temporais.

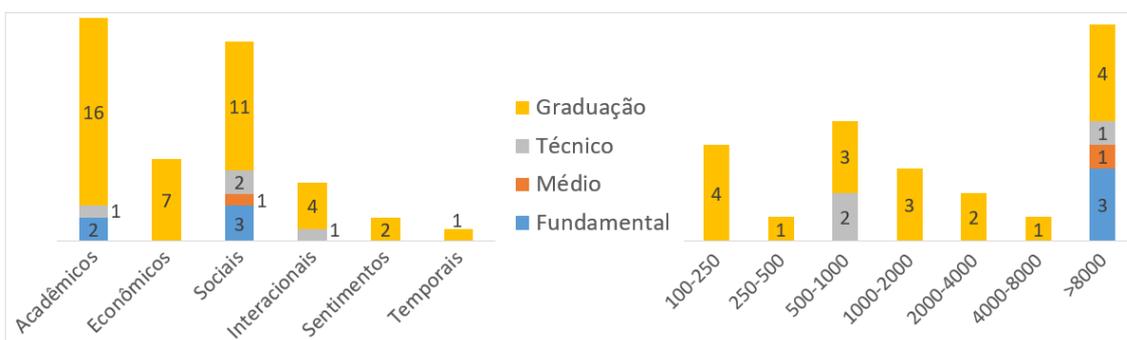


Figura 4(a). Tipos de dados e níveis.

Figura 4(b). Tamanhos dos conjuntos de dados e níveis.

Figura 4. Artigos por níveis e tipos (a) ou tamanhos dos conjuntos (b) de dados.

Na Figura 4(b) é apresentada a relação entre os artigos selecionados e os tamanhos dos conjuntos de dados (nº de alunos) utilizados, considerando cada nível de ensino separadamente. Embora, em geral, os trabalhos se baseiem em conjuntos de dados de tamanhos bastante variados, percebe-se que todos os estudos relacionados à educação básica se encontram na faixa de maior abrangência dos dados, o que também pode ser

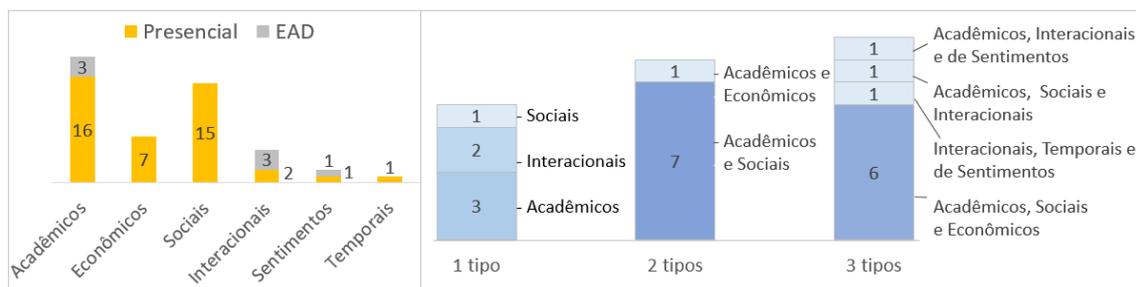


Figura 5(a). Tipos de dados e modalidades.

Figura 5(b). Quantidade de tipos de dados utilizada.

Figura 5. Artigos por tipos de dados e modalidades (a); e por nº de tipos (b).

explicado pela adoção de informações do Censo Escolar, considerando o grande volume de dados disponibilizados.

A relação entre os artigos e a natureza dos dados utilizados, considerando a perspectiva da modalidade de ensino, é apresentada na Figura 5(a). Observa-se que, para o ensino presencial, a tendência de adoção dos tipos de dados se mantém a mesma já identificada na Figura 4(a). Porém, na modalidade EAD, os dados interacionais passam a ter a mesma representatividade dos acadêmicos, o que é natural, já que, como apontado por [Maschio et al. 2018], esses trabalhos costumam explorar dados da interação dos alunos com os ambientes virtuais de aprendizagem. Na Figura 5(b), pode-se visualizar um mapeamento acerca da quantidade de tipos de dados considerada nos trabalhos. Percebe-se que a maioria dos estudos utiliza dois ou três tipos de dados, seguida pelos trabalhos que adotam um tipo. Quando um único tipo é usado, pode-se notar que ele é, em geral, acadêmico ou interacional. Dados acadêmicos e sociais são adotados em conjunto pela grande maioria dos estudos que envolvem dois tipos e, por fim, as pesquisas que consideram 3 tipos utilizam, por padrão, dados acadêmicos, sociais e econômicos.

4.3. Técnicas

Todos os trabalhos selecionados nesta RSL consideram a tarefa de Classificação para identificar alunos em risco ou verificar os atributos e regras com maior influência na evasão. Os algoritmos utilizados foram categorizados em relação aos métodos/técnicas que representam, sendo essas categorias apresentadas na Figura 6(a). Pode-se observar que os algoritmos de árvores de decisão são adotados na maioria dos trabalhos (16), o que pode ser explicado pela facilidade de interpretação dos seus resultados [Han et al. 2012]. Modelos probabilísticos e redes neurais aparecem em dez estudos, sendo seguidos pelas categorias de aprendizagem baseada em instâncias, *ensemble* e máquinas de vetores de suporte, que também apresentam recorrência significativa. O mapeamento das ferramentas e linguagens de programação utilizadas no processo de MDE é apresentado na Figura 6(b). Note que um trabalho pode utilizar mais de uma ferramenta, assim como diversos algoritmos. O gráfico demonstra que a ferramenta Weka é adotada por mais da metade dos estudos, sendo seguida pela linguagem R, que, embora não conste na síntese, pode estar vinculada ao uso de alguns pacotes, como o *caret*. A ferramenta SPSS e a linguagem Python também apresentam recorrência e, assim como em R, o uso da linguagem Python aparece associado a algumas bibliotecas, como *scikit-learn* e *XGBoost*.

Além das características técnicas já apresentadas, faz-se importante posicionar os estudos em relação ao atendimento/satisfação de alguns aspectos adicionais, conforme

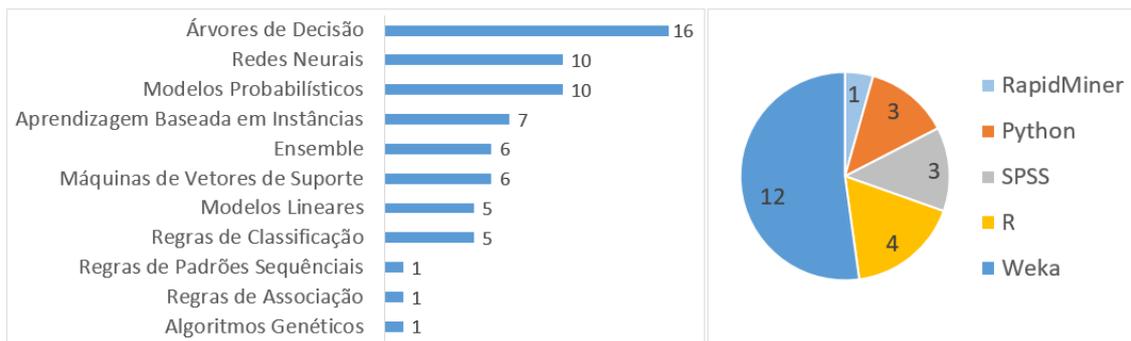
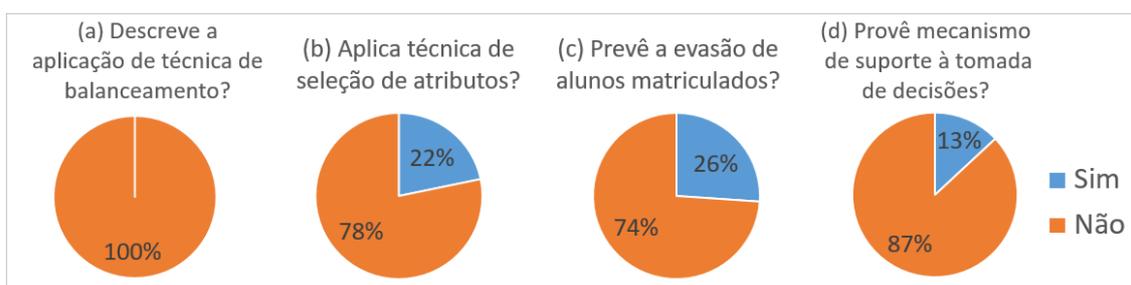


Figura 6(a) - Categorias dos algoritmos utilizados.

Figura 6(b). Ferramentas ou linguagens.

Figura 6. Artigos por categorias de algoritmos (a) e ferramentas (b) utilizadas.

apresentado na Figura 7. As proporções dos trabalhos perante a sinalização de adoção de técnicas de balanceamento e seleção de atributos na etapa de pré-processamento dos dados, são mostradas nas Figuras 7 (a) e (b), respectivamente. Essas análises são necessárias pois dados desbalanceados tendem a ser uma realidade no contexto da evasão (ou seja, a quantidade de alunos evadidos é, em geral, menor que a de persistentes) e podem prejudicar a detecção de padrões da classe sub-representada, refletindo negativamente no desempenho do algoritmo de mineração. Além disso, técnicas de seleção de atributos visam remover dimensões irrelevantes ou redundantes dos dados, reduzindo o custo de processamento e evitando que essas variáveis influenciem negativamente na geração dos modelos de classificação [Han et al. 2012]. Apesar de seus possíveis benefícios, pode-se observar que nenhum trabalho cita a aplicação de técnicas de balanceamento. Já as técnicas de seleção de atributos são consideradas em 22% (5) dos estudos, o que representa uma boa taxa de aplicação ao se considerar que alguns trabalhos já partem de um número reduzido de atributos, por falta de acesso a mais informações ou por realizarem seleção manual, baseada em experiências ou conhecimento prévio, das variáveis.

**Figura 7. Proporções de atendimento a determinadas características técnicas.**

Na Figura 7(c) é apresentada a proporção dos trabalhos que preveem automaticamente o risco de evasão para alunos matriculados (ativos). Pode-se perceber que, embora muitas pesquisas (74%) apliquem e validem técnicas de MDE em dados históricos de evasão, a fim de identificar padrões relacionados a esse fenômeno ou definir o melhor algoritmo de predição, apenas 26% (6) dos trabalhos utilizam os modelos construídos em dados do presente, para prever e detectar automaticamente evasões futuras. Outro fator a ser observado é que nem todos esses trabalhos disponibilizam suas previsões, de forma intuitiva e contínua, aos gestores acadêmicos, suportando a tomada de decisões.

Na Figura 7(d), pode-se perceber que apenas 13% (3) dos estudos disponibilizam mecanismos/ferramentas para consulta ou monitoramento das previsões, facilitando o acompanhamento e o desenvolvimento de estratégias preventivas junto aos alunos em risco. Também se faz importante mencionar que todas essas ferramentas se destinam à rede privada. Esses resultados revelam que poucos trabalhos têm contribuído de forma efetiva na prática educacional, principalmente na rede pública de ensino, o que caracteriza um subaproveitamento dos esforços e potenciais de grande parte das pesquisas.

5. Conclusões e Considerações Finais

Neste trabalho foi apresentada uma revisão sistemática dos estudos publicados no CBIE acerca da aplicação de MDE no âmbito da previsão de evasão estudantil, tendo como finalidade a identificação de aspectos contextuais, técnicos e de dados que vêm sendo explorados no cenário de pesquisa brasileiro.

Em relação ao contexto, os resultados apontaram que a maioria das pesquisas, além de considerar o escopo de evasão em curso, se concentra nas regiões Sul, Nordeste e Sudeste, com destaque para os estados do Rio Grande do Sul e do Rio de Janeiro. Os trabalhos também demonstraram uma maior exploração da evasão no nível de graduação, na modalidade presencial e na rede pública de ensino. Sob a perspectiva das informações utilizadas, ficou evidenciado que, além de apresentarem uma grande variação na quantidade de alunos considerados, as pesquisas adotam de um a três tipos de dados para representá-los, sendo utilizadas, em geral, combinações dos tipos acadêmicos, sociais e econômicos. Considerando as técnicas empregadas, os trabalhos indicam unanimidade da tarefa de classificação e predominância de algoritmos de árvores de decisão, geralmente aplicados com o auxílio da ferramenta Weka. Adicionalmente, a adoção de técnicas de balanceamento não foi citada por nenhum estudo na etapa de preparação dos dados, o que surpreendeu, considerando a natureza, em geral, desbalanceada dos dados de evasão, enquanto técnicas de seleção de atributos foram usadas em 22% das pesquisas, com a finalidade de potencializar o desempenho da classificação.

Ainda sob o âmbito técnico, mas considerando características mais amplas, este trabalho demonstrou que poucas pesquisas aplicam seus modelos de classificação a dados de alunos matriculados, com a finalidade de prever automaticamente os estudantes em risco de evasão. Além disso, apenas metade dos estudos que realizam previsões disponibiliza mecanismos para a consulta de seus resultados, de forma intuitiva e contínua, aos gestores acadêmicos, dando suporte à tomada de decisões e fomentando a implementação de estratégias preventivas. Esses resultados são preocupantes pois indicam um subaproveitamento dos esforços e potenciais de grande parte das pesquisas na prática educacional.

Por fim, este trabalho evidenciou o crescimento, principalmente no último ano, do interesse pela aplicação de técnicas de MDE no contexto da evasão estudantil. Confirmado o potencial dessa temática de pesquisa, espera-se que os resultados apontados nessa revisão contribuam para a realização de novos estudos, instigando, sobretudo, a exploração da educação básica e técnica, a aplicação de modelos preditivos em dados vigentes, e a disponibilização automática e contínua das previsões aos gestores acadêmicos. Essas carências, inclusive, nortearão a continuidade dessa pesquisa.

Agradecimentos: O presente trabalho foi realizado com apoio da Coordenação de

Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001. Os autores agradecem à CAPES, ao PPGC/UFPEL e ao IFFar.

Referências

- Baker, R. S. J. d., Isotani, S., and Carvalho, A. M. J. B. d. (2011). Mineração de dados educacionais: Oportunidades para o Brasil. *Revista Brasileira de Informática na Educação*, 19:3–13.
- Brasil (1996). Diplomação, retenção e evasão nos cursos de graduação em instituições de ensino superior públicas. Technical report, Ministério da Educação, Comissão Especial de Estudos sobre a Evasão nas Universidades Públicas Brasileiras: ANDIFES; ABRUEM; SESu/MEC, Brasília, DF.
- Han, J., Kamber, M., and Pei, J. (2012). *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers, Waltham, MA, 3rd edition.
- INEP (2020). Indicador de fluxo da educação superior 2010-2016. Disponível em: <http://portal.inep.gov.br/indicadores-educacionais>. Acesso em: 17 jun. 2020.
- Kitchenham, B. A. and Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering. Technical Report EBSE-2007-01, School of Computer Science and Mathematics, Keele University, Keele, UK.
- Marques, L. T., Castro, A. F. D., Marques, B. T., Silva, J. C. P., and Queiroz, P. G. G. (2019). Mineração de dados auxiliando na descoberta das causas da evasão escolar: Um mapeamento sistemático da literatura. *Novas Tecnologias na Educação*, 17(3):194–203.
- Maschio, P. d. T., Vieira, M. A., da Costa, N. T., Melo, S., and Júnior, C. P. (2018). Um panorama acerca da mineração de dados educacionais no Brasil. In *Anais do XXIX Simpósio Brasileiro de Informática na Educação (SBIE 2018)*, pages 1936–1940.
- Mduma, N., Kalegele, K., and Machuve, D. (2019). A survey of machine learning approaches and techniques for student dropout prediction. *Data Science Journal*, 18:14:1–10.
- Oliveira Júnior, J. G., Noronha, R. V., and Kaestner, C. A. (2017). Método de seleção de atributos aplicados na previsão da evasão de cursos de graduação. *Revista de Informática Aplicada*, 13(2):54–67.
- Pontili, R., Staduto, J., and Henrique, J. (2018). Abandono e atraso escolar e sua relação com indicadores socioeconômicos: uma análise para a região sul do Brasil. *Gestão & Regionalidade*, 34(101):4–22.
- Romero, C. and Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *WIREs Data Mining and Knowledge Discovery*, 10(3):e1355.
- Silva, G. P. d. (2013). Análise de evasão no ensino superior: uma proposta de diagnóstico de seus determinantes. *Avaliação: Revista da Avaliação da Educação Superior (Campinas)*, 18:311–333.
- Silva Filho, R. L. L., Motejunas, P. R., Hipolito, O., and Lobo, M. B. C. M. (2007). A evasão no ensino superior brasileiro. *Cadernos de Pesquisa*, 37(132):641–659.