

Levantamento e Análise de Obras Recentes Baseada em Mineração de Relações e Clusterização

Warley Leite Fernandes, Cristiano Grijó Pitangui, Alessandro Vivas Andrade,
Luciana Pereira de Assis

PPGED – Universidade Federal dos Vales do Jequitinhonha e Mucuri (UFVJM)
Rodovia MGT 367 - Km 583, nº 5000 - Alto da Jacuba - Diamantina/MG

warleylfernandes@gmail.com, pitangui.cristiano@gmail.com,
prof.alessandrovivas@gmail.com, lupassis@gmail.com

***Abstract.** This paper presents an analysis and survey of publications in the field of Educational Data Mining published by several researchers at events and journals in the period from 2011 to 2016. The purpose of this research is to identify and present what is being researched in performance prediction area using mining relationships and clustering. The work explored exhibit valuable fruit, which reinforce the importance of the search area, and point alternatives have not or little explored. As continuity of research, suggests the use of classification based on association rules in Virtual Learning Environments and the use of concept maps to represent the interactions in the forums.*

***Resumo.** Este trabalho apresenta uma análise e levantamento de publicações na área de Mineração de Dados Educacionais publicados por diversos pesquisadores em eventos e periódicos no período de 2011 a 2016. O objetivo desta investigação é identificar e apresentar o que está sendo pesquisado na área de predição de desempenho usando mineração de relações e clusterização. Os trabalhos explorados exibem frutos valiosos, que reforçam a importância da área de pesquisa, e apontam alternativas ainda não ou pouco exploradas. Como continuidade de pesquisa, sugere-se o uso de classificação baseada em regras de associação em Ambientes Virtuais de Aprendizagem e o uso de Mapas Conceituais para representar as interações nos fóruns.*

1. Introdução

Avanços recentes em diversas áreas da tecnologia possibilitaram um grande crescimento na capacidade de armazenar, coletar, transmitir e gerar dados digitais. O preço acessível dos dispositivos e do alcance a redes de computadores fez com que aumentasse em grande escala o número de usuários em sistemas computacionais de ensino.

Com a evolução da aplicação das Tecnologias da Informação e Comunicação (TIC) na Educação, os ambientes computacionais de ensino e os conteúdos digitais passaram a ser vistos como tecnologias essenciais à condução dos processos educacionais;

Imensos volumes de dados são gerados pela interação de usuário em Ambientes Virtuais de Aprendizagem (AVA), o que torna possível a obtenção de informações sobre o processo de interação dos estudantes com esses ambientes. Nesse sentido, existem pesquisas em Informática na Educação que realizam a análise desses dados com o

objetivo de descobrir conhecimentos que possam ajudar na resolução de problemas educacionais, como exemplo pode-se citar a evasão, reprovação e o desempenho de estudantes, que estão entre as maiores preocupações presentes na educação de forma geral.

De acordo com Romero e Ventura (2010), as informações armazenadas nos bancos de dados do AVA são diversas, podendo incluir simples registros de acesso e interação com o sistema, até dados com ricos significados semânticos. Porém, identificar quais informações são importantes e como manipulá-las de maneira significativa não é uma tarefa simples. Uma solução promissora para extrair informações das bases de dados dos AVA é o uso de técnicas de Mineração de Dados, denominada Mineração de dados Educacionais ou EDM (*Educational Data Mining*) quando aplicadas ao contexto educacional.

O presente trabalho apresenta uma revisão dos trabalhos que utilizam técnicas de Mineração de Relações e Clusterização em EDM, publicados nas principais conferências e periódicos da área de Informática na Educação. O objetivo é apresentar a origem, as técnicas, as ferramentas, o que se analisa, os algoritmos e as fontes dos dados das pesquisas realizadas na área de Predição de Desempenho Acadêmico. Como principal objetivo, a pesquisa aponta informações interessantes a respeito da área, identificando possibilidades pouco (ou ainda não) exploradas.

Este trabalho está organizado em 7 seções. Na próxima seção é apresentada uma breve explicação sobre EDM. A seção três apresenta a metodologia aplicada na busca dos trabalhos. A quarta seção apresenta um pequeno resumo sobre cada trabalho analisado. A quinta seção apresenta os resultados do estudo. A seção seis apresenta a interpretação dos resultados e contribuição adquirida. Por fim, a sétima seção apresenta as conclusões deste trabalho.

2. Mineração de Dados Educacionais

O grande crescimento das tecnologias culmina num avanço na capacidade das pessoas em gerar e coletar dados. Essa enorme quantidade de dados vem ultrapassando a capacidade humana de interpretar e compreender tanta informação. Tal fato é um dos motivos que influenciou o desenvolvimento do processo para a descoberta de conhecimento em Base de Dados.

Da Silva *et al.* (2003), afirmam que KDD ou Descoberta de Conhecimento em Bases de Dados, é o ramo da computação que utiliza ferramentas e técnicas computacionais com a finalidade de sistematizar o processo de extração de conhecimento útil de grandes volumes de dados.

A escolha da técnica a ser utilizada no processo de EDM, depende exclusivamente do tipo de tarefa de KDD a ser efetivamente realizada. Baker *et al.* (2011) apresenta a taxonomia das principais subáreas de pesquisa em EDM: Predição, Agrupamento, Mineração de Relações, Destilação de Dados e Descobrimto com Modelos.

Para Baker *et al.* (2011), na área de Agrupamento, o objetivo principal é encontrar dados que se agrupam naturalmente, classificando os dados em diferentes grupos e/ou categorias. Em Mineração de Relações, afirma que a meta é descobrir possíveis relações entre variáveis das Bases de dados. Na área de Predição, a meta é desenvolver modelos

que deduzam aspectos específicos dos dados, conhecidos como variáveis preditivas, através da análise e fusão dos diversos aspectos encontrados nos dados, chamados de variáveis preditoras. A área de Destilação tem como objetivo apresentar dados complexos de forma a facilitar sua compreensão e expor suas características mais importantes.

Por sua vez, em Descoberta com Modelos, parte-se de um modelo gerado por um método de predição, tal como classificação, ou por um método de agrupamento, ou ainda manualmente, por meio de Engenharia de Conhecimento. Em seguida, esse modelo é utilizado como componente, ou ponto de partida, em outra análise com técnicas de Predição ou Mineração de Relações (Costa *et al.*, 2012).

Neste estudo são apresentadas as principais pesquisas na área de Mineração de Relações e Clusterização (agrupamento), que têm apresentado excelentes resultados relacionados à predição e desempenho de estudantes, contribuindo em diversos pontos nos ambientes educacionais, como: auxílio nas tomadas de decisão, identificação de alunos com dificuldades de aprendizagem, identificação de habilidades cognitivas, desempenho, e perfis de comportamento e aprendizagem.

3. Materiais e Métodos

Esta pesquisa abrange as publicações ocorridas entre 2011 e 2016 por diversos pesquisadores em eventos e periódicos. A busca foi realizada através do *Google Scholar* e em anais dos principais eventos e periódicos de Informática na Educação.

4. Trabalhos Pesquisados

Neto *et al.* (2016) objetivam separar grupos de estudantes com dificuldades de aprendizagem semelhantes. Foi confirmado que a aplicação de técnicas de agrupamento é bastante útil para a formação de grupos homogêneos de alunos, que uma vez identificados, permitem que o professor possa formular estratégias de ensino eficazes.

Coelho *et al.* (2015) objetivaram identificar e quantificar as incidências de termos relacionados às dificuldades de acesso ao AVA, nos chamados registrados nos canais de comunicação, no período de 2010 a 2013. Dessa forma, possibilitou-se uma melhor visualização e monitoramento do AVA e assim, a criação de ações direcionadas e mais efetivas com o intuito de favorecer a usabilidade e acessibilidade ao público.

Nunes *et al.* (2015) têm como objetivo explorar a aplicabilidade de técnicas de MDE em Mundos Virtuais. Assim, foram analisados possíveis padrões de comportamento na base de dados por meio da mineração de Regras de Associação com o uso do algoritmo Apriori, que possibilitou a identificação de mudanças no planejamento pedagógico.

Wegner *et al.* (2015) objetivaram em seu estudo apresentar os resultados de pesquisa envolvendo a técnica de Clusterização na identificação e classificação do Estilo Cognitivo de Aprendizagem (ECA) de estudantes que interagem em um AVA. Os resultados desta pesquisa são comparados aos apresentados por outro trabalho, que utilizou a técnica de Redes Neurais Artificiais.

Wilves *et al.* (2015) têm como objetivo da pesquisa validar em um novo contexto, regras de associação já mineradas, a fim de evidenciar o desânimo do aluno quando realiza atividades individuais e em grupo. A partir da adaptação de algumas Regras de

Associação, foi possível obter um modelo genérico para a descoberta do padrão de comportamento do aluno desanimado, que pode ser usado como subsídio ao professor.

Baker, *et al.* (2014) fizeram um estudo para selecionar as melhores regras de associação. Analisando os dados para determinar quais métricas é a melhor de acordo com estimativas de especialistas sobre interestingness (Merceron *et al.*, 2008). Onde lift e cosine são bons indicadores de interestingness. Além disso, Phi Coefficient, Conviction, e Jaccard também veem a ser bom indicadores de interestingness (Merceron *et al.*, 2008)

Lima *et al.* (2014) investigaram o padrão de comportamento de alunos e professores de cursos a distância em fóruns de discussão. As análises mostraram que para os cursos da área de humanas, são aconselháveis que a oferta das disciplinas fossem modular. Já cursos da área de Exatas demonstraram o inverso, ou seja, que o modo de oferta semestral de disciplinas é mais profícuo à participação nos fóruns.

Romero, C. *et al.* (2014) aplicaram-se um algoritmo de Regras de Associação para reduzir disparidades encontradas nas regras dos algoritmos tradicionais. Com os resultados, conseguiram extrair regras de elevado interesse, descobriram regras de associação frequentes e confiáveis, encontrando lacunas menores no espaço das instâncias, e se comportando melhor que o algoritmo tradicional Apriori.

Baker, *et al.* (2014) apresentaram um estudo que constitui um primeiro passo na investigação visual de design da sala de aula usando um algoritmo de agrupamento. Onde os resultados poderão contribuir na compreensão do impacto das decisões do professor, conduzindo a melhores resultados em termos de engajamento e de aprendizagem para o desenvolvimento de projetos em sala de aula.

França *et al.* (2013) apresentaram um conjunto de atributos para aplicação de técnicas de mineração de dados em avaliações contínuas da aprendizagem, que viabilize realizar o mapeamento do conhecimento de estudantes. Esta pesquisa demonstrou a viabilidade em se aliar técnicas de Mineração de Dados a objetivos educacionais, sendo importante para a elaboração de estratégias pedagógicas que favoreçam a aprendizagem de estudantes com tendência a não alcançar desempenho satisfatório.

Silva *et al.* (2013) têm como objetivo principal analisar o desempenho acadêmico dos alunos, a fim de destacar quais fatores são responsáveis pela retenção e que podem estar impedindo a conclusão do estudante no tempo regular. Foi utilizado o algoritmo Apriori. Os autores chegaram a conclusão que apartir dos dados observados não foi encontrado nenhuma relação causal como solução para diminuir a retenção de alunos, mas como trabalhos futuros pretendem introduzir mais variáveis para análise, como por exemplo, o uso de redes sociais.

Soares (2013) apresentou uma tentativa de agrupar estudantes com dificuldades de aprendizagem similares em uma disciplina de Programação. O resultado permitiu agrupar alunos pelas suas dificuldades de aprendizagem. Foi possível, por exemplo, identificar um grupo de aprendizes com dificuldades em entender Estruturação de Sistemas em Camadas, e outro composto por estudantes com problemas em criar Vetores: Estrutura de dados muito comum na maioria das Linguagens de Programação. Ambos os grupos foram reprovados ao final do semestre.

Romero *et al.* (2013) propõem a aplicação de mineração de regras de associação para melhorar as avaliações e os cursos. Os resultados mostraram um impacto positivo e

significativo sobre os estudantes nas avaliações. E os feedback recolhidos serão muito úteis para melhorar os cursos online.

Webber *et al.* (2013) verificaram o desempenho de Algoritmos de Agrupamento. Esses algoritmos foram testados nas ferramentas *WEKA* e *R* (Mello, 2008). Uma análise preliminar, permitiu concluir que a ferramenta imunológica que utiliza conceitos de Sistemas Imunológicos Artificiais, alcançou os melhores índices de homogeneidade e separação. Em segundo lugar, houve uma variação entre os algoritmos *K-means* e *EM* (Dempster *et al.*, 1976). Por fim, os algoritmos da ferramenta *R* apresentaram índices elevados de homogeneidade e separação para os conjuntos de dados.

Azevedo *et al.* (2012) aplicaram o algoritmo de Agrupamento *EM* para agrupar estudantes que continham conhecimentos prévios diferentes em grupos de acordo com os ganhos de aprendizagem realizados em um Sistema Tutor Inteligente (STI). Os resultados desta análise revelam que o grupo de estudantes de alto desempenho identifica mais rapidamente a relevância do uso do sistema para aprendizagem, sendo assim um primeiro passo na identificação de perfis de grupos de aprendizagem, e então, tornar possível uma adaptação em tempo real do sistema.

Romero *et al.* (2012) têm duplo objetivo: Determinar se a participação dos alunos no fórum do curso pode ser um bom preditor das marcas finais para o curso, e examinar se a classificação proposta via agrupamento pode obter precisão semelhante a Algoritmos de Classificação tradicionais. Pôde - se concluir que a participação dos estudantes no fórum, foi um bom preditor das marcas finais, e obteve precisão semelhante a Algoritmos de Classificação tradicionais.

Martins *et al.* (2012) evidenciaram pontos positivos e negativos de dois trabalhos realizados em 2005 e 2008. Definiu-se utilizar como critério de avaliação a quantidade de acerto na predição da situação final do aluno. Como resultado, apenas o modelo KDD foi capaz de prever qual classe o aluno pertencia (reprovado, aprovado ou desistente), enquanto o novo modelo a probabilidade do aluno ser ou não desistente.

Martinez *et al.* (2012) usaram o Método Hierárquico Aglomerativo (Técnica de agrupamento para dividir os estudantes em grupos). Analisaram-se as interações homem-máquina através dos *logs*, onde se conseguiu resultados expressivos, usando informações adquiridas sobre os grupos para a melhoria da aprendizagem.

Penedo *et al.* (2012) identificaram relações entre os atributos ou a presença de padrões entre os dados que possam influenciar a tomada de decisões. O algoritmo Apriori foi aplicado ao acesso de todos os usuários do sistema. As regras descobertas apontam para uma tendência maior de utilização das ferramentas disponibilizadas pela plataforma que dizem respeito às disciplinas, as relacionadas a aplicativos pouquíssimas utilizadas. Outra descoberta foi relacionada ao horário de maior utilização (tarde e noite).

No seu estudo Peckham *et al.* (2012) descreveram como o algoritmo de Clusterização *K-means* (Dempster *et al.*, 1976) combinado com a Taxonomia de Bloom pode ser usado para determinar os conjuntos de habilidades cognitivas positivas e negativas com respeito à leitura e compreensão de tarefas. Esta informação foi utilizada para modificar os AVAs com intuito de melhorar as habilidades dos estudantes.

Rus *et al.* (2012) utilizaram o algoritmo *K-means* no ambiente *WEKA* (Mello, 2008) para analisar e categorizar os diversos tipos de diálogos existentes em um jogo educacional. O

objetivo era fornecer subsídios para melhor projetar mecanismos de adaptação e *feedback* em STI. Foi realizado o estudo em três conjuntos de dados a partir de três jogos educativos on-line, que através dos dados retirados das conversas dos estudantes pode-se inferir intenções, que por sua vez são cruciais para fornecer *feedback*.

Yoo *et al.* (2012) estudam a viabilidade de analisar Mapas Conceituais para entender a aprendizagem do estudante universitário. Como resultado, verificou-se que a maioria dos alunos na fase inicial do curso se concentra na aprendizagem individual ao invés de aprender através da interação com os outros colegas, e quando usado esse mapa com as técnicas de MDE, pode-se encontrar vários padrões numa sala de aula.

Conti *et al.* (2011) apresentam um estudo de caso cujo foco é a análise de prazos de entrega de atividades no AVA Moodle. Pode-se concluir que nesse contexto que é mais viável a realização de atividades com um prazo menor, pois além do índice de postagem ser maior logo na abertura para postagem, o professor consegue um *feedback* mais rápido do processo de aprendizagem do aluno, podendo esse tomar atitudes corretivas afim de evitar o insucesso ou desistência do aluno.

Fernandes *et al.* (2011) apresentam o uso de técnicas de Mineração de Dados para a formação de turmas mais homogêneas para a aplicação da recuperação paralela. Usando o algoritmo de agrupamento de dados *K-means* conseguiu-se agrupar os alunos pelos problemas de aprendizagem apresentados, inclusive, separando em um grupo específico aqueles alunos com dificuldades de interpretação.

García *et al.* (2011) descreveram uma Ferramenta Educacional de Mineração de Dados colaborativa baseada nas Regras de Associação de Mineração de Dados para a melhoria contínua dos cursos *e-learning*, permitindo que professores possam compartilhar informações descobertas para que outros professores dos cursos semelhantes possam usar essas informações e tomar melhores decisões.

Kardan *et al.* (2011) consideraram o agrupamento não supervisionado através do algoritmo *K-means* para criar um modelo automático para descobrir e reconhecer padrões relevantes durante a interação do estudante com software educacional. Após análise dos resultados, pode-se concluir que as regras geradas são adequadas para orientar nas intervenções de adaptação destinadas a melhorar a eficácia da interação.

Li *et al.* (2011) aplicaram uma técnica de busca automatizada para descobrir modelos de estudante. Para abordar esta questão, foi proposto um método que detecta automaticamente modelos de estudante não dependendo de fatores fornecidos por humanos. Foi usado um sistema agente inteligente que indutivamente aprende habilidades dos estudantes para poder modelá-los.

5. Resultados

Nesta seção são apresentadas algumas informações relevantes sobre os trabalhos encontrados durante a pesquisa como: a fonte das publicações com destaque para os periódicos, que originaram 70% dos trabalhos encontrados, destaca-se também os eventos International Conference on Educational Data Mining e SBIE, com 30% e 26% respectivamente das publicações, a origem dos dados onde o AVA *Moodle* domina essa estatística representando aproximadamente 52% das bases pesquisadas, os algoritmos observou-se o domínio do *K-means* para regras de agrupamento e Apriori para regras

de associação, com respectivamente 30% e 41% das publicações aproximadamente, apesar de um número menor de aplicações, o algoritmo de clusterização *EM*, também merece destaque, e a ferramenta computacional de maior utilização (presente em 81% dos trabalhos) foi o *Weka*. Outro dado observado foi o período em que os trabalhos foram publicados, com o ano de 2012 sendo responsável por aproximadamente 30% das publicações.

6. Discussão

A mineração de dados educacional é uma área de pesquisa muito recente, contudo, há um número importante de contribuições publicadas em revistas, congressos, *workshops* e alguns livros que mostram como a área é promissora.

O presente trabalho realizou uma investigação em diversas pesquisas na área de EDM. Todos os trabalhos analisados se localizam nas duas das principais subáreas apresentado por Baker (2011): Mineração de Relações, onde foi observado que a grande maioria das pesquisas está ligada as Regras de Associação, e a subárea de Clusterização.

Como sugestão para trabalhos futuros seria uma análise de viabilidade do uso de Mapas Conceituais e MDE em interação nos fóruns em ambientes virtuais de aprendizagem. Andriole *et. al* (2005, p. 03) constataram em um fórum que das 144 mensagens trocadas, aproximadamente 91% deram-se entre um integrante do grupo e o professor, sendo praticamente 2% de mensagens trocadas entre pares de aprendizes. Quando um professor quer saber o estado atual de aprendizagem de seus alunos, ele teria que percorrer todos os tópicos e fazer uma análise mais tarde, que seria um processo muito demorado e improvável, dessa forma, para aliviar este problema, poderia criar uma interação visual dos estudantes (Mapa Conceitual), onde contribuiria retratando o conhecimento coletivo de um assunto específico entre um grupo de estudantes e dessa forma entenderia melhor o desenvolvimento do aluno.

Com o Mapa Conceitual da interação, seria feito uma análise da participação dos estudantes nos fóruns, separando-os em grupos através dos Algoritmos de Clusterização, como por exemplo: os alunos que interagiram com outros alunos, alunos que interagiram com o tutor e com outro aluno, alunos que interagiram somente com o tutor, e aqueles com nenhuma interação. Com esses resultados, os tutores e professores já poderiam ter uma ideia de quais alunos teriam pouca participação, e promoveria uma discussão ou atividade, ou criaria salas de chats para os alunos recordarem o que aprenderam para compartilhar com seus colegas durante ou após o horário de aula normal, o conhecimento adquirido, dessa forma, incentivando os alunos que tem pouca ou nenhuma interação a ter uma maior participação.

Com os resultados obtidos no agrupamento, introduziria uma variável informando quais alunos foram aprovados ou reprovados. Nesses dados, seriam aplicados os algoritmos de classificação, em que poderia alcançar grandes resultados, deduzindo aspectos específicos relacionados com aprovação ou reprovação dos estudantes, onde seriam desenvolvidos modelos, que seriam aplicados nas próximas turmas no combate a evasão.

Com resultados preditivos abriria portas para o desenvolvimento e aplicação de ferramentas de software inteligente, que automatizaria todo o processo do uso de Mapas Conceituais e MDE em interações nos fóruns de discussão, onde auxiliaria os

professores a visualizar e analisar, o conhecimento coletivo e a participação dos estudantes nos fóruns no decorrer do curso e em tempo real.

7. Conclusão

Mineração de Dados Educacionais é uma área de investigação ainda nova e se vislumbra a necessidade de um trabalho mais especializado e orientado na área educacional, a fim de obter uma aplicação de semelhante sucesso as outras áreas, como a Mineração de Dados Médicos, Mineração de Dados de E-commerce, etc.

Esta pesquisa teve como objetivo realizar uma revisão dos trabalhos, publicados nos últimos seis anos, que utilizaram as técnicas de Mineração de relações e Agrupamento em MDE, identificando as técnicas, ferramentas empregadas, o objeto de pesquisa bem como a origem das bases de dados e das publicações.

Propõe-se o uso da técnica de classificação associativa, aplicando o algoritmo CBA (Classification Based on Association), que tem como objetivo integrar associação e classificação para construir classificadores mais precisos. Onde analisaria os dados de acesso dos estudantes em ambientes virtuais de aprendizagem com o objetivo de prever a evasão avaliando as interações dos alunos nos três turnos do dia. Esta técnica, de acordo com as pesquisas realizadas, não foi ainda aplicada em bases de dados educacionais.

Outra proposta seria o emprego de Mapas Conceituais e Mineração de Dados Educacionais em fóruns de discussão, onde criaria um novo modelo de perfil do aluno para combater a evasão.

Referências

ANDRIOLE, W. B. ; LOUREIRO, R. Sistematização da avaliação da Aprendizagem em comunidades organizadas no ciberespaço. 2005

AZEVEDO, R.; KINNEBREW, J. S.; BISWAS, G. Identifying Students ' Characteristic Learning Behaviors in an Intelligent Tutoring System Fostering Self-Regulated Learning Franc. p. 65–72, 2012.

BAKER, R.; ISOTANI, S.; CARVALHO, A. Mineração de Dados Educacionais: Oportunidades para o Brasil. Revista Brasileira de Informática na Educação, v. 19, n. 02, p. 03, 2011.

BAKER, RYAN S. J. D.ALMEDA, M. V. et al. Clustering of design decisions in classroom visual displays. Proceedings of the Fourth International Conference on Learning Analytics And Knowledge - LAK '14, p. 44–48, 2014.

BAKER, R. et al. Comparing Expert and Metric-Based Assessments of Association Rule Interestingness. Proceedings of the 7th International Conference on Educational Data Mining (EDM), v. 1, n. 917, p. 44–51, 2014.

COELHO, C. G. et al. Mineração de dados educacionais para identificação de barreiras na utilização da educação a distância. 2015.

CONTI, F. DE; CHARÃO, A. S. Análise de Prazos de Entrega de Atividades no Moodle : um Estudo de Caso Utilizando Mineração de Dados. *Renote*, v. 9, n. 2, 2011.

COSTA, E., BAKER, R. S., Amorim, L. Magalhães, J., & Marinho, T. (2012). Mineração de Dados Educacionais: Conceitos, Técnicas, Ferramentas e Aplicações. *Jornada de Atualização em Informática na Educação*, 1(1), pp. 1-29.

DEMPSTER, A.P, LAIRD, N.M. RDIN, D.B. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society*, v. 39, p. 1-38, 1976.

FERNANDES, C.; PINHEIRO, E. Mineração de Dados na Formação de Turmas para a Recuperação Paralela na Educação Básica. p. 172–175, 2011.

FRANÇA, R. S. DE; JOSÉ, H. Aplicação de Técnicas de Mineração de Dados Baseada na Taxonomia de Bloom para o Mapeamento do Conhecimento na Aprendizagem de Programação. n. *Cbie*, p. 122–131, 2013.

GARCÍA, E. *et al.* A collaborative educational association rule mining tool. *The Internet and Higher Education*, v. 14, n. 2, p. 77–88, 2011.

KARDAN, S.; CONATI, C. A Framework for capturing distinguishing user interaction behaviours in novel interfaces. 2011.

LI, N. *et al.* A Machine Learning Approach for Automatic Student Model Discovery. 2011.

LIMA, E. T. DE; CASTRO, M. F. DE. Uma Análise de Interação em Fóruns de EAD. n. *Cbie*, p. 1203–1212, 2014.

MARTINEZ, R. *et al.* Analysing frequent sequential patterns of collaborative learning activity around an interactive tabletop. 2012.

MARTINS, L. C.; LOPES, D. A.; RAABE, A. Um Assistente de Predição de Evasão aplicado a uma disciplina Introdutória do curso de Ciência da Computação. v. 2, n. *Sbie*, p. 26–30, 2012.

MELLO, C. E. R. Agrupamento de Regiões: Uma abordagem utilizando acessibilidade. 2008.

MERCERON, A., & Yacef, K. (2008). Interestingness Measures for Associations Rules in Educational Data. *Educational Data Mining*, 8, 57-66.

NETO, V. S. M. *et al.* Data Mining to Identify Learning Groups with Difficulties in Programming Education. p. 1–10, 2016.

NUNES, F. B.; VOSS, G. B.; CAZELLA, S. C. Mineração de dados educacionais e Mundos Virtuais : um estudo exploratório no OpenSim. n. *Cbie*, p. 1044–1053, 2015.

PENEDO, J. R.; CAPRA, E. P. Mineração de Dados na Descoberta do Padrão de Usuários de um Sistema de Educação à Distância. n. Sbsi, p. 396–407, 2012.

PECKHAM, T.; MCCALLA, G. Mining Student Behavior Patterns in Reading Comprehension Tasks. 2012.

PERES, S. M.; BOSCARIOLI, C.; SILVA, M. Análise de Logs da Web por meio de Técnicas de Data Mining. Discovery, n. 045, p. 12, 2003.

RUS, V.; MOLDOVAN, C.; GRAESSER, A. C. Automated Discovery of Speech Act Categories in Educational Games. p. 25–32, 2012.

ROMERO, C.; VENTURA, S. Educational data mining: A review of the state of the art. IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews, v. 40, n. 6, p. 601–618, 2010

ROMERO, C. et al. Association rule mining using genetic programming to provide feedback to instructors from multiple-choice quiz data. Expert Systems, v. 30, n. 2, p. 162–172, 2013.

ROMERO, C. et al. Classification via clustering for predicting final marks based on student participation in forums. p. 4–7, 2012

ROMERO, C. et al. Reducing gaps in quantitative association rules: A genetic programming free-parameter algorithm. Integrated Computer-Aided Engineering, v. 21, n. 4, p. 321–337, 2014.

SILVA, C. V. A. et al. Mining Retention Rules from Student Transcripts: A Case Study of the Information Systems programme at a Federal University. n. Cbie, p. 577–586, 2013.

SOARES, R. Mineração de Dados na Identificação de Grupos de Estudantes com Dificuldades de Aprendizagem no Ensino de Programação. n. January, 2013.

WEBBER, C. G.; ZAT, D.; LIMA, M. DE F. W. DO P. Utilização de algoritmos de agrupamento na mineração de dados educacionais. p. 1–10, 2013.

WEGNER, W. et al. Técnica de Clusterização para Determinar o Estilo Cognitivo de Usuários em um Ambiente Virtual de Aprendizagem. p. 444–449, 2015.

WILVES, F. D. S. L. K.; CAZELLA, S. C. Analisando o desânimo de alunos em ambientes virtuais através da mineração de dados educacionais. p. 65–70, 2015.

YOO, J. S.; WAYNE, F.; CHO, M. Mining Concept Maps to Understand University Students' Learning. p. 184–187, 2012.