

Levantamento Bibliográfico sobre Mineração de Dados na Educação: Um Mapeamento Sistemático da Literatura

Martony Demes da Silva¹, James da Luz Dias¹

¹Departamento de Computação -- Universidade Federal do Piauí (UFPI)
Teresina -- PI -- Brazil

{mardemes, jamesdias0}@gmail.com

Abstract. *Educational Data Mining (MDE) is the process of discovering information in a given educational database. This enhances the analysis, presentation and improvement of the educational process. In this scenario, the purpose of this research is to present a bibliographic survey on DEM. To this end, a Systematic Mapping of Literature was built to classify and organize the various published works. The results show which approaches were most explored, the most used tools and algorithms. In addition, it presents evidence of some gaps with possibilities for future research.*

Resumo. *A Mineração de Dados Educacionais (MDE) é o processo de descoberta de informações em uma determinada base de dados educacional. Isso potencializa a análise, a apresentação e a melhoria do processo educacional. Nesse cenário, o propósito desta pesquisa é apresentar um levantamento bibliográfico sobre MDE. Para tanto, foi construído um Mapeamento Sistemático da Literatura para classificar e organizar os diversos trabalhos publicados. Os resultados apresentam quais as abordagens foram mais exploradas, as ferramentas e os algoritmos mais utilizados. Além disso, apresenta evidências de algumas lacunas com possibilidades de pesquisas futuras.*

1. Introdução

Mineração de Dados Educacionais (MDE) tem como objetivo aplicar métodos para explorar uma base de dados em ambientes educacionais [Baker et al. 2011]. A MDE pode ser utilizada para aprender, prever e compreender o comportamento da aprendizagem [Baker 2014]. Assim, é possível identificar relações entre dados e produzir conhecimento.

A MDE é útil em análise e visualização de dados abertos educacionais e dados acadêmicos. Também é abordada para avaliar desempenho [Rabelo et al. 2017] e comportamento [Kampff 2009] no âmbito educacional. Outrossim, a MDE é empregada em mineração de texto [Dionísio et al. 2017]. Dessa forma, há diversas aplicações de mineração de dados (MD) no âmbito educacional.

Nesse contexto, alguns artigos propuseram realizar Pesquisa Bibliográfica (PB) sobre a área de MDE. Dentre esses, não se encontrou trabalho que abrangesse uma quantidade significativa de informações sobre o tema. Investigou-se também que, em nenhum desses trabalhos de PB, abordou-se sobre educação especial.

Diante disso, este trabalho tem o propósito de exibir um levantamento bibliográfico sobre MDE. Objetiva-se, com isso, coletar vários tipos de elementos que classificam e organizam os artigos sobre o assunto. A pesquisa pretende também examinar trabalhos de MDE no contexto de educação especial. A principal relevância

deste trabalho está em levantar e concentrar as vertentes sobre MDE dispersas em vários trabalhos de PB.

Este trabalho está organizado em 5 seções. A seção 2 ilustra o referencial teórico. A Seção 3 apresenta a metodologia aplicada. A Seção 4 mostra os resultados do estudo. Finalmente, a Seção 5 apresenta as conclusões deste trabalho.

2. Referencial Teórico

Mineração de dados educacionais (MDE) é a aplicação de métodos e técnicas de estatísticas e de aprendizado de máquina para descoberta de conhecimento a partir de uma determinada base de dados [Fayyad et al. 1996] no contexto educacional. Nessa perspectiva, verificou-se, em trabalhos relacionados, diversas possibilidades de técnicas de MDE são exploradas: descoberta de informações, classificação em grupos, análise de desempenho, propostas de ferramentas entre outras.

Diante do cenário de crescimento da área, encontrou-se também trabalhos de Pesquisas Bibliográficas (PB) no contexto de MDE. A partir da análise desses trabalhos de PB organizou-se os principais elementos abordados em MD.

O primeiro é o trabalho de [Baker et al. 2011]. Nesta pesquisa de PB, o autor abordou sobre as principais contribuições de MD em apenas duas linhas principais: análise de dados e a criação de modelos para compreender o processo de aprendizagem.

O estudo de [Santos et al. 2017] apresenta uma PB em que são levados em consideração os elementos: os objetivos de pesquisa, as fontes ou bases de dados dos trabalhos e a evolução da pesquisa em determinado tempo.

Na pesquisa de [Maschio et al. 2018] apresenta como metodologia um Mapeamento Sistemático da Literatura (MSL) acerca de MDE. Neste trabalho, os elementos levantados foram: técnicas de MD, objetivos (investigações), nível de escolaridade e tipos de dado.

Já na PB realizada por [Rodrigues et al. 2014] é realizada uma revisão da literatura. Nesta pesquisa, investigou-se tipo de pesquisa utilizada, a classificação da MDE e o objetivo. Além disso, a pesquisa apresentou as ferramentas utilizadas.

Por fim, a abordagem feita no trabalho de [Santos et al. 2016] apresenta levantamento sobre algoritmos, as fontes dos dados e as ferramentas.

Após análise desses trabalhos de PB, verificou-se que em nenhum dos trabalhos apresenta as quatro abordagens são citadas: **i) algoritmos, ii) ferramentas utilizadas, iii) objetivo da MD e o iv) tipo de dado**. Acrescentou-se, nesta pesquisa, abordagem sobre **v) educação especial**, não encontrada em nenhuma das pesquisas de PB. Portanto, esta pesquisa propõe abordar todas essas cinco abordagens levantadas por meio de MSL.

3. Metodologia

Esta pesquisa de revisão bibliografia fez o uso do Mapeamento Sistemático da Literatura (MSL). Segundo [Kitchenham 2004], MSL é um método para identificar, interpretar e sumarizar os trabalhos relevantes para determinada linha de pesquisa, área ou fenômeno de interesse de forma não tendenciosa e replicável. O MSL é elaborado seguindo três etapas: **planejamento, execução e sumarização** (ou resultados).

3.1. Planejamento

O planejamento visa elaborar um protocolo com as informações e procedimentos de como será feito o levantamento dos trabalhos relacionado. Dentro do planejamento são desenvolvidos os itens descritos na Tabela 1 a seguir.

Tabela 1. Itens do planejamento

Item	Descrição do item
Questões de pesquisa (QP)	São as perguntas que serão respondidas com o resultado do MSL
<i>Strings</i> de busca	São as palavras chaves relacionadas ao tema
Bases de dados	Repositórios de trabalhos científicos
CrITÉrios de inclusão (CI) e critérios de exclusão (CE)	CI: trabalho deve abordar MDE no período de 2015 a 2019; CE: negação de todo CIs
Formulários de extração	são as questões de pesquisa descritos na Tabela?

As questões de pesquisas são:

(QP1): Que algoritmos são utilizados?

(QP2): Qual é a ferramenta utilizada para realizar a MDE?

(QP3): Qual o objetivo da MDE?

(QP4): Qual o tipo de dado utilizado na MDE?

(QP5): O trabalho faz abordagem sobre educação especial no contexto de MDE?

As palavras chaves desta pesquisa são *mineração de dados e educação*. Em virtude disso, as *strings* de buscas são as mesmas. Porém, para incluir o máximo de trabalhos relevantes, a consulta foi sintetizada apenas para o termo: *mineração**.

As bases de dados estão disponíveis no portal de publicação da [Comissão Especial de Informática na Educação \(CEIE\)](#) e são estas:

Revista Brasileira de Informática na Educação (RBIE),

Simpósio Brasileiro de Informática na Educação (SBIE),

Workshop do Congresso Brasileiro de Informática na Educação (WCBIE) e

Workshop de Informática na Escola (WIE).

3.2. Execução

Logo após elaborar o Planejamento, descrito na subseção 3.1, a próxima etapa é a **Execução**. Nesta fase há os procedimentos de **seleção de trabalho** e de **extração de dados**.

Seleção: esse procedimento é iniciado com a submissão de *strings* nas bases de dados. Foi encontrado a quantidade de trabalhos: i) RBIE - 9 artigos, ii) SBIE com 46 trabalhos, iii) WCBIE com 29 pesquisas e iv) WIE com apenas 2. O total foram 87 trabalhos. Em seguida, selecionou-se os artigos relevantes e condizentes com a área utilizando os critérios de inclusão e exclusão. O resultado foi **59 artigos selecionados**.

Além disso, foi nesta pesquisa uma busca em profundidade, nos 59 artigos selecionados. Essa busca consiste em coletar trabalhos relevantes citados por esses artigos selecionados. Assim, encontrou-se mais **12 artigos** relevantes sobre MDE. A Figura 1 ilustra um diagrama simplificado da metodologia aplicada.

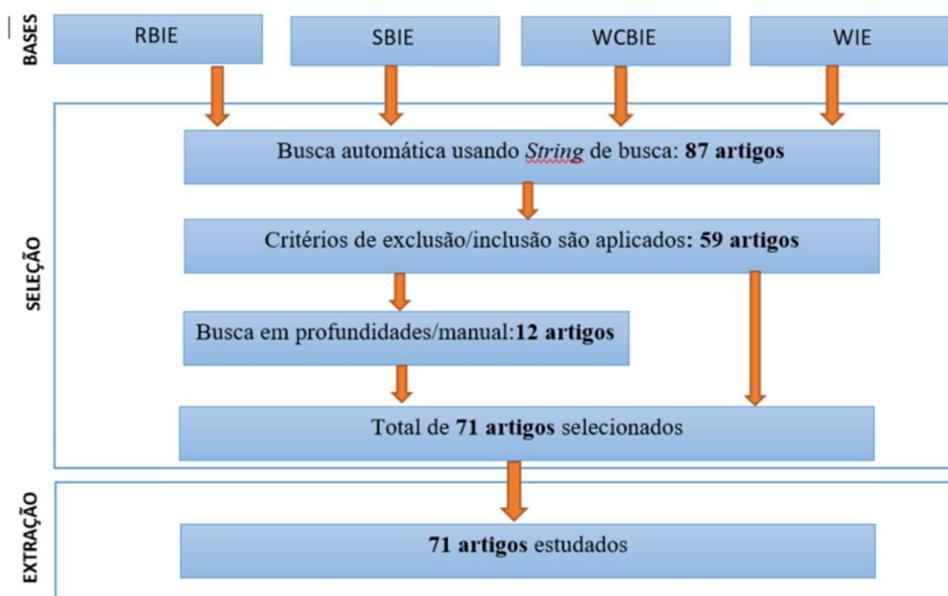


Figura 1. Diagrama da metodologia aplicada

Por conseguinte, realizou-se a **extração de dados** nos 71 artigos. Esse é o procedimento de responder às Questões de Pesquisa (PQ) por meio dos formulários de extração.

4. Sumarização: resultados

A Sumarização é a etapa em que apresenta os resultados. Estes serão apresentados a seguir por meio de gráficos e tabelas. Pela Figura 2, visualiza-se a quantidade de trabalhos sobre MDE. Limitou-se o tempo de 2015 a 2019. Examina-se que o ano de 2017 foi o mais produtivo sobre trabalhos de MDE, com 21 artigos. Salienta-se, na mesma imagem, uma linha contínua que demonstra uma variação do quantitativo por ano. Com efeito, a base de dados com mais artigos é a SBIE, com 32 artigos.

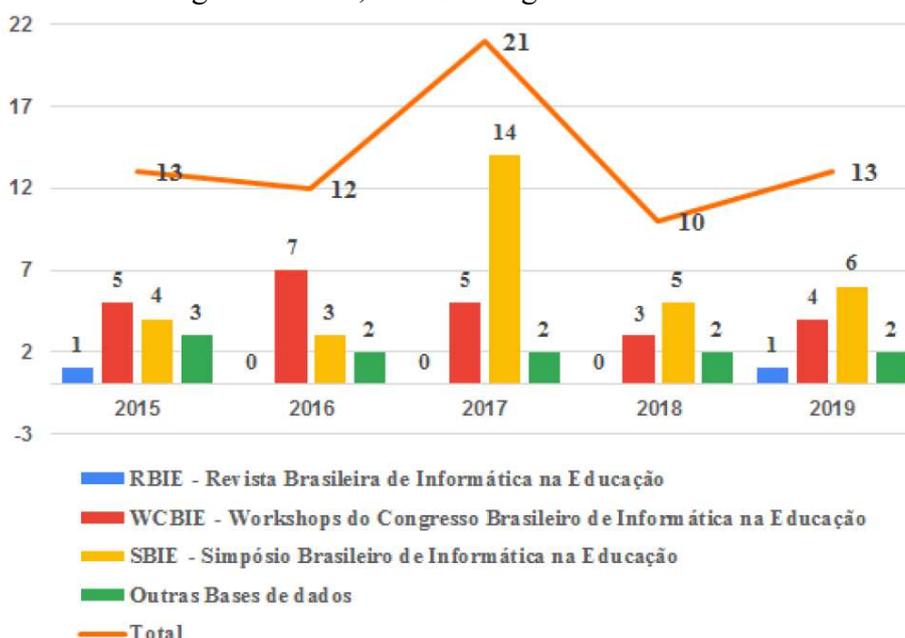


Figura 2. Quantitativo de artigos por anos e suas respectivas bases de dados

As respostas para as questões de pesquisas estão a seguir.

QP1: Quais algoritmos são utilizados?

A resposta, visível na Figura 3 é que o algoritmo *J48* é o mais empregado nos trabalhos científicos de MDE, utilizado em 22 trabalhos. Isso corrobora sobre sua eficiência. No entanto, ressalta-se que 14 artigos estão classificados como **não citado** e 9 como **não utilizado** sobre os algoritmos. Isso exprime que muitos trabalhos não se preocupam em apresentar o algoritmo utilizado para a MD.

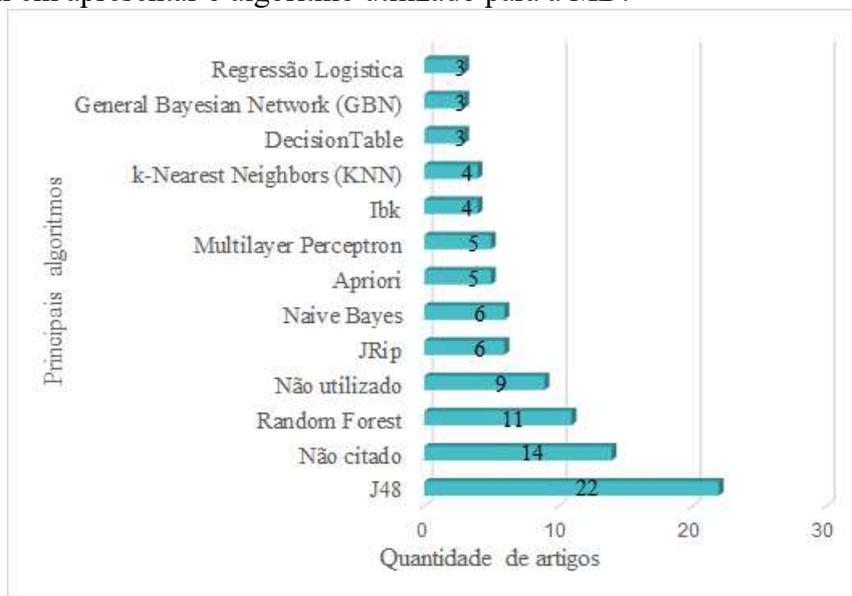


Figura 3. Quantitativo de principais algoritmos utilizados

QP2: Qual é a ferramenta utilizada para realizar a MDE?

Outra classificação relevante são as ferramentas utilizadas. A Tabela 2 enumera o quantitativo desta pesquisa. Assim, o instrumento mais empregado é a *Weka*, com 24 trabalhos. Revela-se, também, que em 11 artigos não citaram uma ferramenta. Até certo ponto, esse dado não favorece para este levantamento.

Tabela 2. Quantitativo de ferramentas utilizadas

Ferramentas	Quantidade
Weka	24
Nao citado	11
Não utilizado	9
Linguagem Python/Anaconda	6
Proposta de Ferramenta	5
RapidMiner	4
Ferramenta estatística R	2
R Studio	2
R Studio e R Project	2
SPSS Modeler	2
Ferramenta de aprendizagem	2

QP3: Qual o objetivo da MDE?

Revela-se, pela Figura 4, que a **mineração de dados no contexto acadêmico** é o objetivo mais utilizado, com 20 artigos. Esse tipo de pesquisa inclui, por exemplo, dados de matrículas, de cursos, de notas, de frequência bem como dados sócio econômico. Logo em seguida, as propostas mais utilizadas são MD abertos educacionais e comportamento com 11 trabalhos. O primeiro, envolve investigação com dados disponíveis pelo MEC como IDEB, ENEM entre outros; o segundo, trata sobre aspecto comportamental do aluno.

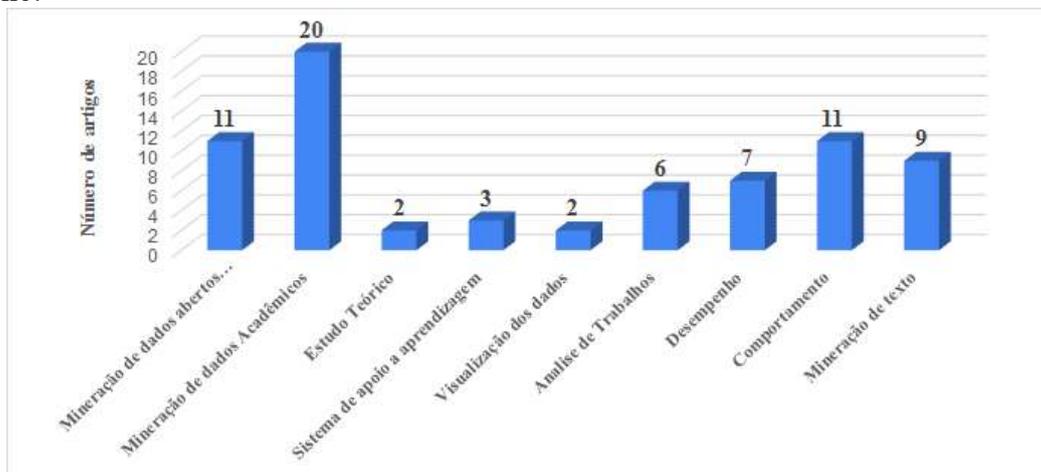


Figura 4. Quantitativo de artigos em relação aos objetivos

Em contraste com o que foi citado, o objetivo sobre **proposta de sistemas de apoio a aprendizagem** mostra-se em quantidade limitada de trabalhos, com 3 trabalhos apenas. Isso indica uma vertente em potencial para pesquisas futuras.

QP4: Qual o tipo de dado utilizado na MDE?

O produto exposto na Figura 5 apresenta os tipos de dados. O resultado é harmônico com o apontado na QP2 (os objetivos da MDE). Naquela Figura 4, ilustra um número de trabalhos com **MD acadêmicos**, 20 artigos, que é o mesmo valor para o tipo de dado **dados acadêmicos de Universidade**. Em proporção similar, o **tipo de dado texto** apresenta mesmo em mesmo valor, com 9 trabalhos, o objetivo **mineração de texto**.

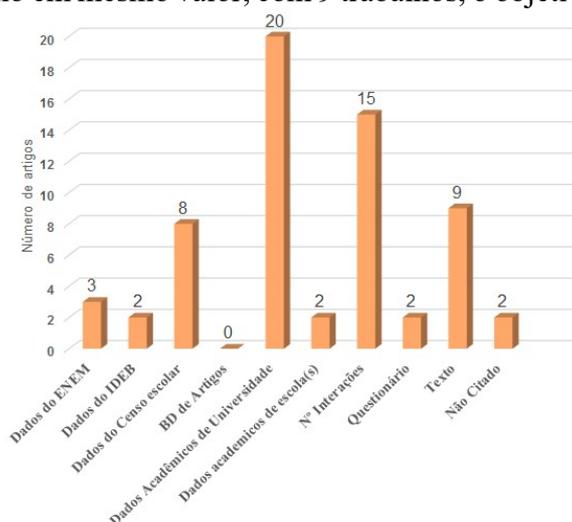


Figura 5. Número de trabalhos classificado pelo tipo de dado utilizado

QP5: O trabalho faz abordagem sobre educação especial no contexto de MDE?

Expõe-se, no gráfico da Figura 6, que há um reduzido número de artigos de MDE no contexto em educação especial. Com apenas um trabalho, visualiza-se um cenário que apresenta uma lacuna para estudos futuros.

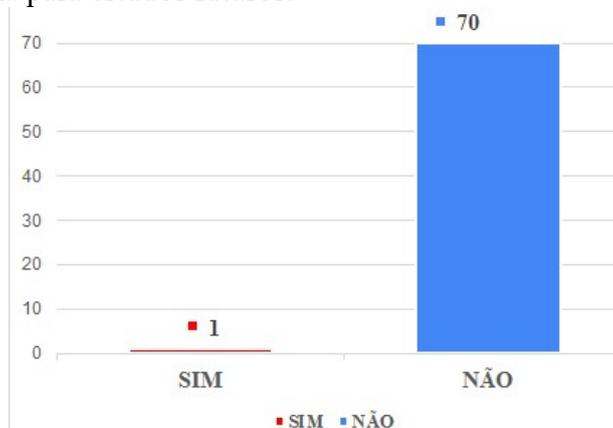


Figura 6. Quantitativo de artigos que aborda sobre educação especial

Além das respostas para as cinco QP's, a pesquisa apresenta outros levantamentos auxiliares. O conjunto completo está no [link da planilha](#).

5. Conclusão

Esta pesquisa apresentou um levantamento bibliográfico de pesquisas sobre mineração de dados na educação. O método proposto foi um mapeamento sistemático da literatura de trabalhos relacionados ao assunto. Com isso alcançou os objetivos apresentando os elementos que classificaram e sistematizaram os trabalhos sobre MDE.

Os principais resultados identificados foram: *i)* O algoritmo mais aplicado é o *J48*; *ii)* a ferramenta mais utilizada é a *weka*; *iii)* os objetivos, verificou-se número relevante de trabalhos de MD aplicado em ambiente acadêmico; *iv)* tipos de dados é condizente com a resposta dos resultados dos objetivos e *v)* inexpressivo número de trabalhos na vertente de educação especial, que aponta para uma potencial oportunidade de pesquisa.

Os resultados implicam mecanismos facilitadores para o desenvolvimento de novos trabalhos científicos. Um objeto de trabalho futuro desta pesquisa é expandir para base de dados internacionais. Acrescenta-se, ainda, que pôde-se explorar as lacunas de pesquisas citadas neste levantamento bibliográfico.

Referências

- [Baker et al. 2011] Baker, R., Isotani, S., and Carvalho, A. (2011). Mineração de Dados Educacionais: Oportunidades para o Brasil. *Revista Brasileira de Informática na Educação*, 19(02):3–13.
- [Baker 2014] Baker, R. S. (2014). Educational data mining: An advance for inteligente systems in education. *IEEE Intelligent Systems*, 29(3):78–82.
- [Dionísio et al. 2017] Dionísio, M., Ferreira, R., Cavalcanti, A., Carvalho, R., and Neto, S. (2017). Mineração de Texto Aplicada à Identificação de Colaboração em Fóruns Educacionais. *Anais do Simpósio Brasileiro de Informática na Educação (SBIE 2017)*.

- [Fayyad et al. 1996] Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, 17(3):37–37.
- [Kampff 2009] Kampff, A. J. C. (2009). Mineração de dados educacionais para geração de alertas em ambientes virtuais de aprendizagem como apoio à prática docente.
- [Kitchenham 2004] Kitchenham, B. (2004). Procedures for performing systematic reviews. Keele, UK, Keele University, 33(2004):1–26.
- [Maschio et al. 2018] Maschio, P., Vieira, M. A., Costa, N., Melo, S. D., and Júnior, C. P. (2018). Um Panorama acerca da Mineração de Dados Educacionais no Brasil. *Anais do XXIX Simpósio Brasileiro de Informática na Educação (SBIE 2018)*, 1(Cbie):1936.
- [Rabelo et al. 2017] Rabelo, H., Burlamaqui, A., Valentim, R., Rabelo, D. S. d. S., and Medeiros, S. (2017). Utilização de técnicas de mineração de dados educacionais para predição de desempenho de alunos de EaD em ambientes virtuais de aprendizagem. *Anais do XXVIII Simpósio Brasileiro de Informática na Educação (SBIE 2017)*, 1.
- [Rodrigues et al. 2014] Rodrigues, R. L., Ramos, J. L. C., Silva, J. C. S., and Gomes, A. S. (2014). A literatura brasileira sobre mineração de dados educacionais. In *Anais dos Workshops do Congresso Brasileiro de Informática na Educação*, volume 3, page 621.
- [Santos et al. 2017] Santos, P., Ferreira, R., and Miranda, P. (2017). Dados Abertos Educacionais: Uma Revisão da Literatura Brasileira. *Anais do XXVIII Simpósio Brasileiro de Informática na Educação (SBIE 2017)*, 1(Cbie):11.
- [Santos et al. 2016] Santos, R., Pitangui, C., Vivas, A., and Assis, L. (2016). Análise de Trabalhos Sobre a Aplicação de Técnicas de Mineração de Dados Educacionais na Previsão de Desempenho Acadêmico. *Anais dos Workshops do V Congresso Brasileiro de Informática na Educação (CBIE 2016)*, 1(Cbie):960.