

PPGTER/TEC.05.2019.REL

Mineração de Dados Educacionais: Uma Revisão da Literatura

Autores

Thiago Siqueira Sonnenstrahl
thiago.svs@hotmail.com

Solange de Lurdes Pertile
solangepertile@gmail.com

Antônio Guilherme Schmitz Filho
schmitzg@gmail.com

Versão 1.0
Status: Final
Distribuição: Externa
JULHO 2019



2019 PPGTER – Programa de Pós-Graduação em Tecnologias Educacionais em Rede

Atribuição-Não Comercial 4.0 Internacional (CC BY-NC 4.0)

Você tem o direito de compartilhar, copiar e redistribuir o material em qualquer suporte ou formato; adaptar, remixar, transformar, e criar a partir do material, de acordo com o seguinte: você deve dar o crédito apropriado, prover um link para a licença e indicar se mudanças forem feitas. Você deve fazê-lo em qualquer circunstância razoável, mas de nenhuma maneira que sugira que o licenciante apoia você ou seu uso. Você não pode usar o material para fins comerciais.

Programa de Pós-Graduação em Tecnologias Educacionais em Rede - PPGTER

Editoria Técnica do PPGTER

Universidade Federal de Santa Maria

Av. Roraima n. 1000

Centro de Educação, Prédio 16, sala 3146

Santa Maria – RS – CEP 97105-900

Fone / FAX: 55 3220 9414

ppgter@ufsm.br

edtec.ppgter@gmail.com

ISSN: 2675-0309

Relatórios Técnicos do Programa de Pós-Graduação em Tecnologias Educacionais em Rede / Programa de Pós-Graduação em Tecnologias Educacionais em Rede, Universidade Federal de Santa Maria. – Vol. 1. n. 1 (2019) Jan/Jul. – Santa Maria: PPGTER/UFSM, 2019.

Periodicidade semestral.

1. Tecnologia Educacional. 2. Desenvolvimento de Tecnologias Educacionais. 3. Gestão de Tecnologias Educacionais. I. Universidade Federal de Santa Maria. Programa de Pós-Graduação em Tecnologias Educacionais em Rede.

Resumo

Este relatório tem como objetivo apresentar os dados dos trabalhos correlatos da proposta de dissertação, intitulada “utilização da mineração de dados para compor indicadores da tendência da evasão nos cursos EaD do Instituto Federal de Educação, Ciência e Tecnologia Farroupilha” que está sendo desenvolvida no Programa de Pós-Graduação em Tecnologias Educacionais em Rede, ligada à linha de pesquisa: Gestão de Tecnologias Educacionais em Rede, com a orientação da professora Dra. Solange de Lurdes Pertile e Coorientadora Giliane Bernardi. Seu objetivo geral é, por meio da Mineração de Dados Educacionais (MDE), propor e avaliar indicadores da tendência de evasão em cursos do Instituto Federal Farroupilha na modalidade à distância.

1. INTRODUÇÃO

Há muito tempo se discute os problemas da evasão nos cursos de Educação a Distância (EaD) do Brasil, principalmente nas Instituições públicas de ensino.

Este é um tema tão abordado e fomentado em diversos contextos históricos, que devemos abordar com maior cautela e tratar como assunto primordial dentro da educação. De acordo com Censo EaD de 2018, as taxas de evasão da EaD ainda são superiores às taxas dos cursos presenciais, temos 5% do total de instituições com taxas entre 0 e 5%, 5% das instituições com taxas entre 6 e 10%, e somente 1% das instituições registraram taxas de mais de 50% (CENSO EAD 2017).

A evasão está presente em todas as modalidades de ensino, seja presencial, semipresencial ou a distância (BITTENCOURT e MERCADO, 2014). No contexto atual, temos diversos trabalhos que buscam identificar as causas da evasão ou até mesmo, que proponham metodologias e ferramentas para mitigá-las.

Nesse contexto, estão inseridos os Ambientes Virtuais de Aprendizagem (AVAs), estes os responsáveis pela interação do aluno e professor, sendo possível compartilhar materiais, realizar tarefas, interagir com outros alunos, participar de fóruns e Wikis e outras atividades inerentes ao processo de aprendizagem EaD.

Entretanto, dentre as diversas interações entre aluno e professor através da ferramenta, ainda se encontram dificuldades para acompanhar o desempenho do aluno em cursos oferecidos de forma virtual, ou seja, pelo AVA.

Desta forma, para que a tecnologia possa auxiliar na tomada de decisão e verificar os fatores que têm influenciado na evasão, com base nos registros de dados dos alunos no sistema, se faz necessário recorrermos a uma técnica chamada Mineração de Dados (MD).

Sendo assim, a mineração surge como excelente alternativa, pois possibilita identificar quais os fatores que contribuem para o insucesso dos estudantes dentro dos ambientes virtuais de aprendizagem. De forma a entendermos melhor sobre a definição de MD, dizemos que ela consiste na exploração de grandes quantidades de dados com o objetivo de detectar padrões que permitem a extração de novos conhecimentos (SILBERSCHATZ, KORTH e SUDARSHAN, 2006).

2. TRABALHOS CORRELATOS

Esta seção busca apresentar alguns dados de pesquisas relacionadas à evasão e previsão de desempenho, bem como a análise dos trabalhos mencionados. A busca foi realizada através do Google Acadêmico utilizando termos como 'Mineração de Dados Educacionais', 'Previsão de Desempenho Acadêmico', 'AVEA', "AVA" e 'Evasão EaD".

2.1 DESCRIÇÃO DOS TRABALHOS

Rabelo et al. 2017, tomando como base as interações armazenadas no repositório do ambiente de pesquisa MOODLE foram encontradas 64 ações associadas a 24 módulos (atributos action e module), realizadas por 514 usuários com perfil de aluno em 13 Turmas de Cursos Graduação da UFRN, perfazendo um montante de 11.310.848 interações armazenadas no log do MOODLE. Oito indicadores de desempenho foram selecionados: Ação login de usuário (login); Ação de visualização do Curso (view); Ação de visualização de Recursos do Curso (view); Ação de Visualização de discussão em fórum (view discussion); Ação de Adição de postagem em fórum (add post); Ação de visualização de tarefa – View; Ação de enviar tarefa (Submit) e Ação de responder questionário (Attempt). Para aplicar as Técnicas de Mineração de Dados, fez-se o emprego da ferramenta computacional Weka. A técnica de classificação empregada neste trabalho é a de "árvore de decisão". Utilizou-se dois algoritmos de classificação baseados em árvores de decisão, ID3 e J48. O melhor caso foi obtido através do algoritmo J48, com 96,5% de acertos e 3,5% de erros de classificação. O algoritmo J48 obteve acurácia de 96,5%, classificando corretamente o desempenho de 496 dos 514 alunos.

Burgos et al. (2017) propõem a utilização de mineração de dados e a criação de modelos de predição utilizando regressão logística linear para prever o risco de evasão de alunos. Foram utilizados dados de 104 alunos de diversos cursos na modalidade a distância. Os resultados apresentam valores de até 100% de acurácia geral já na quarta semana do curso. Segundo os autores, a aplicação desta técnica junto a um plano de tutorial diminuiu em 14% o abandono escolar nos cursos em que foram aplicados.

Queiroga et al. (2018), apresentam uma proposta de metodologia para geração de modelos de predição de alunos em risco de evasão que utiliza a contagem de interações. Para o desenvolvimento da proposta foram utilizados dados de quatro cursos técnicos na modalidade a distância do Câmpus Visconde de Graça (CaVG), do Instituto Federal Sul-rio-grandense (IFSul). Os autores utilizaram a contagem de interações dos estudantes no AVA como a principal informação para a geração dos modelos de predição. Para avaliação dos modelos foram utilizados cinco diferentes algoritmos: Bayes Net, Simple Logistic, Multilayer Perceptron, Random Forest e J48. A biblioteca utilizada foi o WEKA. Desde a primeira semana as taxas de acerto foram superiores a 75% em quase todos os algoritmos, excetuando-se os resultados obtidos pelo Simple Logistic. Com o passar das semanas as taxas de acerto se elevam para 87%. No segundo semestre do curso são obtidos resultados próximos a 94% antes do final do mesmo, sendo possível afirmar com quase 95% de exatidão se um aluno irá terminar o curso antes do final do primeiro ano.

Detoni, Araujo e Cechinel (2015), buscaram detectar os alunos por meio da utilização da contagem de interações no AVEA MOODLE em disciplinas na EaD. O experimento foi realizado na Universidade Federal de Pelotas (UFPEL), nos cursos de Licenciatura em

Educação do Campo e Licenciatura em Pedagogia. A partir das interações foram calculadas as médias de utilização do AVEA por parte dos alunos, tutores e professores e, posteriormente, foram aplicados os algoritmos de Redes Bayesianas, Redes Neurais, J48 e RandomForest, onde foram obtidos resultados de até 67% de acurácia na predição do desempenho do aluno.

Neste mesmo sentido, Rodrigues et al. (2013) têm como objetivo 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. Este trabalho utilizou para análise dos dados o software estatístico R. Os autores afirmam, com uma confiança de 95%, que 79,2% dos alunos que interagem via fórum de discussão, à medida que aumenta a quantidade de interação, tende a aumentar o desempenho acadêmico.

Já Da Costa et al. (2014) buscam identificar o perfil de alunos com potencial de evasão. Para isto, utilizam as notas e o número de interações no ambiente de cursos de especialização à distância de cada aluno. Para uso da ferramenta WEKA, estes dados foram transformados e posteriormente 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. Os dados utilizados foram de 98 alunos de um curso semipresencial, extraídos do banco de dados do AVA MOODLE. Os atributos utilizados são: desempenho final (nota) e número de interações. Para analisar o desempenho do perfil dos alunos, foram utilizados 7 algoritmos, onde o J48 obteve o melhor desempenho, alcançando 74% de acurácia.

Schimitt (2018) buscou apresentar aos gestores dados estratégicos por meio da mineração de dados educacionais, de modo que pudessem avaliar e gerar ações para mitigar a evasão na Universidade Federal de Santa Maria (UFSM). Sua pesquisa abrangeu dois experimentos, em três cursos de graduação, onde foram empregados dados de interações dos alunos no AVEA e dados do sistema de gestão acadêmico da UFSM. Para o primeiro experimento, foi utilizado exclusivamente os dados provenientes de logs do AVEA MOODLE, sendo suas simulações efetuadas com dois formatos de dados, não normalizados e normalizados. Foram aplicados sete algoritmos diferentes, abrangendo dados sobre as interações dos alunos do curso de Letras – Espanhol. Todas as simulações realizadas com dados não normalizados apresentaram bons resultados podendo-se observar que o melhor deles foi obtido pelo algoritmo Naive Bayes com 58,39% de Medida-F. No entanto, algoritmos como o próprio Naive Bayes, Random Forest e IBk obtiveram respectivamente 51,59%, 54,34% e 43,56% de Medida-F para o formato normalizado.

3. ANÁLISE DOS TRABALHOS

Ao analisar os resultados, observa-se uma maior utilização de algoritmos de Árvore de Decisão. Outro ponto observado na pesquisa é as ferramentas computacionais mais utilizadas. Neste processo de descoberta de conhecimento, o Weka foi a de maior utilização nos trabalhos lidos e pesquisados.

Os objetivos das pesquisas são bem variados. As investigações abrangem estudos de desempenho em fóruns, em disciplinas específicas, em cursos, na graduação à distância e presencial. 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. Também, vários trabalhos têm como foco a busca pelos melhores atributos.

Dentre as pesquisas, percebe-se a constante busca pelos motivos que levam os alunos a evadirem nas diferentes modalidades de ensino, no entanto, é visível a falta de pesquisa nos cursos técnicos, foco da proposta de dissertação.

A utilização em larga escala do ambiente MOODLE ficou evidente, sendo a maior fonte de dados para as pesquisas. No que tange a classificação dos algoritmos com melhor taxa de acerto, o J48 é um dos mais citados pelas pesquisas, obtendo altas taxas de acertos na maioria delas.

Para melhor sintetizar as diferenças entre os trabalhos analisados, são apresentadas as características de cada trabalho no Quadro 1.

Quadro 1 - Análise dos trabalhos correlatos

Autor	Modalidade	Técnica/Algoritmos utilizados	Ferramenta de KDD	Técnica de mineração utilizada	Atributos utilizados
Rabelo et. al (2017)	Curso de graduação - EaD	ID3 e J48	WEKA	Classificação	Login; Visualização do curso, recursos, discussão em fórum, tarefas; Adição de postagem; Questionário
Burgos et. al (2017)	Cursos EaD – Não especificado a modalidade	FEED Forward Neural Network (FFNN), Support Vector Machine, ARTMAP	Não especificada.	Regressão	Contagem de interações no AVEA.
Queiroga, Lopes, Araujo, Cechinel (2018)	Cursos técnicos EaD	Bayes Net, Simple Logistic, Multilayer Perceptron, Random Forest e J48	WEKA	Classificação	Contagem de interações no AVEA
Detoni, Araújo e Cechinel (2015)	Curso de graduação EaD	Redes Bayesianas, Redes Neurais, J48 e Random Forest	WEKA	Classificação	Contagem de interações nas disciplinas no AVEA.
Rodrigues et al. (2013)	Curso de graduação EaD	Não especificado	Estatístico R2	Regressão	Modelo de Regressão Linear aplicado à previsão de desempenho de estudantes em ambiente de aprendizagem.
Da Costa et al. (2014)	Curso de especialização EaD	J48	WEKA	Classificação	Nota; Contagem de acesso ao ambiente.
Santana, Maciel e Rodrigues (2014)	Curso semipresencial	Random Forest, Multilayer Perceptron (MLP), Naive Bayes, SVM, KNN, J48 e RBF	WEKA	Classificação	Nota; Número de acesso ao fórum; Interações e tempo médio de acesso ao ambiente
Schimitt (2018)	Curso de graduação EaD	J48, Naive Bayes, MultiLayer Perceptron, Random Forest, OneR, SMO e IBk.	WEKA	Classificação, Precisão, Revocação e Medida-F	Contagem de interações. (Fórum, tarefas, materiais, links, pasta, wiki, questionário)

Fonte: Próprio autor (2019).

REFERÊNCIAS

- BITTENCOURT, I. M.; MERCADO, L. P. L. Evasão nos cursos na modalidade de educação a distância: estudo de caso do Curso Piloto de Administração da UFAL/UAB. **Revista Ensaio: Avaliação de Políticas Públicas em Educação**. Rio de Janeiro, v. 22, n. 83, p. 465-504, abr./jun. 2014.
- BURGOS, C. et al. Data mining for modeling students' performance: A tutoring action plan to prevent academic dropout. **Computers & Electrical Engineering**. p. 1-16, mar. 2017.
- Censo EAD.BR 2017 - Relatório Analítico da Aprendizagem a Distância no Brasil**. São Paulo, 2017.
- 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**. RENOTE.
- DETONI, D; ARAÚJO, R.; CECHINEL, C. Modelling and Prediction of Distance Learning Students Failure by using the Count of Interactions. **Revista Brasileira de Informática na Educação**, v. 23, p. 1-11, 2015.
- QUEIROGA, E.; CECHINEL, C.; ARAÚJO, R. Predição de estudantes com risco de evasão em cursos técnicos a distância. In: **Anais do XXVIII Simpósio Brasileiro de Informática na Educação (SBIE 2017)**. Recife: Sociedade Brasileira de Computação, 2017. v. 1, p. 1547–1556.
- RABELO, H. et al. **Utilização de técnicas de mineração de dados educacionais para predição de desempenho de alunos de EaD em ambientes virtuais de aprendizagem**. n. Cbie, p. 1527, 2017.
- 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**.
- SANTANA, L. C.; MACIEL, A. M.; RODRIGUES, R. L. Avaliação do perfil de uso no ambiente moodle utilizando técnicas de mineração de dados. In: **Anais do Simpósio Brasileiro de Informática na Educação**. [S.l.: s.n.], 2014. v. 25, n. 1, p. 269.
- SCHMITT, Jäder Adiel. **Identificação de alunos com tendência a evasão nos cursos de graduação a distância por meio de mineração de dados educacionais**. 2018. 175 p. Dissertação (Mestrado Profissional em Tecnologias Educacionais em Rede) – Universidade Federal de Santa Maria, Santa Maria, 2018.
- SILBERSCHATZ, Abraham; KORTH, Henry F.; SUDARSHAN, S. **Sistema de Banco de Dados**. Tradução de Daniel Vieira. Rio de Janeiro: Elsevier, 2006.