

Análise de Trabalhos Sobre a Aplicação de Técnicas de Mineração de Dados Educacionais na Previsão de Desempenho Acadêmico

Rodrigo Magalhães Mota dos Santos¹, Cristiano Grijó Pitangui², Alessandro Vivas Andrade¹, Luciana Pereira de Assis¹

¹ Universidade Federal dos Vales do Jequitinhonha e Mucuri

² Universidade Federal de São João del Rei

{rodrigo.mota, lpassis}@ufvjm.edu.br,
pitangui.cristiano@ufsj.edu.br, alessandro.vivas@gmail.com

Abstract. *This work presents an analysis and a classification of publications in the field of Educational Data Mining published by Brazilian researchers in events and periodicals in Latin America over the past four years. The objective of the present work is to point and present what is being researched in academic performance prediction area. The results of the classification of publications are presented and we show interesting information about the area of research, identifying some possibilities less (or not yet) explored.*

Resumo. *O presente trabalho apresenta uma análise e classificação de publicações na área de Mineração de Dados Educacionais publicados por pesquisadores brasileiros em eventos e periódicos na América Latina nos últimos quatro anos. O objetivo deste trabalho é apontar e apresentar o que está sendo pesquisado na área de predição de desempenho acadêmico. Os resultados da classificação das publicações são apresentados e mostram informações interessantes a respeito da área de pesquisa, identificando possibilidades pouco (ou ainda não) exploradas.*

1. Introdução

A tecnologia está cada vez mais presente no cotidiano das pessoas, trazendo mudanças significativas em toda a estrutura da sociedade atual, seja na política, na economia, na esfera jurídica, na área social, e como não poderia ser diferente, na educação. A evolução tecnológica tem permitido um maior acesso à informação, trazendo mudanças profundas na área acadêmica, e o uso das Tecnologias de Informação e Comunicação (TIC) se torna um fator transformador.

Atualmente, é crescente a utilização de TIC's. Este fato permite uma maior integração entre as Instituições de Ensino, docentes e discentes, que passaram a utilizar a tecnologia para produzir e gerenciar conteúdos didáticos, elaborar disciplinas presenciais, semipresenciais e cursos totalmente a distância, com o objetivo final de gerar e adquirir conhecimento, tanto em caráter individual como coletivo.

O uso destas tecnologias, como os Ambientes Virtuais de Aprendizagem (AVA), gera um enorme volume de dados, resultantes de interações e de registros de informações de professores, tutores, alunos, gestores e demais atores do sistema

educacional. Este enorme volume de dados armazenado está sendo estudado por profissionais da área de Informática na Educação (IE).

Esta pesquisa apresenta uma revisão dos trabalhos que utilizam técnicas de Mineração de Dados Educacionais (do inglês, *Education Data Mining* – EDM) aplicadas na previsão de desempenho acadêmico, publicados nas principais conferências e periódicos da área de IE. 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 previsão de desempenho acadêmico. Como principal contribuição, a pesquisa aponta informações interessantes a respeito da área, identificando possibilidades pouco (ou ainda não) exploradas.

Este trabalho está organizado em 5 seções. A seção dois apresenta a metodologia aplicada na busca, seleção e análise dos trabalhos. A terceira seção apresenta um pequeno resumo sobre cada trabalho analisado. A quarta seção apresenta e interpreta os resultados do estudo. Finalmente, a quinta seção apresenta as conclusões deste trabalho, indicando algumas perspectivas de continuidade da pesquisa.

2. Materiais e Métodos

Esta pesquisa engloba as publicações realizadas por pesquisadores brasileiros em eventos e periódicos ocorridos na América Latina de 2012 a 2015. A busca foi realizada através do *Google Scholar* e em anais dos principais eventos e periódicos de IE.

Durante as pesquisas foram utilizados termos como ‘Mineração de Dados Educacionais’, ‘Previsão de Desempenho Acadêmico’, ‘Classificação de Dados Acadêmicos’ e ‘Previsão Utilizando Dados Educacionais’. Foram pré-selecionados 67 trabalhos, após a filtragem e seleção de trabalhos que abordam a previsão de desempenho acadêmico utilizando as técnicas de classificação e/ou regressão restaram apenas 30.

Os critérios empregados para analisar os trabalhos foram fonte das publicações, fonte dos dados analisados, algoritmos e ferramentas utilizadas, além do objetivo geral. Os trabalhos são apresentados em ordem cronológica crescente.

3. Trabalhos Pesquisados

Nesta seção são apresentados breves resumos dos trabalhos analisados nesta pesquisa.

Santos *et al.* (2012) relatam um estudo de caso que permite identificar alunos que têm maior risco de reprovação. Resultados preliminares mostram que os modelos criados permitem a identificação da propensão à reprovação com taxa de acerto em torno de 69%. Por meio dos experimentos realizados, os autores comprovaram que quanto maior a dedicação nas atividades presenciais e semipresenciais, melhor o desempenho do aluno nas avaliações somativas.

Manhães *et al.* (2012) têm como objetivo identificar, apresentar e quantificar as variáveis que representam os principais fatores que influenciam a conclusão, a evasão e permanência, além do tempo médio para concluir o curso de graduação. Os autores afirmam que a análise dos atributos permite identificar os fatores de sucesso e insucesso específicos para cada curso.

Gottardo *et al.* (2012) têm como objetivo principal a obtenção de um modelo de previsão de desempenho de estudantes a partir de dados coletados em séries temporais durante o período de realização de um curso a distância. Os resultados obtidos com os

experimentos apontam para viabilidade de se realizar inferências relativas ao desempenho de estudantes, obtendo-se taxas de acurácia acima de 74%, mesmo em etapas iniciais de realização do curso.

Martins *et al.* (2012) buscam aprimorar um Assistente de Predição de Evasão (APE) por meio da aplicação do processo KDD acrescentando novas variáveis obtidas com os dados do Sistema Tutor Inteligente (STI) e de outras fontes. No processo de KDD foram utilizados os algoritmos *OneR* e *NNge* presentes na ferramenta *Weka*. Os autores observaram que o modelo baseado no processo de KDD resultou em uma melhor identificação e maior precisão nos resultados.

Gottardo *et al.* (2012a) propõem a definição de um conjunto de atributos amplo e generalizável a ser utilizado para realizar inferências relativas ao desempenho de estudantes. Experimentos realizados demonstraram a viabilidade desta proposta, que alcança índices de 76% de acerto na previsão de desempenho.

Rodrigues *et al.* (2013) têm como objetivo principal a obtenção de um modelo de previsão de desempenho de estudantes a partir de dados de interação durante o período de realização de um curso na modalidade à distância. Os autores afirmam, com uma confiança de 95%, que 79,2% dos alunos que interagem via fórum de discussão, a medida que aumenta a quantidade de interação, tende a aumentar o desempenho acadêmico.

Os resultados obtidos por Gottardo *et al.* (2013) apontam para a viabilidade da aplicação da técnica de balanceamento de classes para identificar grupos de estudantes com maior risco de reprovação. Os autores avaliaram o desempenho de estudantes em uma disciplina por meio de dados extraídos na plataforma *Moodle*, utilizando os algoritmos *Multilayer Perceptron* e *Random Forest* com o auxílio da ferramenta *Weka*. O melhor resultado obteve 80,7% de acurácia.

Gotardo *et al.* (2013) apresentam uma abordagem que usa algoritmos de aprendizado acoplados para integrar diferentes técnicas de aprendizado e Mineração de Dados. Segundo os autores, os experimentos realizados demonstraram que a abordagem proposta auxilia a obtenção de recomendações sobre o desempenho num ambiente com poucos dados sobre os alunos.

Gottardo (2013) tem como objetivo geral investigar a possibilidade de geração de inferências relativas ao desempenho futuro de estudantes. Durante a realização dos experimentos, o autor obteve taxas de acurácia que ficaram entre 72% e 80%, e concluiu que a principal contribuição do trabalho foi demonstrar a viabilidade e apresentar os procedimentos que podem ser utilizados para obter inferências relativas ao desempenho de estudantes em um curso EAD.

Guércio *et al.* (2014) propõem a criação de um modelo que auxilie o professor na análise do comportamento dos alunos no decorrer da oferta de determinada disciplina. Segundo os autores, a acurácia média encontrada foi de 73% na classificação do desempenho e foi possível observar a possibilidade de transformar os dados armazenados na base de dados da plataforma *Moodle* em conhecimento, gerando regras muito úteis para o apoio a tomada de decisões.

Dos Santos *et al.* (2014) acreditam que é importante identificar primeiramente quais são as disciplinas que mais influenciam na evasão do curso de graduação. Depois, utilizam essas disciplinas para predição de desempenho dos alunos ao longo das semanas. A partir dos experimentos realizados, os autores constataram que já ao final do primeiro período, é possível obter previsões com acurácia superior a 80%.

Detoni *et al.* (2014) têm como objetivo prever a reprovação de alunos de educação a distância utilizando o registro de interações entre alunos, tutores e professores. Os melhores resultados foram alcançados pelas Redes Bayesianas, 80% de acurácia média. Segundo os autores, o modelo preditivo gerado é mais generalizável, podendo ser aplicado em qualquer situação onde é possível contar interações.

Marques (2014) tem como objetivo identificar padrões de acesso dos alunos que evadem de cursos na modalidade de EAD e gerar regras que caracterizem o perfil de acesso desses alunos. Segundo a autora, durante o estudo, foi possível observar a relação existente entre alguns atributos, demonstrando a importância da interação constante entre aluno e professor.

Da Costa *et al.* (2014) buscam identificar o perfil de alunos com potencial de evasão. Para isto, utilizam dados, disponibilizados em planilhas, de cursos de especialização à distância. Os dados pré-processados e transformados foram classificados utilizando o algoritmo *J48* que obteve 97,6% de precisão.

Santana, Maciel e Rodrigues (2014) tiveram como objetivo realizar a avaliação da dimensão perfil de uso no ambiente *Moodle*. Foram utilizados dados de um curso ofertado na modalidade semipresencial extraídos do banco de dados do AVA *Moodle*. Foram utilizados sete algoritmos para analisar o desempenho do perfil, onde o *J48* obteve o melhor desempenho, alcançando 74% de acurácia.

Pequeno *et al.* (2014) utilizam técnicas de Estatística e de Mineração de Dados para analisar a distribuição de mensagens de alunos e professores/tutores dentro do período de aplicação de cada fórum. Segundo os autores, os experimentos realizados identificaram regras de associação com altos níveis de confiança, e as análises permitiram identificar que tipo de oferta de disciplina (modulares ou semestral) é mais adequado a cada área de conhecimento (Exatas, Linguística, etc).

Santana *et al.* (2014) tiveram como propósito identificar quais são as ferramentas mais utilizadas e o impacto das mesmas sobre o desempenho dos estudantes nas disciplinas. Foram utilizados quatro algoritmos, os melhores resultados foram obtidos pelo *J48*, onde o *recall* foi de 0,769 e a precisão foi de 77,8%. Segundo os autores, os resultados obtidos nesse estudo apontam a viabilidade de realizar inferências relativas ao uso das ferramentas disponíveis no AVA *Moodle*.

De Brito *et al.* (2014) tiveram como objetivo identificar os estudantes que necessitam de apoio didático. Foi analisado o desempenho acadêmico no primeiro período de um curso de graduação em Ciência da Computação. O algoritmo *SMO* obteve a maior precisão, com média de 74,7% de acerto. Segundo os autores, os resultados podem ser utilizados para que medidas sejam tomadas para a diminuição da retenção ou evasão acadêmica.

Por meio de técnicas de Mineração de Dados, Kampff *et al.* (2014) buscam identificar perfis de alunos com risco de evasão ou reprovação, visando à geração de alertas para sensibilizar o professor sobre possíveis problemas. Segundo os autores, nos resultados obtidos, há evidências de que o índice de evasão nas turmas com os alertas tenha sido significativamente inferior ao índice de evasão observado na amostra de dados históricos.

Rigo *et al.* (2014) apresentam um estudo sobre aspectos da evasão escolar e descrevem a aplicação e os resultados de uso de sistemas de EDM e *Learning Analytics* (LA). Segundo os autores, a taxa de acerto foi na ordem de 87% na identificação de

perfis de alunos evasores, e a realização de intervenções pedagógicas nas turmas participantes do estudo resultaram na redução da evasão.

Adeodato *et al.* (2014) trazem uma solução de Mineração de Dados para avaliar e prever a qualidade da educação secundária privada no Brasil a partir das bases do ENEM e do Censo Escolar. Os autores afirmam que os resultados deixaram claro que os principais fatores que influenciam a boa qualidade das escolas estão nos aspectos econômico-financeiros, seja direta ou indiretamente, ou em aspectos culturais da família.

Gottardo *et al.* (2014) apresentam um estudo preliminar do uso de algumas técnicas de Mineração de Dados para identificar quais parâmetros da base dados do AVA *Moodle* possibilitariam uma inferência sobre desempenho futuro dos estudantes. Segundo os autores, este estudo se torna relevante, pois aponta para a possibilidade de utilizar um conjunto amplo de atributos que representem de forma abrangente e generalizável um estudante.

Warley *et al.* (2014) descrevem uma proposta para uso de técnicas de Mineração de Dados para identificação de causas de evasão e de levantamento do perfil do aluno evadido. Como se trata de uma proposta, o autor não cita claramente como extrairá a base de dados, que algoritmos utilizarão e de quais ferramentas fará uso.

Queiroga *et al.* (2015) têm como objetivo a predição precoce da evasão de acadêmicos de um curso ofertado no AVA *Moodle*. Durante o processo de classificação dos dados, foram utilizados nove algoritmos presentes na ferramenta *Weka*, sendo 81,25% a taxa mais alta de acertos obtida. Os autores chegaram a conclusão, através dos resultados obtidos, que é viável a predição precoce do risco de evasão analisando os *log's* das 4 primeiras semanas do curso.

Ferreira (2015) tem como foco a identificação de fatores relacionados à conclusão do Ensino Fundamental utilizando os Microdados do Censo Escolar da Educação Básica do INEP de 2014. Foram aplicados os algoritmos de classificação *J48* e o filtro *CfsSubsetEval*, ambos da ferramenta *Weka*. A autora concluiu que é interessante a aplicação de filtros, pois permite uma melhor visualização das regras obtidas, mesmo diante de uma diminuição da acurácia dos resultados.

Laisa e Nunes (2015) têm como objetivo analisar uma base de dados de alunos do Ensino Médio. Utilizando o algoritmo *J48*, os experimentos chegaram a taxas de 77,9% de acertos. Segundo as autoras, a técnica de classificação aplicada permitiu compreender as informações armazenadas dos estudantes de ensino médio.

Santos *et al.* (2015) relatam uma pesquisa que identifica o aluno desanimado em um AVA. Os dados foram obtidos através da aplicação de questionários e dos *logs* dos alunos no AVA *Moodle*. A classificação dos dados foi realizada com o algoritmo *J48* da ferramenta *Weka*, e o modelo gerado obteve acerto em cerca de 91% dos dados.

Costa *et al.* (2015) têm como objetivo fornecer um método (utilização de algoritmos classificadores) para a identificação de estudantes cotistas com possibilidade de terem menor rendimento acadêmico durante o curso. Foram utilizados os algoritmos *J48*, *Naive Bayes*, *SMO*, *IBk* e *Multilayer Perceptron* da ferramenta *Weka*. O estudo obteve taxas de acerto de até 88% e, segundo os autores, o método pode ser uma abordagem viável na identificação de estudantes cotistas com possibilidades de terem menor desempenho acadêmico.

Silva *et al.* (2015) mostram a versão funcional de uma ferramenta para apresentar as tendências de evasão geradas por Mineração de Dados em AVA's a partir

de interações de alunos em fóruns de discussão. Os melhores resultados foram obtidos pelo algoritmo *J48* e, segundo os autores, a partir dos modelos gerados por árvore de decisão, foi possível criar um diagnóstico mais detalhado das tendências de evasão e reprovação, para auxiliar as partes interessadas na tomada de decisão.

Manhães (2015) propõe uma arquitetura chamada EDM WAVE que engloba todo o processo de descoberta de conhecimento em dados para predição do desempenho acadêmico de graduandos. Segundo a autora, os resultados obtidos mostram que a arquitetura EDM WAVE fornece suporte para os gestores educacionais realizarem o monitoramento do progresso dos graduandos a cada semestre letivo, e identifica aqueles que apresentam maiores riscos de abandonar o sistema educacional.

4. Resultados

Nesta seção são apresentadas algumas informações relevantes sobre os trabalhos encontrados durante a pesquisa como, a fonte das publicações, a origem dos dados, os algoritmos e ferramentas utilizados, o que se analisou, e o ano das publicações.

Na Tabela 1 pode ser observada a fonte das publicações, com destaque para as Conferências (Simpósios, Workshops, Congressos), que originaram aproximadamente 74% dos trabalhos encontrados.

Tabela 1. Fonte das publicações

Fonte das Publicações	Quantidade
Conferências (Simpósios, Workshops, Congressos)	22
Periódicos	5
Teses, Dissertações e TCC's	3
Total	30

Na Tabela 2 pode-se observar a fonte das Bases de Dados, que ultrapassa a quantidade total de publicações, devido ao fato de uma pesquisa ter mais de uma fonte de dados. O AVA *Moodle* domina essa estatística, representando aproximadamente 57% das bases pesquisadas, seguido dos Sistemas de Gerenciamento Acadêmico (SGA).

Tabela 2. Fonte das bases de dados

Fonte da Base de Dados	Quantidade
AVA Moodle	17
SGA	5
Não Mencionado	3
AVA não identificado	2
Censo Escolar	2
Sistema Tutor Inteligente	1
AVA SOLAR	1
Outros	2

A utilização em larga escala do ambiente *Moodle* ficou evidente, sendo a maior fonte de dados para as pesquisas. Este ambiente está presente em 230 países com mais de 76 mil sites registrados [Moodle, 2016].

Os algoritmos utilizados nos trabalhos estão listados na Tabela 3. Observa-se o domínio das árvores de decisão com o *J48* e *Random Forest* presentes, respectivamente, em 60% e 40% das publicações.

Tabela 3. Algoritmos utilizados nas publicações

Algoritmos	Quantidade
J48	18
Multilayer Perceptron	14
Random Forest	12
Naive Bayes	9
Outros	4
SVM; Simple Cart; Bayes Net; Decision Table	4
OneR; IBk; Não Mencionado	3
Ada Boost; Simple Logistic; RBF; SMO	2
ADTree; Apriori; ID3; Jrip; Random Tree; REPTree	1

Ao analisar os resultados, observa-se uma maior utilização de algoritmos de Árvore de Decisão. Este fato se deve a indução de árvores de decisão ser uma das formas mais simples, e ainda assim mais bem-sucedidas de Algoritmos de Aprendizagem [Russel and Norvig, 2004].

As ferramentas computacionais mais utilizadas estão listadas na Tabela 4. O *Weka* foi a de maior utilização (presente em aproximadamente 77% dos trabalhos).

Tabela 4. Ferramentas utilizadas nas publicações

Ferramentas	Quantidade
Weka	23
Não Mencionado	2
RapidMiner	2
Pentaho	1
R	1
Linguagens	1

A ferramenta mais utilizada foi o *Weka*, ferramenta já consolidada que reúne diversos algoritmos de Mineração de Dados.

Segundo o levantamento, Tabela 5, o que mais se analisa é o desempenho utilizando dados de disciplinas, o que representa o foco da metade das pesquisas.

Tabela 5. Tipo de análise das publicações

Tipo de Análise	Quantidade
Desempenho em Disciplinas	15
Desempenho em um Curso	10
Desempenho em Fóruns	2
Desempenho em Escola	2
Desempenho de Alunos Cotistas	1

Os objetivos das pesquisas são bem variados. As investigações abrangem estudos de desempenho em fóruns, em disciplinas específicas, em cursos, na graduação de escolas particulares e do ensino fundamental. Outras visam gerar alertas a estudantes com baixo desempenho ou com tendências a evasão, e têm-se, ainda, os que procuram traçar o perfil dos estudantes. Pode-se observar uma busca incessante pela melhor forma

de se representar o objeto a ser analisado. Vários trabalhos têm como foco a busca pelos melhores atributos.

No Gráfico 1 observase o período em que os trabalhos foram publicados, com o ano de 2014 sendo responsável por aproximadamente 47% das publicações.

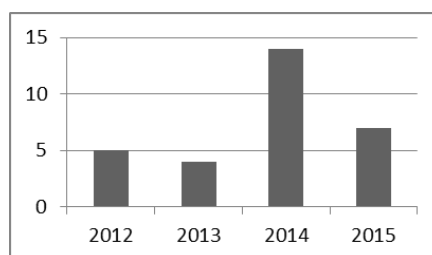


Gráfico 1. Número de publicações por ano

Dos trabalhos analisados poucos procuraram utilizar uma abordagem diferente de apenas aplicar o algoritmo de classificação na base de dados. Foi identificado apenas um, baseado no conceito de *co-training*, que utiliza algoritmos acoplados para integrar diferentes técnicas de aprendizado (Aprendizado Bayesiano e Árvore de Decisão).

Outro dado interessante é que dos 30 trabalhos analisados, apenas 8 mencionaram o uso de alguma técnica de Seleção de Atributos automática, ou seja, por meio de algoritmos. Sendo que não foi identificado nenhum trabalho que utilizasse a abordagem Cápsula (*Wrapper*), onde o algoritmo de indução é utilizado para avaliar, e consequentemente selecionar o melhor subconjunto de atributos [Kohavi *et al.*, 1996].

5. Conclusão

Com o avanço das TIC's, e o crescimento da oferta de cursos nas modalidades à distância e semipresenciais, e da utilização de AVA's e SGA, o volume de dados gerado e armazenado tem aumentado em larga escala. Este aumento vertiginoso de dados acadêmicos abriu a possibilidade do uso de técnicas de Mineração de Dados com o objetivo de encontrar padrões e, por consequência, descobrir conhecimento útil que possa ser utilizado na gestão do ensino.

Esta pesquisa teve como objetivo realizar uma revisão dos trabalhos, publicados nos últimos quatro anos, que utilizaram técnicas de EDM para prever o desempenho acadêmico, identificando as técnicas e ferramentas empregadas, o objeto de pesquisa bem como a origem das bases de dados e das publicações.

Através dos trabalhos relacionados neste estudo pode-se perceber uma tendência de evolução nesta área de pesquisa o que pode acarretar um avanço no desenvolvimento de novas funcionalidades em ambientes de ensino e uma consequente melhora na gestão acadêmica.

Como sugestão de continuidade de pesquisa, poderia ser utilizado alguma abordagem que utilize diferentes técnicas de Aprendizado de Máquina para derivar novos atributos a partir de atributos originais no intuito de facilitar a extração de informações. Poderia ainda, de forma contrária a anterior, utilizar técnicas de Seleção de Atributos automática que fazem emprego da abordagem Cápsula (*Wrapper*).

Referências

- Adeodato, P. J., Santos Filho, M. M., & Rodrigues, R. L. (2014). *Predição de desempenho de escolas privadas usando o ENEM como indicador de qualidade escolar*. In: Anais do XXV Simpósio Brasileiro de Informática na Educação.
- Costa, F., dos Santos Silva, A. R., de Brito, D. M., & do Rêgo, T. G. (2015). *Predição de sucesso de estudantes cotistas utilizando algoritmos de classificação*. In: Anais do XXVI Simpósio Brasileiro de Informática na Educação.
- da Costa, S. S., Cazella, S., & Rigo, S. J. (2014). *Minerando dados sobre o desempenho de alunos de cursos de educação permanente em modalidade EAD: Um estudo de caso sobre evasão escolar na UNA-SUS*. RENOUE, 12(2).
- de Brito, D. M., de Almeida Júnior, I. A., Queiroga, E. V., & do Rêgo, T. G. (2014). *Predição de desempenho de alunos do primeiro período baseado nas notas de ingresso utilizando métodos de aprendizagem de máquina*. In: Anais do XXV Simpósio Brasileiro de Informática na Educação.
- Detoni, D., Matsumura, R. A., & Cechinel, C. (2014). *Modelagem e predição da situação acadêmica de alunos de educação a distância*. In: Anais dos Workshops do III Congresso Brasileiro de Informática na Educação.
- dos Santos, R. N., de Albuquerque Siebra, C., & Oliveira, E. S. (2014). *Uma Abordagem Temporal para Identificação Precoce de Estudantes de Graduação a Distância com Risco de Evasão em um AVA utilizando Árvores de Decisão*. In: Anais dos Workshops do III Congresso Brasileiro de Informática na Educação.
- Ferreira, G. (2015). *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 IV Congresso Brasileiro de Informática na Educação.
- Gotardo, R., Cereda, P. R. M., & Junior, E. R. H. (2013). *Predição do Desempenho do Aluno usando Sistemas de Recomendação e Acoplamento de Classificadores*. In: Anais do XXIV Simpósio Brasileiro de Informática na Educação.
- Gottardo, E. (2013). *Estimativa de desempenho acadêmico de estudantes em um AVA utilizando técnicas de mineração de dados* (Dissertação Mestrado). Universidade Tecnológica Federal do Paraná.
- Gottardo, E., Kaestner, C., & Noronha, R. V. (2014). *Estimativa de Desempenho Acadêmico de Estudantes: Análise da Aplicação de Técnicas de Mineração de Dados em Cursos a Distância*. Revista Brasileira de Informática na Educação, 22(01), 45.
- Gottardo, E., Kaestner, C., & Noronha, R. V. (2013). *Aplicação de Técnicas de Mineração de Dados para Estimativa de Desempenho Acadêmico de Estudantes em um AVA Utilizando Dados com Classes Desbalanceadas*. In: Anais do ICBL2013 - International Conference on Interactive Computer Aided Blended Learning.
- Gottardo, E., Kaestner, C., & Noronha, R. V. (2012). *Avaliação de Desempenho de Estudantes em Cursos de Educação a Distância Utilizando Mineração de Dados*. In: Anais do I Workshop de Desafios da Computação Aplicada à Educação.
- Gottardo, E., Kaestner, C., & Noronha, R. V. (2012a). *Previsão de Desempenho de Estudantes em Cursos EAD Utilizando Mineração de Dados: uma Estratégia Baseada em Séries Temporais*. In: Anais do XXIII Simpósio Brasileiro de Informática na Educação.
- Guércio, H., Marques, P., Ströele, V., Pereira, C. K., & Barrere, E. (2014). *Análise do Desempenho Estudantil na Educação a Distância Aplicando Técnicas de Mineração de Dados*. In: Anais dos Workshops do III Congresso Brasileiro de Informática na Educação.
- Kampff, A. J. C., Ferreira, V. H., Reategui, E., & de Lima, J. V. (2014). *Identificação de Perfis de Evasão e Mau Desempenho para Geração de Alertas num Contexto de Educação a Distância*. Revista Latinoamericana de Tecnología Educativa-RELATEC, 13(2), pp. 61-76.
- Kohavi, R., Sommerfield, D., Odugherty, J. *Data Mining Using MLC++, A Machine Learning Library in C++*. In Proceedings of the 8th International Conference on Tools with Artificial Intelligence (ICTAI '96). IEEE Computer Society, Washington, DC, USA, 234-245, 1996.

- Laisa, J., & Nunes, I. (2015). *Mineração de Dados Educacionais como apoio para a classificação de alunos do Ensino Médio*. In: Anais do XXVI Simpósio Brasileiro de Informática na Educação.
- Manhães, L. M. B. (2015). *Predição Do Desempenho Acadêmico De Graduandos Utilizando Mineração De Dados Educacionais* (Tese Doutorado). Universidade Federal do Rio de Janeiro.
- Manhães, L. M. B., Cruz, S. M. S., Costa, R. J. M., Zavaleta, J., & Zimbrão, G. (2012). *Identificação dos fatores que influenciam a evasão em cursos de graduação através de sistemas baseados em mineração de dados: Uma abordagem quantitativa*. In: Anais do VIII Simpósio Brasileiro de Sistemas de Informação.
- Marques, J. L. D. Q. (2014). *Mineração de dados educacionais: um estudo de caso utilizando o ambiente virtual do SENAI* (Trabalho de Conclusão de Curso). Universidade Estadual da Paraíba.
- Martins, L. C., Lopes, D. A., & Raabe, A. (2012). *Um Assistente de Predição de Evasão aplicado a uma disciplina Introdutória do curso de Ciência da Computação*. In: Anais do XXIII Simpósio Brasileiro de Informática na Educação.
- MOODLE. Moodle.net Courses and Content. 2016. <https://moodle.net/stats/?lang=pt_br>. [Acesso em: 16-Mai-2016]
- Pequeno, H., de Ávila, R. L., Cruz, E., Alexandre, M., de Lima, E. T., & de Castro, M. F. (2014). *Uma Análise de Interação em Fóruns de EAD*. In: Anais do XXV Simpósio Brasileiro de Informática na Educação.
- Queiroga, E., Cechinel, C., & Araújo, R. (2015). *Um Estudo do Uso de Contagem de Interações Semanais para Predição Precoce de Evasão em Educação a Distância*. In: Anais dos Workshops do IV Congresso Brasileiro de Informática na Educação.
- Rigo, S.J., Cambuzzi, W., Barbosa, J.L.V., Cazella, S.C. (2014). *Aplicações de Mineração de Dados Educacionais e Learning Analytics com foco na evasão escolar: oportunidades e desafios*. Revista Brasileira de Informática na Educação, 22(1).
- Rodrigues, R. L., de Medeiros, F. P., & Gomes, A. S. (2013). *Modelo de Regressão Linear aplicado à previsão de desempenho de estudantes em ambiente de aprendizagem*. In: Anais do XXIV Simpósio Brasileiro de Informática na Educação.
- Russel, S., & Norvig, P. (2004). *Inteligência artificial*. Editora Campus.
- Santana, L. C., Maciel, A. M., & Rodrigues, R. L. (2014). *Avaliação do Perfil de Uso no Ambiente Moodle Utilizando Técnicas de Mineração de Dados*. In: Anais do XXV Simpósio Brasileiro de Informática na Educação.
- Santana, M. A., dos Santos Neto, B. F., & de Barros Costa, E. (2014). *Avaliando o Uso das Ferramentas Educacionais no Ambiente Virtual de Aprendizagem Moodle*. In: Anais do XXV Simpósio Brasileiro de Informática na Educação.
- Santos, H., Camargo, F., & Camargo, S. (2012). *Minerando Dados de Ambientes Virtuais de Aprendizagem para Predição de Desempenho de Estudantes*. Latin American Conference on Learning Objects and Technologies (LACLO), 3(1).
- Santos, F. D., Bercht, M., & Wives, L. (2015). *Classificação de alunos desanimados em um AVEA: uma proposta a partir da mineração de dados educacionais*. In: Anais do XXVI Simpósio Brasileiro de Informática na Educação.
- Silva, F., Fonseca, L. C., Silva, R., & Silva, J. D. (2015). *Uma ferramenta para visualização das tendências de evasão geradas por mineração de dados a partir das interações em fóruns de discussão*. Conferências LACLO, 6(1), p. 237.
- Warley Barbosa, D. M., Jatobá, A., Leite, A., & Soares, E. (2014). *Uma Proposta para Identificação de Causas da Evasão na Educação a Distância através de Mineração de Dados*. In: Anais da ERBASE - Escola de Computação Bahia-Alagoas-Sergipe.