

A Descoberta das Causas da Retenção Acadêmica Utilizando Mineração de Dados: Uma Revisão Sistemática da Literatura

Leonardo Torres Marques, UECE, leonardo.torresmarques@gmail.com, ORCID: <https://orcid.org/0000-0001-9733-3626>

Bruno Torres Marques, UFC, brunotores@alu.ufc.br, ORCID: <https://orcid.org/0000-0002-0240-0716>

Carlos Alexandre Morais Silva, UFC, carlosalexandresilva100@gmail.com, ORCID: <https://orcid.org/0000-0002-4180-0181>

Resumo: Nos últimos anos as instituições de ensino superior têm enfrentado grandes problemas ocasionados pela retenção acadêmica. Diante disso, diversos pesquisadores têm dedicado seus esforços a descoberta das causas e a prevenção contra a retenção acadêmica e a mineração de dados educacionais pode auxiliar nesse processo. Porém, para obter resultados satisfatórios com a mineração de dados aplicada ao contexto da retenção acadêmica é necessário conhecer os melhores recursos de mineração de dados a serem utilizados. Dessa forma, no presente trabalho foi realizada uma revisão sistemática da literatura a fim de identificar as principais técnicas e ferramentas da mineração de dados que podem ser aplicadas nesse processo, além de identificar os principais fatores indutores à retenção acadêmica.

Palavras-Chave: revisão sistemática, mineração de dados, educação.

Discovering the Causes of Academic Retention Using Data Mining: A Systematic Review of the Literature

Abstract. In recent years, higher education institutions have faced major problems caused by academic retention. In view of this, several researchers have dedicated their efforts to discovering the causes and preventing academic retention and mining of educational data can help in this process. However, to obtain satisfactory results with data mining applied to the context of academic retention, it is necessary to know the best data mining resources to be used. Thus, in the present work, a systematic literature review was carried out in order to identify the main data mining techniques and tools that can be applied in this process, in addition to identifying the main factors that induce academic retention.

Keywords: systematic review, data mining, education.

1. Introdução

Nos últimos anos, o ensino superior brasileiro tem passado por grandes transformações. As diversas ações de apoio ao setor acadêmico como, por exemplo, o Sistema de Seleção Unificada (SISU), têm impulsionado o crescimento do ensino superior no país. Porém, apesar da evolução ainda existe uma grande preocupação com alguns problemas, tais como a retenção e evasão acadêmica (MORAES, 2015).

A retenção acadêmica é o termo utilizado para se referir ao processo que resulta no atraso da conclusão de curso pelos discentes do ensino superior, levando-os a uma permanência prolongada nas Instituições de Ensino Superior (IES) (MORAES, 2015). Para Moraes (2015), um elevado número de reprovações e/ou trancamentos podem ocasionar uma desmotivação do discente, levando-o a evasão acadêmica.

Para Espíndola (2016), além da retenção acadêmica ocasionar problemas na vida pessoal e profissional dos discentes, ela também afeta as instituições, pois elas não atingem os índices esperados em relação à formação de seus discentes. Segundo Santos (2015), a retenção acadêmica gera grandes problemas para as IES, pois essas necessitam desembolsar mais recursos para custear os discentes retidos. Espíndola (2016), ressalta ainda os impactos sofridos pela sociedade, pois a retenção acadêmica retarda a disponibilidade de profissionais no mercado de trabalho.

Uma possibilidade plausível de identificação dos fatores motivadores para a retenção acadêmica é o emprego da Mineração de Dados Educacionais (MDE). Segundo Santos (2015), a MDE consiste na aplicação de técnicas de mineração de dados e/ou modelos probabilísticos em dados da educação.

No entanto, para se aplicar a MDE no contexto apresentado (retenção acadêmica) é fundamental utilizar as ferramentas adequadas dentre as tantas que existem como, por exemplo: *Waikato Environment for Knowledge Analysis (WEKA)*, *Jupyter Notebook* e *Python* (MARQUES *et al.*, 2019), além disso, é importante saber quais técnicas de Mineração de Dados (MD) se aplicam melhor ao contexto (MARQUES *et al.*, 2019). Outra questão, é saber quais fatores serão estudados para entender o que leva os discentes a ficarem retidos.

Tendo em vista o que foi apresentado, uma solução para conhecer as principais ferramentas, técnicas e fatores explorados no estudo do tema retenção acadêmica, é a realização de uma Revisão Sistemática de Literatura (RSL). Segundo Cruz e Nunes (2012), a RSL é uma análise geral de trabalhos, que têm preestabelecidos os materiais e métodos, além disso em seu desenvolvimento ela torna explícitos os métodos de forma que possam ser reproduzidos. A RSL faz uso de um protocolo válido e explícito, que reduz o viés e faz a inclusão de resumos qualitativos e quantitativos (CRUZ; NUNES, 2012).

Diante desse contexto, os objetivos desta Revisão Sistemática da Literatura são: (i) identificar os principais fatores contribuintes para a retenção acadêmica a serem estudados; (ii) avaliar ferramentas e algoritmos da MD que apresentam melhores desempenhos com dados da retenção acadêmica; (iii) identificar as técnicas de MD que apresentam melhor desempenho, quando aplicadas a retenção acadêmica.

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 nesta RSL. Salienta-se, que esta RSL se baseia no protocolo definido por Biolchini *et al.* (2007), em que são aludidos os objetivos da pesquisa e 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 extração e apresentação dos dados.

2.1 Questões de Pesquisa

De acordo com os objetivos apresentados na Seção 1, foram definidas as seguintes questões de pesquisa que direcionaram esta RSL:

- **Questão primária (QP1):** Quais os fatores indutores têm sido investigados para a descoberta das causas da retenção acadêmica?
- **Questão primária (QP2):** Quais ferramentas de MD têm sido utilizadas para descoberta das causas da retenção acadêmica?
- **Questão primária (QP3):** Quais técnicas de MD têm sido utilizadas para descoberta das causas da retenção acadêmica?

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

A determinação das palavras-chave em RSLs é estabelecida a partir das questões de pesquisa sob investigação. Preliminarmente, foram realizados testes com as strings de busca que seriam utilizadas nas buscas. No Quadro 1, são apresentadas as palavras-chave e seus sinônimos utilizados para montar a *string* de busca.

Quadro 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>Academic Retention</i>	<i>Academic Retention, retention in higher education, delay in completing higher education</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 (“academic retention” OR “retention in higher education” OR “delay in completing higher education”)

2.3 Intervalo de Tempo para Pesquisa

No presente trabalho, considerou-se estudos realizados no período dos últimos dez (10) anos, compreendido entre dois mil e onze (2011) e (dezembro de) dois mil e vinte e um (2021), por se considerar que, as constantes evoluções, no que se trata de fatores, abordagens, ferramentas e técnicas tem aperfeiçoado a busca das possíveis causas da retenção.

2.4 Bases de Dados de Pesquisa

Para a RSL em questão, foram definidas quatro (4) bases de dados para pesquisa conforme se apresenta no Quadro 2. 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 de revistas científicas e seus respectivos trabalhos, assentes por qualidade e originalidade (DANTAS *et al.* 2018).

Quadro 2 – Lista de fontes.

Fontes	Endereço Eletrônico (link)
---------------	-----------------------------------

<i>IEEEExplore</i>	http://ieeexplore.ieee.org/Xplore/home.jsp
<i>Science Direct</i>	http://www.sciencedirect.com/
<i>Periódicos da CAPES</i>	https://www-periodicos-capes-gov-br.ez11.periodicos.capes.gov.br/

2.5 Critérios de Inclusão e Exclusão

Os critérios de inclusão e exclusão são definidos para auxiliar na condução de uma RSL com o intuito de apoiar a classificação de relevância dos estudos (BIOLCHINI *et al.*, 2007). 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.

Os seguintes Critérios de Inclusão (CI) de trabalhos foram definidos:

- **CI1:** trabalho discute, aplica, propõe ou investiga, fatores ou dados relacionados ao problema da retenção acadêmica no contexto educacional;
- **CI2:** trabalho discute, aplica, propõe ou investiga, ferramentas de MD relacionado ao problema da retenção acadêmica no contexto educacional;
- **CI3:** trabalho discute, aplica, propõe ou investiga, técnicas de exploração ou descoberta de conhecimento em bases de dados relacionado ao problema da retenção acadêmica no contexto educacional;
- **CI4:** trabalho aborda estudos experimentais para validar os resultados que tratam do problema da retenção acadêmica.

Os seguintes Critérios de Exclusão (CE) de trabalhos foram definidos:

- **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 os demais.

2.6 Procedimentos para Seleção dos Estudos

Para a seleção dos trabalhos, foram seguidas as três (3) fases apresentadas 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.

Para garantir um grupo de trabalhos de qualidade, essas fases foram executadas por dois pesquisadores experientes.

3. Condução da RSL

Nesta seção, são apresentados os detalhes do 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.

Quadro 3 – Resultados das buscas.

Base	Fase 1	Fase 2	Fase 3
<i>IEEEExplore</i>	95	3	3
<i>Science Direct</i>	75	2	2
<i>Periódicos CAPES</i>	316	4	1
Total	486	9	6

Pode-se observar no Quadro 3 os quantitativos dos trabalhos analisados em cada fase. Na Fase 1 foram retornados quatrocentos e oitenta e seis (486) trabalhos. Na Fase 2, após a leitura dos títulos e resumos, nove (9) trabalhos foram selecionados para serem avaliados na fase seguinte. Na fase 3, após a leitura completa dos trabalhos e aplicando-se os critérios (inclusão/exclusão), foram selecionados seis (6) trabalhos.

No Quadro 4 são apresentados os trabalhos selecionados, com os respectivos anos de publicação e o continente em que a pesquisa foi explorada.

Quadro 4 – Relação dos trabalhos selecionados.

Índice	Trabalhos	Ano Publicação	Continentes
#1	(NANDESHWAR <i>et al.</i> , 2011)	2011	América do Norte
#2	(POKRAJAC <i>et al.</i> , 2016)	2016	América do Norte
#3	(SHUQFA; HAROUS, 2019)	2019	Ásia
#4	(CARDONA <i>et al.</i> , 2019)	2019	América do Norte
#5	(SARASWATI <i>et al.</i> , 2020)	2020	Ásia
#6	(ULIYAN <i>et al.</i> , 2021)	2021	Ásia

Em relação à distribuição por continentes, é possível observar que o interesse e pesquisas em relação à retenção acadêmica, tem se concentrado em apenas dois continentes América do Norte e Ásia. Vale ressaltar, que os Estados Unidos tem sido o país que mais investiga sobre o assunto.

4. Resultados e Discussões

Nesta Seção são apresentadas as respostas para as questões de pesquisa, que foram encontradas a partir da análise dos trabalhos selecionados. Por meio de formulários de extração de dados, buscaram-se pontos importantes dos trabalhos que respondessem às questões de pesquisa.

4.1 Quais fatores indutores têm sido investigados para a descoberta das causas da retenção acadêmica?

Na Figura 1, são exibidos os fatores que influenciam na retenção acadêmica, com os índices dos respectivos trabalhos retornados na RSL, a relação de índices e trabalhos é

apresentada no Quadro 4. Salienta-se que em quase todos os trabalhos, foram explorados dados inerentes aos fatores internos da instituição e as características do aluno. Destaca-se que apenas no trabalho (#6) foram explorados apenas dados inerentes às características individuais do aluno, enquanto nos outros trabalhos foram explorados mais de um fator.

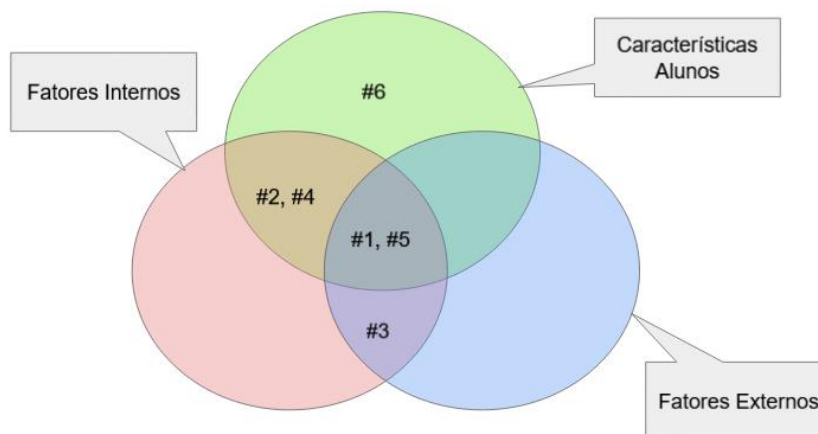


Figura 1 - Fatores Indutores para à retenção acadêmica.

Os fatores relacionados às características individuais do aluno, abordados nos trabalhos (#1, #2, #4, #5 e #6) foram principalmente dados relacionados ao desempenho acadêmico como notas e reprovações. Além disso, foram identificados tais fatores motivantes para a retenção: fatores pertinentes à adoção precoce da profissão, formação escolar anterior, insatisfação ou desinteresse dos alunos com cursos elegidos em segunda (2ª) ou terceira (3ª) opção, embaraços pessoais de adequação à vida universitária, falta de informação em relação a natureza dos cursos, descoberta de novos interesses que levam ao cumprimento de novo processo seletivo, dificuldades financeiras do aluno, divergências entre a vida acadêmica e as exigências do mundo do trabalho, fraco desempenho e dificuldades de aprendizagem, constantes reprovações e/ou baixa frequência às aulas, habilidade de estudo, personalidade do estudante, entre outros.

Em relação aos fatores associados às instituições, abordou-se nos trabalhos (#1, #2, #3, #4 e #5) dados relativos a apoio aos discentes, acompanhamento aos discentes, ações de incentivo e oferta de bolsas. Outros fatores que valem destacar: conectados à questões acadêmicas (por exemplo: currículos ultrapassados e distendidos, cadeia de pré-requisitos inteiriçados, além da falta de clareza sobre o próprio projeto pedagógico do curso); cultura institucional de desmerecimento da docência na graduação; estrutura débeis de apoio ao ensino de graduação (laboratórios, equipamentos de informática, bibliotecas, etc.); didático-pedagógicas (métodos inadequados de avaliação do desempenho do aluno, deficiência de formação pedagógica ou desinteresse do professor, desprovimento ou número reduzido de programas institucionais para o estudante, como iniciação científica, monitoria, entre outros); deficiência do sistema público nacional que proporcione a racionalização da utilização das vagas, suprimindo a possibilidade da matrícula em duas universidades.

Tratando-se de fatores externos às instituições nos trabalhos (#1, #3 e #5) foram estudados dados concernentes ao meio em que o aluno está inserido. São exemplos: incentivo da família e da sociedade e empregos. Outros fatores: concernentes ao mercado de trabalho, à qualidade do ensino fundamental e médio associados às condições econômicas, ao reconhecimento social da profissão escolhida. Além disso, ainda se constata a depreciação das profissões, a exemplo, nos cursos de licenciaturas, às dificuldades de modernizar a universidade frente aos avanços tecnológicos, econômicos

e sociais da atualidade; à falta de políticas públicas plausíveis e continuadas direcionadas para o ensino de graduação.

4.2 Quais ferramentas de MD têm sido utilizadas para descoberta das causas da retenção acadêmica?

No Quadro 5 é apresentado o quantitativo de uso das ferramentas de MD nos trabalhos selecionados. Destaca-se que em apenas três trabalhos foi possível identificar as ferramentas que foram utilizadas.

Quadro 5 – Ferramentas utilizadas na MD.

Índice	Ferramentas	Quantidade de trabalhos
1	WEKA	1
2	JMP	1
3	STATISTICA 12	1
4	Não identificados	3

A ferramenta JMP foi utilizada no trabalho de Pokrajac *et al.* (2016). JMP é um conjunto de programas de computador para análise estatística desenvolvido pela unidade de negócios JMP do SAS *Institute*. Tal ferramenta permite lidar com problemas estatísticos de rotina e difíceis. Desde o acesso fácil aos seus dados de várias fontes até o uso de ferramentas de preparação de dados rápidas e confiáveis e a realização de análises estatísticas de escolha, permitindo aproveitar ao máximo os dados de um determinado *dataset* em qualquer situação.

A ferramenta WEKA foi utilizada no trabalho de Shuqfa e Harous (2019). A ferramenta é uma coleção de algoritmos de Aprendizado de Máquina (AM) para tarefas de MD, oferece aparatos para preparação de dados, Classificação, Regressão, Agrupamento, Mineração de Regras de Associação e Visualização. O WEKA está implementado na linguagem Java, que tem como principal característica a sua portabilidade, desta forma é possível utilizá-la em diferentes sistemas operacionais, além de aproveitar os principais benefícios da orientação a objetos. O WEKA é um software livre, ou seja, está sob domínio da licença GPL.

A ferramenta STATISTICA 12 foi utilizada no trabalho de Cardona *et al.* (2019). STATISTICA é um conjunto de produtos e soluções de software de análise originalmente desenvolvidos pela *StatSoft* e adquiridos pela Dell em março de dois mil e quatorze (2014). O software inclui uma série de procedimentos de análise de dados, gerenciamento de dados, visualização de dados e mineração de dados; bem como uma variedade de Modelagem Preditiva, Agrupamento, Classificação e Técnicas Exploratórias.

Nos trabalhos de Nandeshwar *et al.* (2011), Saraswati *et al.* (2020) e Uliyan *et al.* (2021), não foi possível identificar as ferramentas que foram utilizadas.

4.3 Quais técnicas de MD têm sido utilizadas para descoberta das causas da retenção acadêmica?

No Quadro 6 são exibidas as técnicas de MD que foram exploradas nos trabalhos encontrados, enfatiza-se as técnicas de classificação que foram utilizadas em quatro (4) trabalhos.

Quadro 6 – Técnicas e algoritmos utilizados na MD.

Índice	Técnicas	Quantidade de trabalhos
1	<i>Classificação</i>	4
2	<i>Regressão</i>	1
3	<i>Redes Neurais</i>	1

A técnica de Classificação foi explorada nos trabalhos de Nandeshwar *et al.* (2011), Pokrajac *et al.* (2016), Cardona *et al.* (2019) e Saraswati *et al.* (2020). 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. Quanto aos algoritmos que foram utilizados nesses trabalhos, destacam-se os algoritmos baseados em Árvores de Decisão (onde os nós internos são rotulados com uma *feature* de entrada, e os nós folhas são rotulados com a classe a ser preterida). Ainda sobre a técnica de Classificação, também foram empregados algoritmos de *Naive Bayes* (que é um classificador que trabalha com probabilidades de ocorrência de cada classe para cada valor de atributo, supondo que as variáveis são independentes).

A técnica de Regressão abordada no trabalho de Shuqfa e Harous (2019), é utilizada para se prever um valor real. A ideia é prever um valor numérico; ou em outros termos: identificar uma categoria em escala contínua (pode ser até uma probabilidade). Neste caso, o modelo pode aprender uma função para prever em qual período do curso um aluno irá evadir do curso, o preço de um imóvel, uma demanda de venda, probabilidade de superlotação de um estoque, o tempo que levará para uma máquina apresentar defeito, uma nota de um cliente para aumentar seu limite de crédito, ou qualquer outro valor quantitativo.

A técnica de Rede Neural (baseadas em redes de neurônios artificiais interconectados – os *perceptrons*, onde os pesos das camadas ocultas são ajustados a cada iteração) foi utilizada no trabalho de Uliyan *et al.* (2021), essa técnica é inspirada no sistema nervoso central de um animal, para realização de AM e reconhecimento de padrões.

5. Conclusões e Considerações Finais

Neste trabalho foram apresentados o planejamento, a condução e os resultados de uma RSL sobre a retenção acadêmica, em que se buscou elencar ferramentas, técnicas e fatores indutores utilizados na aplicação da MD na retenção acadêmica nos últimos dez (10) anos. A busca pelos trabalhos resultou na pré-seleção de nove (9) trabalhos, dentre os quais seis (6) foram incluídos para a extração de dados.

A partir dos resultados, foi possível verificar que a técnica de Classificação vem sendo bastante utilizada e tem retornado alta precisão nas previsões de tendência de retenção acadêmica. Quanto aos algoritmos de Classificação utilizados, vale destacar que foram utilizados vários e algoritmos utilizados para diversos propósitos (*Árvores de Decisão* e *Naive Bayes*).

Em relação as ferramentas, em apenas cinquenta por cento (50%) dos trabalhos utilizados na pesquisa, foi possível identificar a ferramenta utilizada. Contudo, verificou-se que as ferramentas JMP, WEKA e STATISTICA 12 vem sendo utilizadas para desvendar as causas da retenção acadêmica. Por mais que não seja possível inferir qual(is) são as mais utilizadas e eficientes, com tal descoberta é possível nortear pesquisadores sobre quais ferramentas utilizarem em suas pesquisas.

Quanto aos fatores indutores para a retenção acadêmica, constatou-se que em quase todos os trabalhos são investigadas possíveis causas relacionadas a fatores internos, ligados a instituição. Tais fatores como por exemplo os programas de apoio e incentivo

aos discentes, mostraram-se ser de grande influência (positiva ou negativa) para a retenção acadêmica. Além disso, em cinco (5) dos seis (6) trabalhos utilizados na pesquisa, foram abordados fatores relacionados as características dos alunos, principalmente fatores relacionados ao desempenho acadêmico individual (como as notas das disciplinas e índices de aprovação e reprovação), esses fatores também apresentaram excelentes resultados na previsão da retenção acadêmica.

É importante enfatizar que qualquer pesquisa desta natureza possui riscos à validade dos seus resultados. Para esta RSL em particular, detectam-se os seguintes pontos de fragilidade: a pré-seleção dos trabalhos, devido ao conjunto limitado de bases consultadas, ao número de trabalhos 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 a 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 uma RSL, podendo conduzir a uma diferente captura de resultados nas bases consultadas, em diferentes datas.

Ainda vale destacar as contribuições diretas deste trabalho, mostrou-se um panorama sobre o que vem sendo pesquisado e empregado em relação ao tema retenção acadêmica. Assim, com a garimpagem feita neste estudo abrisse um caminho para descobrir ainda de início, discentes com alto potencial de ficar retidos e, conseqüentemente, providenciasse medidas que visem amenizar tal problema.

Por fim, como trabalhos futuros sugere-se aplicar a *string* de busca em outras bases de dados, acrescentar novas palavras-chave, buscar possíveis soluções para o problema abordado neste estudo e desenvolver um módulo inteligente que possa detectar previamente os discentes com tendência de ficarem retidos.

Referências Bibliográficas

- BIOLCHINI, J. C. de A.; MIAN, P. G.; NATALI, A. C. C.; CONTE, T. U.; TRAVASSOS, G. H. Scientific research ontology to support systematic review in software engineering. **Adv. Eng. Inform.**, Elsevier Science Publishers B. V., Amsterdam, The Netherlands, The Netherlands, v. 21, n. 2, p. 133–151, abr. 2007. ISSN 1474-0346. Disponível em: <http://dx.doi.org/10.1016/j.aei.2006.11.006>. Acesso em: 14 set. 2021.
- CARDONA, T. A. et al. Predicting student retention using support vector machines. **Procedia Manufacturing**, Elsevier, [S.I.], v. 39, p. 1827–1833, 2019.
- CRUZ, A.; NUNES, H. Prevalência e fatores de risco de dores nas costas em adolescentes: uma revisão sistemática da literatura. **Revista de Enfermagem Referência**, Escola Superior de Enfermagem de Coimbra, v. 3, n. 6, p. 131–146, 2012.
- ESPÍNDOLA, Q. C. **Retenção acadêmica nos cursos de Licenciatura em Ciências da Natureza da UNIPAMPA**. Pampa: Universidade Federal do Pampa, 2016.
- MARQUES, Leonardo Torres et al. Mineração de dados auxiliando na descoberta das causas da evasão escolar: Um mapeamento sistemático da literatura. **RENOTE**, v. 17, n. 3, p. 194-203, 2019.
- MORAES, J. d. P. B. **Retenção discente no Centro de Ciências Humanas e Naturais da UFES: o caso dos cursos letras-português e geografia**. Vitória: Universidade Federal do Espírito Santo, 2015.

- NANDESHWAR, A.; MENZIES, T.; NELSON, A. Learning patterns of university student retention. **Expert Systems with Applications**, Elsevier, v. 38, n. 12, p. 14984–14996, 2011.
- POKRAJAC, D. D.; SUDLER, K. R.; EDAMATSU, P. Y.; HARDEE, T. Prediction of retention at historically black college/university using artificial neural networks. In: IEEE. 2016 13th **Symposium on Neural Networks and Applications (NEUREL)**. [S. l.], 2016. p. 1–6.
- SANTOS, M. **Uma análise probabilística da retenção na Universidade Federal da Bahia**: Um estudo de caso no curso de ciência da computação. Salvador: Instituto de Matemática. Departamento de Ciência da Computação, 2015.
- SARASWATI, N.; WIDIARTHA, K.; PRAPITASARI, L. Vector machine to predict student retention: A computerized approach. In: IOP PUBLISHING. **Journal of Physics: Conference Series**. [S. l.], v. 1469, n. 1, p. 012045, 2020.
- SHUQFA, Z.; HAROUS, S. Data mining techniques used in predicting student retention in higher education: A survey. In: IEEE. **2019 International Conference on Electrical and Computing Technologies and Applications (ICECTA)**. [S. l.], 2019. p. 1–4.
- ULIYAN, D.; ALJALOU, A. S.; ALKHALIL, A.; AMER, H. S. A.; MOHAMED, M. A. E. A.; ALOGALI, A. F. M. Deep learning model to predict students retention using blstm and crf. **IEEE Access**, IEEE, [S.l.], v. 9, p. 135550–135558, 2021.