Dados Educacionais Abertos: associações em dados dos inscritos do Exame Nacional do Ensino Médio

Tancicleide C. S. Gomes[1], Roberta M. M. Gouveia[2], Conceição Moraes Batista[2]
[1]Centro de Informática – Universidade Federal de Pernambuco (UFPE)
Recife – PE - Brasil
[2]Departamento de Estatística e Informática – Universidade Federal Rural de Pernambuco (UFRPE)
Recife – PE - Brasil

tancicleide.gomes@gmail.com, {robertammg, cecamoraes}@gmail.com

Abstract: *Educational data mining provides managers and educators with inputs for decision-making as well as perspectives to mitigate the challenges of the teaching-learning process. In this way, this work seeks to apply data mining techniques and methods in order to discover association rules in educational statistical data from the National High School Examination (ENEM) in the Northeast region. The results obtained may be interesting to support the proposal of public policies and initiatives aimed at the entry and permanence of young people in higher education.*

Resumo. *A mineração de dados educacionais oferece aos gestores e educadores subsídios para a tomada de decisão, bem como perspectivas para mitigar desafios do processo de ensino-aprendizagem. Deste modo, este trabalho busca aplicar técnicas e métodos de mineração de dados no intuito de descobrir regras de associação em dados estatísticos educacionais oriundos do Exame Nacional do Ensino Médio (ENEM) no âmbito da região Nordeste. Os resultados obtidos podem ser interessantes para apoiar proposição de políticas públicas e iniciativas visando o ingresso e a permanência dos jovens no ensino superior.*

1. Introdução

A mineração de dados (MD) tem o propósito de filtrar e correlacionar dados e oferecer informação. Ferramentas, técnicas de mineração e análise de dados, anteriormente confinados apenas em sofisticados laboratórios de investigação agora têm sido utilizados em diversos ambientes [ibidem, USDE 2012]. O uso da mineração de dados na educação é denominada mineração de dados educacionais (MDE). Na MDE, os métodos e técnicas de estatística, aprendizado de máquina e banco de dados são utilizados para analisar os dados coletados durante o processo de aprendizagem, agregando métodos de computação para compreender como os estudantes aprendem, por exemplo.

Muitas instituições de ensino passaram a adotar a análise de dados para auxiliar na tomada de decisão. Estes padrões podem ser construídos em modelos de mineração de dados e usados para prever comportamentos individuais com precisão para auxiliar na alocação de recursos e pessoal de maneira mais eficaz [USDE 20112].

Sistemas tutores inteligentes, simulações, jogos, sistemas adaptativos, entre outros,

DOI: 10.5753/cbie.wie.2017.895

permitem a coleta e análise de dados dos estudantes para identificar padrões e tendências, possibilitando novas descobertas e testes de hipóteses sobre como eles aprendem. Os dados gerados por sistemas como estes possuem diversas variáveis que os algoritmos de mineração podem explorar na construção de novos modelos. Com olhares individualizados e voltados para os dados de progresso de cada aluno, os educadores podem prever o desempenho dos alunos e desenvolver estratégias para mantê-los no caminho adequado [Luan 2007, USDE 2012, West 2012].

Neste contexto, o presente trabalho buscou identificar regras de associação em dados do questionário socioeconômico (QSE) dos inscritos no Exame Nacional do Ensino Médio (ENEM) no ano de 2013. O objetivo foi encontrar possíveis relações entre o desempenho na prova de Matemática do Exame e o local de residência do estudante (INTERIOR/ CAPITAL), a renda familiar, a escola que o estudante cursou Ensino Fundamental e Médio. Os resultados preliminarmente obtidos sugerem relações entre o desempenho dos candidatos e a renda familiar, sobretudo candidatos provenientes de escolas públicas. Este trabalho está organizado da seguinte maneira: os construtos teóricos que alicerçam este trabalho são apresentados na seção 2. Os principais trabalhos relacionados são apresentados na seção 3, e os métodos empregados na seção 4. Os resultados obtidos são relatados e discutidos na seção 5. Por fim, na seção 6 são descritas as considerações finais.

2. Bases Conceituais

Segundo Rodrigues *et al.* (2014), a mineração de dados pode ser descrita como o processo automatizado de descoberta de informações a partir de grandes volumes de dados de um contexto previamente definido, com o objetivo de apoiar a tomada de decisão. O processo de descoberta de padrões inicia-se pela escolha dos dados. Os dados são integrados e pré-processados para que sejam estruturados, limpos, selecionados e padronizados para a tarefa de mineração de dados. Na tarefa de mineração aplicam-se técnicas inteligentes que possibilitam o encontro de soluções que auxiliam os especialistas na descoberta de respostas. Os resultados desta tarefa devem ser pós-processados para se apresentem análises qualitativas/quantitativas dos elementos encontrados e, quando possível, apresentados de maneira que possam ser interpretadas para auxiliar na tomada de decisão (Rodrigues *et al.* 2014).

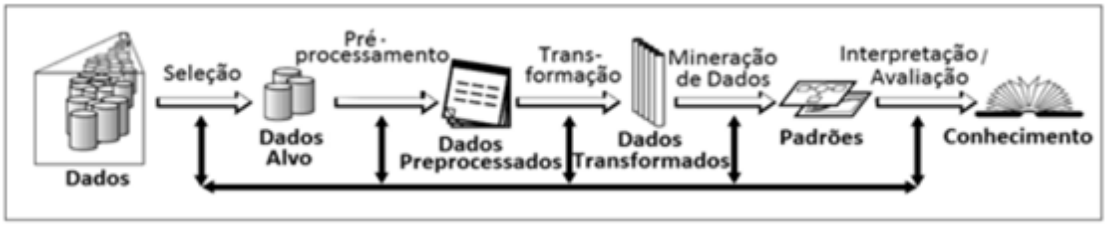

Figura 1. Processo de KDD

Este processo é denominado Descoberta de Conhecimento em Bases de Dados (do inglês *Knowledge Discovery in Databases* - KDD), sendo referenciado apenas por KDD ao longo deste trabalho. O KDD abrange diversas fases que compõem o caminho que os dados percorrem até tornarem-se conhecimento útil. Segundo Fayadd (1996) citado por Camilo e Silva (2009), o KDD é

um processo de identificação de padrões válidos, geralmente desconhecidos, potencialmente uteis e interpretáveis (Figura 1).

Nos contextos educacionais, a MD oferece subsídios para uma melhor compreensão de diversos aspectos, tais como respostas a perguntas do tipo: (1) *Como os alunos aprendem?* (2) *Qual o papel do contexto da aprendizagem?* (3) *Quais fatores influenciam a aprendizagem?* - dentre outras questões que permitem ensaiar estratégias e métodos que promovam melhorias no processo de ensino-aprendizagem. Deste modo, torna-se possível encontrar indicadores sobre a motivação do aluno, oferecer personalização do ambiente e dos métodos de ensino proporcionados no intuito de ofertar melhores condições de aprendizagem.

Embora recebam menos destaque que as outras bases mais comuns, os dados abertos possuem uma ampla gama de informações relevantes para a tomada de decisão e para a criação de políticas públicas, por exemplo. Na subseção a seguir é apresentada uma breve explanação sobre o que são dados abertos, como eles podem ser classificados e como eles podem ser fontes relevantes para a aplicação de mineração e análise de dados.

2.1. Dados abertos

Os dados abertos são importantes bases de dados para aplicação de técnicas de mineração, consistindo em uma iniciativa amplamente fomentada pela *Open Knowledge Foundation* (OPK)[1]. São considerados dados abertos os que são provenientes de fontes diversas, tais como[2]: (1) Cultura, (2) Ciência, (3) Finanças, (4) Estatísticas, (5) Tempo, (6) Ambiente, (7) Transporte. Segundo a OPK, o conceito de dados abertos consiste na possibilidade de promover a transparência, a participação da sociedade e o engajamento social. Além de agregar valor social e comercial de tal modo que os dados possam ser livremente utilizados, modificados e compartilhados por quaisquer pessoais com qualquer finalidade[3].

Neste contexto, todos os anos, o Índice Global de Dados Abertos avalia de maneira independente a abertura de dados de diversos locais e estabelece um ranking. São considerados dados como: horários dos meios de transporte, orçamento governamental, resultados eleitorais, mapas nacionais, estatísticas nacionais, emissão de poluentes e outros. Em 2014[4], o Brasil caiu duas posições neste ranking (26º posição) em relação a 2013 (24º posição). Este é um aspecto interessante porque o Brasil avançou de 48% para 54% na abertura de dados abertos, o que demonstra que este avanço não foi expressivo e assevera que a disponibilização dos dados de qualquer forma não é suficiente.

3. Trabalhos Relacionados

No panorama nacional, os esforços em pesquisa relacionados à temática de mineração de dados educacionais são crescentes ao longo dos últimos anos, evoluindo de um tímido cenário com menos de 20 trabalhos relatados nos principais periódicos e conferências de 2006 a 2010 [Rodrigues et al. 2014], para aproximadamente 50 trabalhos publicados[5] apenas nos anais do

[1] https://okfn.org/
[2] https://okfn.org/opendata
[3] http://opendefinition.org/re
[4] http://index.okfn.org/dataset/
[5] Estes dados foram obtidos através de uma revisão introdutória da literatura realizada pela própria autora nos anais do CBIE nos anos de 2014, 2015 e 2016.

Congresso Brasileiro de Informática na Educação (CBIE) nos últimos quatro anos.

Em uma recente revisão da literatura que abrangeu 68 artigos envolvendo MDE publicados nas principais conferências e periódicos do país, Rodrigues *et al.* (2014) relatam que os trabalhos concentram-se nos seguintes temas: (1) mineração de dados, (2) mineração de textos e (3) visualização de dados. Comumente, os trabalhos têm seus dados oriundos de diversos contextos educacionais[1]:

- *ambientes virtuais de aprendizagem* (Herpich *et al.*, 2016; Queiroga, Cechinel e Araújo, 2015; Santos, Bercht e Wives, 2015; Silva, L. *et al.*, 2015; Silva, R. *et al.*, 2015);
- *instituições de ensino* (Bernardini, Costa e Artigas, 2015; Santos *et* al. 2016; Silva e Nunes, 2015);
- *dados estatísticos públicos* (Ferreira, 2015; Silva, Morino e Sato, 2014).

No ensejo de dados abertos - temática deste trabalho-, Silva, Morino e Sato (2014) relatam uma experiência de mineração de dados do ENEM do ano de 2010, buscando analisar relações de causa e efeito entre o relacionamento do desempenho no ENEM e fatores socioeconômicos com dados de alunos apenas das capitais da região Sudeste. Os resultados obtidos através do conhecimento extraído permitiram observar que a renda familiar baixa, a escolaridade dos pais de nível primário e uma grande quantidade de pessoas que residem com os estudantes são aspectos que contribuem para o baixo desempenho do aluno.

Ferreira (2015), por sua vez, utilizando os dados do Censo Escolar da Educação Básica fornecido pelo Instituto Nacional de Estudos e Pesquisas (INEP), descreve uma experiência de mineração utilizando árvores de decisão com o objetivo de identificar os fatores relacionados à conclusão do ensino fundamental. Os resultados demonstraram que recursos como internet banda larga, laboratório de ciências, auditório na escola e ensino privado estão associados à maior chance do aluno concluir o ensino fundamental, bem como evidenciou que necessidades especiais estão estreitamente relacionadas à não conclusão do ensino fundamental.

Namen e Soares (2011) relatam a aplicação do processo de mineração de dados para a identificação de associações entre variáveis relacionadas ao ensino de Língua Portuguesa para alunos do 5º ano do ensino fundamental do estado do Rio de Janeiro, utilizando os microdados da Prova Brasil do ano de 2007 disponibilizados pelo INEP. A partir dos resultados obtidos notou-se que alguns fatores como: falta de incentivo dos pais, reprovação prévia do aluno e atuação do aluno em trabalho doméstico e/ou fora de casa, entre outros, exerceram influência negativa sobre o aprendizado do estudante.

Mediante o exposto, nota-se a diversidade de aplicações de mineração de dados em contextos educacionais. Este trabalho se aproxima do trabalho de Silva, Morino e Sato (2014), ao usar dados abertos educacionais. No entanto, tem como principal contribuição uma visão local (sobre o estado de Pernambuco e da região Nordeste) sobre as possíveis relações entre o desempenho na prova de Matemática e aspectos como local de residência do candidato, renda familiar e outros de inscritos no ENEM nos anos de 2013 e 2014.

4. Mineração dos Dados

Nesta seção apresentamos os processo aplicado: seleção dos dados, tecnologias aplicadas e técnicas utilizadas, bem como uma visão geral de todas as fases do processo de descoberta de

conhecimento. Na etapa de pré-processamento são realizadas as fases de seleção, limpeza e transformação dos dados. Os dados do QSE do ENEM foram organizados em quatro bases distintas considerando a região Nordeste e apenas os dados provenientes do estado de Pernambuco nos anos de 2013 e 2014: (1) PE 2013, (2) PE 2014, (3) NE 2013, (4) NE 2014. O objetivo foi trazer as discussões para um contexto mais regional e local, assim optou-se por atuar com os dados da região Nordeste e do estado de Pernambuco, sem redução de dimensionalidade.

Foram identificados e selecionados os atributos mais relevantes para encontrar padrões condizentes com os objetivos deste trabalho. Os demais valores não pertencentes ao domínio foram removidos. Depois da limpeza e transformação dos dados, a base ficou constituída por um total de 24 atributos que abrangem: dados pessoais do inscrito, como idade, município de residência; dados advindos do questionário socioeconômico; as notas obtidas em cada uma das quatro provas (Ciências Humanas e suas tecnologias, Linguagens, códigos e suas tecnologias, Ciências da Natureza e suas tecnologias, Matemática e suas tecnologias) e na prova de redação; e o ano de realização do exame.

Na fase de mineração, foi utilizado o software livre *Weka*. Esta ferramenta foi escolhida por dispor de vários algoritmos do aprendizado de máquina supervisionado e não supervisionado, que contemplam tanto métodos preditivos quanto métodos descritivos, além de possuir uma interface gráfica de fácil utilização. No intuito de descobrir relacionamentos dentro da base de dados foi utilizado o algoritmo *Apriori* por ser amplamente utilizado em diversos trabalhos correlatos descritos nas seções anteriores.

5. Resultados e Discussões

Conforme anteriormente mencionado, os dados foram organizados em quatro bases distintas: (1) PE 2013, (2) PE 2014, (3)NE 2013, (4) NE 2014. Buscou-se encontrar associações nos dados, considerando apenas o desempenho, posteriormente buscamos encontrar associações considerando o desempenho e o gênero do candidato. Finalmente, buscamos encontrar padrões na redações com notas zero, considerando as redações de todas as regiões do país no ano de 2014. Em todas as análises dos dados foi utilizado o algoritmo J48, que contempla o método descritivo de análise de dados e faz parte do aprendizado de máquina não supervisionado.

5.1. Análise de desempenho utilizando o algoritmo *Apriori*

Os parâmetros aplicados foram: suporte mínimo de 10%, parâmetro de confiança de 80% e parâmetro de quantidade de regras, 30. Esta análise foi realizada na base de dados ENEM NE 2014, contemplando 23 atributos. Entre as regras de associação geradas, destacaram-se as seguintes:

- Inscritos que realizaram o ensino fundamental somente em escola pública apresentam forte tendência de terem realizado o ensino médio apenas em escolas públicas (grau de confiança de 96%);
- Inscritos que cursaram o ensino médio apenas em escolas públicas apresentam forte tendência de serem oriundos de famílias de classe E, ou seja, com renda familiar de até dois salários mínimos (grau de confiança de 94%);
- Inscritos que indicaram ter realizado seus estudos na modalidade regular têm forte

tendência de residirem na zona urbana (grau de confiança de 83%);
- Inscritos oriundos do interior têm forte tendência de pertenceram a famílias de classe E (grau de confiança de 83%).

Com objetivo de comparar as regras de associação obtidas no cenário regional com o cenário estadual, optamos por aplicar o mesmo algoritmo e parâmetro de confiança sobre os dados da base ENEM PE 2014. Foram encontradas regras muito similares que alcançaram suporte mínimo de 30%, demonstrando que o estado não está em situação diferenciada do restante da região:

- Inscritos pertencentes a famílias de classe E que concluíram o ensino fundamental somente em escola pública têm forte tendência de terem concluído o ensino médio também apenas em escola pública (grau de confiança de 97%);
- Inscritos que cursaram o ensino fundamental somente em escolas públicas tem forte tendência de terem cursado o ensino médio também apenas em escolas públicas (grau de confiança de 96%);

Com base nos resultados encontrados, buscamos verificar se existiria alguma associação entre a escola cursada pelo indivíduo durante o ensino médio, as notas obtidas nas quatro provas do ENEM e a renda familiar. Foi utilizada inicialmente a base de dados ENEM NE 2014. Não consideramos o desempenho na nota de redação porque o seu cálculo é feito de maneira diferenciada. Foram encontradas as seguintes regras:

- Inscritos que concluíram o Ensino Médio em escolas públicas apresentam forte tendência de pertencerem a famílias de classe E (suporte de 40%, confiança de 84%);
- Inscritos com nota entre 400 e 500 pontos em Ciências da Natureza têm forte tendência de terem renda de até 2 salários mínimos (suporte de 30%, confiança de 83%);
- Inscritos com nota entre 400 e 500 pontos em Matemática apresentam forte tendência de pertencerem a famílias de classe E (suporte de 20%, confiança de 82%);
- Inscritos com nota entre 400 e 500 pontos em Ciências Humanas que cursaram ensino médio apenas em escolas públicas apresentam forte tendência de pertencerem a famílias de classe E (suporte de 10%, confiança de 86%).

Foram aplicados mesmos parâmetros na base de dados ENEM PE 2014 e os resultados obtidos não se distinguiram do cenário regional:

- Inscritos que concluíram o ensino médio somente em escola pública possuem forte tendência de pertencerem a famílias de renda familiar de até dois salários mínimos (suporte de 30%, confiança de 82%);

Observando-se as regras de associação encontradas, nota-se diversas associações entre notas de até 500 pontos com a renda familiar baixa (classe E – até dois salários mínimos), bem como associações com a conclusão do ensino médio somente em escola pública como indicador dos inscritos serem oriundos de família de classe E. Deste modo, buscamos verificar se haveriam associações entre o desempenho apenas na prova de Matemática, a renda familiar e o local de residência (se capital ou interior). O local de residência foi levado em consideração, pois embora no cenário regional as famílias com baixa renda estejam concentradas no interior do estado, em

Pernambuco esta distribuição ocorre tanto na capital e regiões metropolitanas quanto em cidades do interior. Complementarmente, consideramos a nota do exame de Matemática por ser a disciplina com a maior quantidade de notas baixas. Foram encontradas cinco regras na base ENEM NE 2014, dentre as quais se destacam:

- Inscritos residentes no interior têm forte tendência de pertencerem a famílias com renda de até 2 salários mínimos (83% de confiança, suporte de 40%);
- Inscritos com notas variando entre 400 e 500 pontos no ENEM têm forte tendência de pertencerem a famílias de classe E (82% de confiança, suporte de 20%);
- Inscritos com notas variando entre 300 e 400 pontos no ENEM têm forte tendência de pertencerem a famílias de classe E (87% de confiança, suporte de 10%);

Na base ENEM PE 2014, foram encontradas regras muito similares:

- Inscritos com notas variando entre 300 e 400 pontos no ENEM têm forte tendência de pertencerem a famílias de classe E (84% de confiança, suporte de 10%);

Convém mencionar que devido a método psicométrico de correção das provas, a Teoria de Resposta ao Item, nenhum inscrito recebe nota zero, mas sim uma nota mínima que varia todos os anos. As notas mínimas, médias e máximas do ano de 2014 são apresentadas na Tabela 1.

Tabela 1. Proficiência dos participantes do ENEM 2014: provas objetivas

	Nota mínima	Nota máxima	Nota média
Ciências Humanas (CH)	324,8	862,1	546,5
Ciências da Natureza (CN)	330,6	876,4	482,2
Linguagens e Códigos (LC)	306,2	814,2	507,9
Matemática (MT)	318,5	973,6	473,5

Embora o NE possua a segunda maior quantidade de participantes, perdendo apenas para região SE, o desempenho médio dos inscritos do NE foi inferior à média nacional em todas as áreas no ano de 2014 (Tabela 2). Outro aspecto observado refere-se a renda dos participantes: aproximadamente 82% dos inscritos oriundos do NE (2.014.143 pessoas) pertencem a classe E, ou seja, declararam possuir renda até 2 salários mínimos, dentre os quais 85.810 participantes afirmaram não possuir qualquer renda. No cenário de Pernambuco, cerca de 80% (342.788) dos inscritos (432.966 pessoas) declararam pertencer a classe E, dentre os quais 12.413 pessoas afirmaram não possuir qualquer fonte de renda. Em uma análise mais detalhada, focando apenas no cenário de Pernambuco e observando apenas notas não nulas de Matemática (304.955), nota-se que aproximadamente 70% (215.062) das notas estão na faixa entre 300 e 500

pontos, a partir do que é possível observar que:

- 30% (90.999) de todos participantes obtiveram notas entre 300 e 400 pontos, o que é abaixo da nota média nacional (473,5) e até mesmo da média regional (442,7).
    - 88% (80.301) são pertencentes a famílias de classe E, isto representa aproximadamente 23,4% de todo o total de inscritos que declararam ser pertencentes a famílias de classe E;
    - 46% (42.485) declararam terem cursado todo ele ou a maior parte do ensino médio em escolas públicas;
    - 41% (38.003) são oriundos de famílias de classe E e declararam terem cursado o ensino em escolas públicas ou a maior parte dele.
- 40% (124.063) de todos participantes obtiveram notas entre 400 e 500 pontos.
    - 82% (102.521) são pertencentes a famílias de classe E, isto representa aproximadamente 30% de todos os inscritos que declararam renda de até dois salários mínimos;
    - 45% (55.869) declararam terem cursado todo o ensino médio ou a maior parte dele em escola pública;
    - 38% (47.989) são pertencentes a famílias de classe E e declararam terem cursado todo o ensino médio ou a maior parte dele em escola pública.

Tabela 2. Desempenho no ENEM 2014 por região

Região	Participantes	Ciências Humanas	Ciências da Natureza	Linguagens e códigos	Matemática	Redação
Centro-Oeste	8,4%	542,6	480,7	503,3	467,3	437,6
Nordeste	33,7%	533,9	471	495,9	456,1	434,9
Norte	10,9%	529,9	464,8	487,1	442,7	417,5
Sudeste	34,9%	561,2	495,8	523,7	496,5	486,9
Sul	11,9%	557,7	491,2	517,8	487,8	468,9
Média BR	100%	546,5	482,2	507,9	473,5	455,4

Mediante o exposto, constata-se que aproximadamente 53% (182.822) de todos os inscritos que declararam renda familiar de até 2 salários mínimos obtiveram desempenho de no máximo 500 pontos na prova de Matemática. Considerando que houve 215.062 notas entre 300 e 500 pontos, 85% das notas mais baixas de Matemática são de inscritos oriundos de famílias de classe E. O que assevera a forte relação entre a renda familiar e o desempenho dos estudantes.

Todas estas análises foram aplicadas nas bases ENEM NE 2013 e ENEM PE 2013, gerando regras de associação quando não idênticas, muito similares. Por este motivo os resultados obtidos nestas bases não foi minuciosamente abordado neste trabalho, por considerar-se que apenas confirmavam as descobertas encontradas.

6. Considerações Finais

Este trabalho buscou identificar regras de associação a partir dos dados socioeconômicos dos inscritos oriundos da região Nordeste no Exame Nacional do Ensino Médio (ENEM) nos anos de 2013 e 2014. O principal objetivo foi encontrar relações entre aspectos como a renda familiar, a escola em que os candidatos cursaram o Ensino Fundamental e Médio, e local de residência do estudante com o desempenho nas provas do Exame.

Os resultados preliminares obtidos endossam uma relação significativa entre a renda familiar dos candidatos e o desempenho deles nos Exames, especialmente os candidatos provenientes de escolas públicas. Outro aspecto observado foi a tendência do desempenho dos alunos de sexo feminino concentrar-se entre a nota mínima e média apenas. O presente trabalho tem como principal contribuição a oferta de subsídios para a criação e/ou fortalecimento de iniciativas que: (1) visem a diminuição da evasão dos estudantes recém-ingressos; (2) avaliação e diagnóstico do corpo discente; (3) auxiliar os gestores de instituições de ensino superior públicas na criação/ consolidação de ações afirmativas como cotas, programas de nivelamento, programas de bolsas de estudo e de apoio acadêmico.

Trabalhos futuros contemplam a integração dos dados da base nacional do ENEM com outras bases como o Censo Escolar e o Censo de Educação Superior, no intuito de identificar possíveis padrões e modelos que deem suporte à criação de estratégias para acolher o recém-ingresso na universidade.

Referências

Bernardini, F., Costa, J., e Artigas, D. Proposta de uma técnica de Mineração em Grafos para identificação de gargalos em currículos de graduação. In: Anais do Simpósio Brasileiro de Informática na Educação. 2015. p. 1062.

Camilo, C. O., & Silva, J. C. D. (2009). Mineração de dados: Conceitos, tarefas, métodos e ferramentas. Universidade Federal de Goiás (UFC), 1-29.

Fayyad, U., Piatetsky-Shapiro, G., e Smyth, P. (1996). From data mining to knowledge discovery in databases. AI magazine, v. 17, n. 3, p. 37.

Ferreira, G. Investigação acerca dos fatores determinantes para a conclusão do Ensino Fundamental utilizando Mineração de Dados Educacionais no Censo Escolar da Educação Básica do INEP 2014. In:Anais dos Workshops do Congresso Brasileiro de Informática na Educação. 2015. p. 1034.

Herpich, F., Nunes, F. B., Tarouco, L. M. R., e Cazella, S. Mineração de Dados Educacionais: uma análise sobre o Engajamento de Usuários em Mundos Virtuais. In: Anais dos Workshops do Congresso Brasileiro de Informática na Educação. 2016. p. 910.

Luan, J. (2007). Data Mining Applications in Higher Education. Namen, A. A., e Soares, A. (2011). Mineração de dados relacionados ao aprendizado de Língua Portuguesa: um estudo exploratório. Encontro de Modelagem Computacional, v. 14, p. 295- 304, 2011.

Queiroga, E., Cechinel, C., e Araújo, R. Um Estudo do Uso de Contagem de Interações Semanais para Predição Precoce de Evasão em Educação a Distância. In: Anais dos Workshops do Congresso Brasileiro de Informática na Educação. 2015. p. 1074.

Rodrigues, R. L., Ramos, J. L. C., Silva, J. C. S., e Gomes, A. S. A literatura brasileira sobre mineração de dados educacionais. In: Anais dos Workshops do Congresso Brasileiro de Informática na Educação. 2014. p. 621.

Santos, F. D., Bercht, M., e Wives, L. Classificação de alunos desanimados em um AVEA: uma proposta a partir da mineração de dados educacionais. In: Anais do Simpósio Brasileiro de Informática na Educação. 2015. p. 1052.

Santos, R., Pitangui, C., Vivas, A., e Assis, L. Análise de Trabalhos Sobre a Aplicação de Técnicas de Mineração de Dados Educacionais na Previsão de Desempenho Acadêmico. In: Anais dos Workshops do Congresso Brasileiro de Informática na Educação. 2016. p. 960.

Silva, J., e Nunes, I. Mineração de Dados Educacionais como apoio para a classificação de alunos do Ensino Médio. In: Anais do Simpósio Brasileiro de Informática na Educação. 2015. p. 1112.

Silva, L. A., Morino, A. H., e Sato, T. M. C. (2014). Prática de Mineração de Dados no Exame Nacional do Ensino Médio. In: Anais dos Workshops do Congresso Brasileiro de Informática na Educação. 2014. p. 651.

Silva, L. A., Trindade, D., de Paula, C. e Pinto, S. Mineração de Dados em publicações de Fóruns de Discussões do Moodle como geração de Indicadores para aprimoramento da Gestão Educacional. In: Anais dos Workshops do Congresso Brasileiro de Informática na Educação. 2015. p. 1084.

USDE. United States - Department of Education, Office of Educational Technology, Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief, Washington, D.C., 2012.

West, D. M. (2012). Big data for education: Data mining, data analytics, and web dashboards. Governance Studies at Brookings, 1-10.