

## Artigo Original

# Proposta de melhoria dos dados de relatórios de uma plataforma de MOOCS brasileira

*Proposal to improve reporting data for a brazilian MOOC platform*

Vanessa Faria de Souza<sup>1</sup> e Gabriela Trindade Perry<sup>2</sup>

## Resumo

A maior parte dos Ambientes Virtuais de Aprendizado (AVAs) não foi projetada para fazer registros de navegação com foco na mineração de dados, mas sim, para informar professores e alunos sobre objetivos pedagógicos. Contudo, vem ganhando relevância o entendimento que as informações que podem ser extraídas dos registros de navegação podem ser bastante relevantes, o que motiva o vigoroso interesse na área de mineração de dados educacionais. Nesse sentido, neste artigo, analisa-se como os dados coletados na plataforma Lúmina, AVA baseado em Moodle e que disponibiliza *Massive Open Online Courses* (MOOCs) da Universidade Federal do Rio Grande do Sul (UFRGS), podem ser melhorados, permitindo que mais informações sobre os estudantes sejam extraídas. Como modelo desta análise, usou-se o repositório DataShop do *Pittsburgh Science of Learning Center* (PSLC), que foi projetado a princípio para fornecer recursos educacionais online, e que tem apresentando bons resultados na aplicação de processos de mineração. O objetivo final é propor adequações à plataforma da UFRGS.

---

<sup>1</sup> Programa de Pós-Graduação em Informática na Educação – Universidade Federal do Rio Grande do Sul – Av. Paulo Gama, 110, Farroupilha – Porto Alegre – RS – Brasil. [vanessa.souza@ibiruba.ifrs.edu.br](mailto:vanessa.souza@ibiruba.ifrs.edu.br)

<sup>2</sup> Programa de Pós-Graduação em Informática na Educação – Universidade Federal do Rio Grande do Sul – Av. Paulo Gama, 110, Farroupilha – Porto Alegre – RS – Brasil.

**Palavras-Chave:** mineração de dados educacionais, análise de dados educacionais, ambientes virtuais de aprendizagem, cursos online massivos e abertos.

## Abstract

Most Virtual Learning Environments (VLEs) are not designed to make navigation records focused on data mining, but rather to inform teachers and students about pedagogical goals. However, the understanding that the information that can be extracted from the navigation records can be very relevant is gaining relevance, which motivates the vigorous interest in the area of educational data mining. In this sense, this article analyzes how data collected on the Moodle-based Lúmina platform, AVA, which provides Massive Open Online Courses (MOOCs) from the Federal University of Rio Grande do Sul (UFRGS), can be improved, allowing more information about students being extracted. As a model for this analysis, we used the Pittsburgh Science of Learning Center (PSLC) DataShop repository, which was originally designed to provide online educational resources and has been successful in applying the mining process. The goal is to propose adjustments to the UFRGS platform.

**Keywords:** educational data mining, educational data analysis, virtual learning environments, massive and open online courses.

## I. Introdução

Com a ampla difusão das tecnologias da informação e da comunicação (TIC), diversas instituições de ensino têm explorado as possibilidades da educação superior a distância e, as modalidades já consolidadas vão aos poucos compartilhando espaços com os Ambientes Virtuais de Aprendizagem (AVA). Nesse âmbito, os *Massive Open Online Courses* (MOOCs) despontam como uma modalidade de cursos abertos para qualquer pessoa, sem acompanhamento por tutores ou professores, geralmente de forma gratuita e sem exigência de conhecimentos prévios. Grande parte desses cursos é oferecido por renomadas instituições de ensino e disponibilizados em plataformas que reúnem cursos em diversas áreas do conhecimento.

Evidencia-se, que essa configuração de curso tem a potencialidade para modificar a educação para os alunos, professores e gestores educacionais, visto que tem competência para oferecer às instituições de ensino grande visibilidade, por meio do acesso a um público global de alunos a uma pequena parcela do custo da educação tradicional, bem como incluir e/ou alterar novos conhecimentos a um curso e reciclar conteúdos antigos (BASTOS e BOTTENTUIT, 2017). Na percepção de Bastos e Bottentuit (2017) quer se julgue que este modelo é exitoso ou não, os cursos em massa transformaram o cenário educacional.

Diante da expressividade da Educação a Distância e dos MOOCs, muitas instituições educacionais fazem uso de AVAs, que são plataformas que suportam a distribuição dos recursos educacionais, a interação (entre alunos, professores e materiais), a organização dos tópicos e do andamento do curso, bem como atividades de avaliação. Por disponibilizarem tantas funcionalidades, costumam produzir um grande volume de dados, os quais são armazenados por esses sistemas sob a forma de relatórios (também conhecidos como “logs”).

Nesse cenário, tem crescido nos últimos anos a intenção em utilizar a mineração de dados em outras áreas, como a educacional (BAKER, ISOTANI e CARVALHO, 2011; SWEENEY et al., 2016), configurando o conjunto de algoritmos, técnicas e procedimentos reconhecidos como Mineração de Dados Educacionais (MDE), cujo objetivo geral

está alinhado com a descoberta, criação e utilização de modelos e padrões sobre os conjuntos de dados coletados pelos AVAs. Contudo, como salientam Krüeger, Merceron e Wolf (2010) muitos AVAs não foram projetados para análise e mineração dados. Mesmo que muitas destas plataformas armazenem dados de uso, seu objetivo é o de apoiar o ensino e o aprendizado, não analisar ou minerar dados. Contudo, o campo da MDE tem avançado precisamente porque valiosas informações são obtidas com a análise e mineração de dados armazenados pelos AVAs (ROMERO e VENTURA, 2007; BAKER e YACEF, 2009). Como consequência, o processo de MDE requer um longo pré-processamento (MACERON e YACEF, 2008).

Por esse ponto de vista, e motivados por estudos realizados no ambiente Lúmina, Perry et al (2019); Souza e Perry (2019); Souza et al (2019); Souza, Wives e Perry et al (2018), plataforma de MOOCs da Universidade Federal do Rio Grande do Sul, percebeu-se que os dados extraídos para realização da MDE não são adequados para a identificação de padrões de uso, modelagem de perfis de estudantes e análises exploratórias e inferenciais. Por este motivo se fazem necessárias melhorias na coleta das interações entre usuário e a plataforma.

Diante deste desafio - melhorar bases de dados para MDE - pesquisadores do Centro de Ciências e Aprendizagem de Pittsburgh (PSLC – *Pittsburgh Science of Learning Center*), criaram o Datashop, um repositório de dados citado por Koedinger et al. (2010) e Baker, Isotani e Carvalho (2011) como sendo completo e capaz de gerar informações e conhecimentos relevantes sobre alunos e também sobre metodologias de aprendizado. Por este motivo, o Datashop será usado como modelo de excelência.

## 2. O PSLC DATASHOP 508538

O DataShop<sup>3</sup>, do *Pittsburgh Science of Learning Center* (PSLC) é um repositório aberto de dados com interações entre estudantes e softwares educacionais (desde que suportem o seu esquema de dados XML).

---

<sup>3</sup> <http://pslcdatashop.org/>

Foi criado como parte do *Pittsburgh Science of Learning Center*, fundado, em 2004, como parte do “*National Science Foundation’s Science of Learning Center*”. Segundo Koedinger et al. (2010) este repositório teve um papel importante na mineração de dados educacionais nos últimos 10 anos (2008-2018).

Os conjuntos de dados nele disponíveis foram usados em 14% dos artigos publicados nos dois primeiros anos (2008 e 2009) da conferência Internacional sobre Mineração de Dados Educacionais (BAKER e YACEF, 2009). O repositório continua a desempenhar um papel importante desde então, e seus dados já foram usados em mais de 100 artigos (KOEDINGER et al., 2009).

Ele agora contém quase 100 milhões de interações entre estudantes e softwares educacionais, que ocorreram durante mais de 200.000 horas de interação. Tem apoiado pesquisas em várias áreas, incluindo pesquisa sobre o domínio da estrutura do conhecimento (KOEDINGER, et al. 2013), comparação entre diferentes abordagens de modelagem de alunos (GOWDA et al., 2011), estudo das diferenças de conteúdo ao longo de um ano letivo (BAKER et al. 2009) e o desenvolvimento de modelos de novas construções de aprendizagem (ALEVEN et al. 2006).

Este repositório fornece dois serviços principais para a comunidade científica de aprendizagem: (i) um repositório central para proteger e armazenar dados de pesquisa e (ii) um conjunto de ferramentas de análise e relatórios. Os pesquisadores podem acessar rapidamente os relatórios padrão, como curvas de aprendizado, bem como navegar pelos dados usando um aplicativo *Web* interativo. Para dar suporte a outras análises, o pesquisador também pode exportar dados para o formato “csv” (*comma separated value*) compatível com softwares estatísticos e outros pacotes de análise.

O DataShop tem ligações com dois projetos na área de MDE e Sistemas de Tutores Inteligentes da Universidade de Carnegie Melon: CTAT<sup>4</sup> (*Cognitive Tutor Authoring Tools*) e OLI<sup>5</sup> (*Open Learning*

<sup>4</sup> <http://ctat.pact.cs.cmu.edu/>

<sup>5</sup> <http://oli.cmu.edu/>

*Initiative*), que geram seus registros de interação no formato do DataShop, um esquema XSD (*XML Schema Definition*) chamado *Tutor Message Format*, que permite que registros de sistemas de terceiros possam fazer uso das ferramentas disponíveis.

Esse esquema contém quatro tipos de “mensagens”: contexto, que estabelece o contexto das mensagens seguintes; ferramenta, que descreve a interação com a ferramenta (sistema); tutor (software), resposta do tutor ao estudante; mensagem, para outros tipos de comunicação não adequadamente descritas. A documentação<sup>6</sup> deste formato contém explicações detalhadas sobre cada um dos elementos e atributos (XML) que podem ser inseridos em cada uma das mensagens.

O repositório contém dados com três atributos que o tornam particularmente útil para MDE (KOEDINGER, et al. 2013): os dados são refinados, do menor tamanho que sejam semanticamente significativos, “transações” entre o aluno e o software, incluindo a ação do aluno e a resposta do software são armazenados. Os dados são longitudinais, são captados dados durante um longo tempo de interação, envolvendo o comportamento e a aprendizagem do aluno, em muitos casos durante o semestre ou o ano inteiro de estudo. Os dados são extensos, muitos dados são mantidos por aluno, envolvendo milhões de transações para alguns dos pacotes para os quais o DataShop possui dados. Essas três características tornaram o DataShop útil para muitos mineadores de dados educacionais.

## 2.1 Pesquisas que utilizaram os dados em *Tutor Message Format*

Muitas análises de dados no formato deste esquema foram realizadas em uma variedade de domínios, como, por exemplo, investigações em leitura (LESZCZENSKI e BECK, 2007), física (NWAIGWE et al. 2007) e geometria (RAFFERTY e YUDELSON, 2007). Estas análises têm objetivo de encontrar maneiras de avaliar ou melhorar o aprendizado dos alunos, usando o *Cognitive Tutor*.

---

<sup>6</sup> <https://pslcdatashop.web.cmu.edu/dtd/guide/index.html>

Cen, Koedinger e Junker (2007) realizaram uma análise da curva de aprendizado usando os fatores de aprendizagem em algoritmos de análise, baseado em dados da unidade do tutor cognitivo de geometria. Eles notaram que, embora os alunos precisassem praticar alguns objetivos fáceis de conhecimento, eles praticavam pouco habilidades mais complexas. Com base em observações e análises adicionais, eles criaram uma nova versão do tutor de geometria, redefinindo parâmetros que determinam com que frequência as habilidades são praticadas. Eles fizeram um experimento em sala de aula, no qual, os alunos de um curso foram pré e pós-testados e designados aleatoriamente para usar a versão anterior ou a nova versão do tutor. Alunos que usaram a nova versão levaram 20% menos tempo para concluir as mesmas unidades curriculares. O resultado mais relevante dessa pesquisa foi o refinamento de um modelo de aprendizagem de máquina denominado *Learning Factors Analysis*.

O trabalho de Baker et al. (2006) também envolveu MDE, estes fizeram observações do comportamento dos alunos em laboratórios de informática enquanto usavam o sistema de tutoria inteligente Cognitive Tutor, durante as aulas de um professor de matemática do ensino médio. Os autores usaram técnicas de aprendizado de máquina para criar um agente capaz de processar o registro de informações do aluno em tempo real, para determinar quando os alunos estavam tentando enganar o sistema. O agente tornou-se base para um sistema de intervenção, um “metatutor”, projetado para desencorajar este comportamento e envolver os alunos em instrução suplementar sobre os tópicos que eles haviam tentado não estudar.

Baker e Gowda (2010) utilizaram dados do *Cognitive Tutor* para comparar escolas dos EUA com 3 perfis diferentes (urbano, suburbano e rural), de acordo com 3 métricas (porcentagem de tempo fora da tarefa, porcentagem do tempo gasto jogando e probabilidade média de desvio). Essa análise foi realizada aplicando detectores automatizados de comportamentos aos dados de alunos, que usaram o Cognitive Tutor em aulas de geometria do ensino médio, durante todo o ano letivo. Concluíram que os alunos da escola urbana ficam mais tempo sem realizar a tarefa e são mais descuidados do que os alunos das escolas

rurais e suburbanas. Essas informações sugerem que algumas das diferenças no desempenho por tipo de escola podem resultar de problemas comportamentais de engajamento.

Usando também ferramentas de análise do DataShop, Koedinger e Stamper (2010) propuseram um modelo cognitivo (i.e., um modelo consistente com o comportamento do estudante, que consegue prever o grau de dificuldade da tarefa e o desempenho em cada teste). Também os autores Rajni e Malaya (2014) utilizaram um conjunto de dados do DataShop, com o objetivo de medir o desempenho dos alunos, crenças, interesses em determinados assuntos, e prever o engajamento no ambiente de aprendizagem usando Regras de Associação. Da mesma forma, os pesquisadores Corbi e Burgos (2015) realizaram um estudo no qual aplicaram um modelo de recomendação, construído com base em um conjunto dados do *Cognitive Tutor*, este modelo de recomendação fornece suporte a estudantes, tutores e professores.

Esses são alguns dos trabalhos que utilizaram dados no formato *Tutor Message*, armazenados no PSLC DataShop como fonte de dados. Com o aumento da procura e oferta de MOOCs, que inovou de forma considerável a produção de dados em meios *online* de aprendizagem, o foco na utilização de grandes massas de dados se volta para plataformas como edX, Coursera, MitX, Veduca, Udacity.

### 3. Comparação do Lúmina com o PSLC Datashop

Como salientado na introdução, os AVAs não foram criados para o armazenamento de dados com vistas à mineração, contudo, aprimorar essa função no Lúmina é importante, pois, conhecer melhor os usuários da plataforma, pode implicar em avanços significativos dos cursos disponibilizados.

O Lúmina (uma instalação do Moodle) tem um grande potencial de armazenamento de dados atualmente não explorado, pois a plataforma somente armazena dados de pesquisas, questionários, textos que são postados nos fóruns, mensagens e comentários, e datas de realização

de cada interação. Há diversas ações que não são registradas: interações com vídeos e textos e rolagem de página. Há ainda elementos que são difíceis de configurar e integrar aos cursos, como ferramentas de trabalho colaborativo (como wikis, que poucos estudantes sabem usar) e de revisão por pares. Além disso, faltam formatos que permitam a proposição de atividades avaliativas que não resultem em produção de texto ou de questões de múltipla escolha, como galerias de imagens.

Por este motivo, atualmente, o conjunto de dados do Lúmina é limitado, o que torna ineficaz a aplicação de algoritmos de MDE. Deste modo, readequações seriam de grande valia para melhorar o entendimento que se tem sobre os estudantes usuários da plataforma. Por isso, a importância da comparação com repositórios que tem sucesso na aplicação de técnicas de MDE. A Tabela 01 realça a comparação entre o Lúmina e do PSLC DataShop.

**Tabela 01:** Comparação PSLC DataShop e Lúmina

CATEGORIA DE COMPARAÇÃO	DATASHOP	LÚMINA	ADEQUAÇÕES
CARACTERÍSTICAS DOS DADOS	Transações dos alunos – O histórico de interação aluno-computador em um nível mais refinado, “todas as interações como cliques e rolagem de páginas”	Implementa, mas para ações menos refinadas, como uma atividade ou questionário	Aumentar o escopo de dados armazenados
	Longitudinais – dados de períodos longos de interações	Implementa	-
	Extensos – muitas transações aluno-computador, todas as interações do aluno	Implementa, mas como os dados mantidos são pouco refinados, não são tão extensos	Aumentar o escopo de dados armazenados
	Possui uma definição de tipo de documento no formato XML para os dados mantidos	Não possui, gera relatórios em outros formatos	Implementar adaptações para manutenção de dados compatíveis com padrões internacionais

DADOS ARMAZENADOS PARA CADA PROBLEMA/TESTE	Sessão	Implementa	-
	Hora	Implementa	-
	Nome do problema/ teste	Implementa	-
	Número da tentativa	Implementa	Armazenar para cada transação do aluno
	Tipo de entrada	Implementa	Implementar diferentes tipos de ações dos alunos e armazenar
	Reposta do software/ AVA	Implementa	Implementar interações da plataforma com o aluno e armazenar
	Número da dica – se houve solicitação	Não fornece dicas	Implementar recurso de dicas
	Condições atribuídas a etapa do problema/ teste	Não implementa problemas ou testes em etapas, nem condicionantes a estes	Implementar condicionantes, dependendo do curso ofertado
DADOS ARMAZENADOS PARA CADA ATIVIDADE	Clique em segmento de áudio e vídeo. Marcador do tempo que ouviu ou assistiu ao recurso	Implementa, mas não possibilita verificar se o aluno ouviu ou assistiu o recurso até o final	Implantar estrutura para verificar
	Navegação pelas páginas de conteúdo on-line do curso, marcador de Scroll das páginas	Implementa, mas não possibilita verificar se aluno rolou a página até o final	Implantar estrutura para verificar
	Clique em baixar material	Implementa, mas não permite verificar se o aluno baixou o material	Implantar marcador para verificar
	Ações de log sem realização de atividades	Implementa	-
RESPOSTAS ARMAZENADAS DO SOFTWARE/AVA	Respostas do software quanto as ações dos alunos – como feedbacks de acertos e erros	Implementa	Sugestão - mais interatividade da plataforma com alunos

FERRAMENTAS DE ANÁLISE	Mapeamento entre etapas de problemas propostos – verificando aumento de complexidade	Não implementa – verifica-se o aumento na complexidade, mas não mapeia	Realização de mapeamento de problemas/teste
	Informações do conjunto de dados: fornece métricas do conjunto de dados	Implementa – gera relatórios	-
	Relatório de erros: apresenta a primeira tentativa de cada aluno em um problema/teste	Implementa – os dados ficam expressos nos relatórios	-
	Performance Profiler: ferramenta multiuso que visualiza o desempenho do aluno	Não implementa	Implementação de ferramenta de comparação da evolução dos alunos
	Curva de aprendizado: visualiza as mudanças de aprendizado dos alunos ao longo do tempo	Não implementa	Implementação de ferramenta de comparação da evolução dos alunos

Fonte: Autoras

Como pode ser observado na Tabela 01, o principal gargalo para MDE no Lúmina é que a plataforma não armazena informações de todas as ações que o usuário pode executar, todas as transações devem ser armazenadas como ocorre no DataShop. Apenas informações de atividades são mantidos na plataforma, o que é insuficiente para extração de conhecimentos mais aprofundados sobre os usuários. Desta forma, não é possível verificar o quanto o aluno está engajado no curso, se ele realmente lê os materiais, assiste as videoaulas, enfim, como é seu processo de aprendizagem.

Outras sugestões quanto a comparação com o DataShop, que podem ser implantadas no Lúmina se o professor ofertante propuser, seriam a inserção de um pré e um pós teste para analisar a evolução dos alunos, e a implementação do componente de conhecimento<sup>7</sup>. Em cada etapa

<sup>7</sup> <http://www.learnlab.org/research/wiki>

de um problema, os tutores inteligentes (softwares) exigem que o aluno saiba algo - um conceito ou habilidade relevante - para executar a etapa corretamente, essa pequena unidade é denominada “componente de conhecimento”, esta unidade codifica uma capacidade geral do aluno para realizar determinadas etapas da tarefa, esses componentes poderiam ser utilizados na geração de métricas relevantes.

## 4. Considerações finais

Como salientado por Krüeger, Merceron e Wolf (2010), muitos AVAs não foram projetados para análise e mineração de dados - eles armazenam dados, mas não foram projetados com essa finalidade, e sim para apoiar o ensino e o aprendizado. Contudo, a MDE tem avançado justamente devido aos dados armazenados nestas plataformas, por isso, a melhoria em seus processos de armazenamento é um avanço científico que precisa ser levado em consideração. Nesta perspectiva, este trabalho teve como objetivo realizar uma comparação do PSLC Datashop com o Lúmina, salientando pontos que este último poderia ser melhorado.

O DataShop foi criado para atender muitos anseios da comunidade acadêmica dos EUA, primeiramente, como uma plataforma que aloca tutores inteligentes, disponibilizando recursos online de aprendizado bastante avançados. A partir dessa percepção, o DataShop foi aperfeiçoado para, além de disponibilizar recursos de aprendizagem, também armazenar dados para realização de pesquisas, e assim, produzir conhecimentos novos. Para tanto, os dados mantidos pelo repositório são refinados, longitudinais e extensos, requisitos indispensáveis ao processo de aplicação dos algoritmos de aprendizagem de máquina, que precisam de muitos dados de qualidade para poderem aprender.

Como descrito anteriormente, os dados extraídos do Lúmina são insuficientes para a aplicação da MDE, não possibilitam alcançar maior entendimento sobre os alunos matriculados em seus cursos, o que dificulta identificar quais motivos levam os alunos a permanecer ou evadir um curso, ou se o nível de dificuldade está adequado, por exemplo. Melhorias na forma de registro dos dados e na quantidade dos atributos mantidos são fatores que podem melhorar a qualidade dos dados gerados.

A análise do formato de registro do PSLC DataShop possibilitou destacar alguns requisitos a serem implementados, que causariam um avanço para a plataforma Lúmina, no que tange a mineração de dados educacionais que são: manutenção de um histórico de interação aluno-computador em um nível mais refinado, ou seja, armazenar ações como, por exemplo, clicar em recursos, mesmo que sejam ações pequenas. Também, armazenar dados que possam fornecer informações quanto à visualização de vídeos e dos materiais de apoio, pois é importante saber se o aluno assistiu ou leu um material, deste modo, pode-se analisar seu envolvimento com o curso.

Em suma, evidencia-se que o Lúmina deve ser capaz de armazenar as interações com todos os seus recursos. A inserção destes atributos aumenta o conjunto de dados que podem ser minerados e, consequentemente, melhora a compreensão sobre como os alunos aprendem, quais fatores influenciam em sua permanência no curso e, principalmente, ajuda na percepção de quais alunos são mais propensos a evadir. Se os algoritmos de MDE forem bem treinados, logo nas primeiras semanas do lançamento de um curso, é possível analisar os dados de alunos e identificar aqueles inclinados a desistir e ações individuais podem ser implementadas, já que promover essas ações para todos os alunos de um MOOC é inviável.

## Referências bibliográficas

ALEVEN, V.; MCLAREN, B.; ROLL, I.; KOEDINGER, K. R. Toward Meta-Cognitive Tutoring: A Model of Help Seeking with a Cognitive Tutor. In: *Int' J. Artificial Intelligence and Education*, vol. 16, no. 2, 2006, p. 101–128. Disponível em [<https://content.iospress.com/articles/international-journal-of-artificial-intelligence-in-education/jai16-2-02>]

BAKER, R.; CORBETT, A.; KOEDINGER, K. R.; EVENSON, S.; ROLL, I.; WAGNER, A.; NAIM, M.; RASPAT, J.; BAKER, D.; BECK, J. *Adapting to when students game an intelligent tutoring system*. M. Ikeda, K. D. Ashley, T.-W. Chan (Eds.). In: 8th International Conference on Intelligent Tutoring Systems, 2006, p. 392-401. Disponível em [[https://link.springer.com/chapter/10.1007/11774303\\_39](https://link.springer.com/chapter/10.1007/11774303_39)]

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. 03-13, 2011. Disponível em [<http://dx.doi.org/10.5753/rbie.2011.19.02.03>]

BAKER, R. S. J. D.; CARVALHO, A. M. J. B.; RASPAT, J.; ALEVEN, V.; CORBETT, A. T.; KOEDINGER, K. R. *Educational Software Features that Encourage and Discourage “Gaming the System”*. In: 14th Int’ Conf. Artificial Intelligence in Education, 2009, p. 475–482. Disponível em [<https://dl.acm.org/doi/10.5555/1659450.1659522>]

BAKER, R. S. J. D.; CORBETT, A. T.; ALEVEN, V. *More Accurate Student Modeling Through Contextual Estimation of Slip and Guess Probabilities in Bayesian Knowledge Tracing*. In: 9th International Conference on Intelligent Tutoring Systems, 2008, p. 406-415. Disponível em [[https://link.springer.com/chapter/10.1007/978-3-540-69132-7\\_44](https://link.springer.com/chapter/10.1007/978-3-540-69132-7_44)]

BAKER, R. S. J. D.; GOWDA, S. M. *An Analysis of the Differences in the Frequency of Students’ Disengagement in Urban, Rural, and Suburban High Schools*. In: International Conference on Educational Data Mining, 3rd, Pittsburgh, PA, USA, 2010, p. 11-20. Disponível em [[https://educationaldatamining.org/EDM2010/uploads/proc/edm2010\\_submission\\_25.pdf](https://educationaldatamining.org/EDM2010/uploads/proc/edm2010_submission_25.pdf)]

BAKER, S.J.D.R.; YACEF, Y. The State of Educational Data Mining in 2009: A Review and Future Visions. *JEDM – Journal of Educational Data Mining*, 1(1), p. 3-17, 2009. Disponível em [<https://jedm.educationaldatamining.org/index.php/JEDM/article/view/8>]

BASTOS, I. M. M.; BOTTENTUIT, J. B. Jr. Os Caminhos da Aprendizagem Virtual Aberta: Apresentação de Plataformas Nacionais e Internacionais de Cursos Massivos Online e Abertos (MOOC) e sua Expansão no Brasil. *Revista Tecnologias na Educação*, Edição Temática III, v. 18, no. 9, p. 1-17, 2017. Disponível em [<http://tecedu.pro.br/wp-content/uploads/2017/02/Art14-vol18-edi%C3%A7%C3%A3o-tematica-III-I-SNTDE-2016.pdf>]

CEN, H.; KOEDINGER, K.; JUNKER, B. *Learning Factors Analysis - A general method for cognitive model evaluation and improvement*. In: 8th International Conference on Intelligent Tutoring Systems, 2006, p. 2-13. Disponível em [[https://link.springer.com/chapter/10.1007/11774303\\_17](https://link.springer.com/chapter/10.1007/11774303_17)]

CEN, H.; KOEDINGER, K.; JUNKER, B. Is over practice necessary? – Improving learning efficiency with the cognitive tutor through educational data mining. In: LUCKIN, Rose and KOEDINGER, Ken (Eds.) *Proceedings of the 13th International Conference on Artificial Intelligence in Education*, 2007, p. 511-518. Disponível em [<https://dl.acm.org/doi/10.5555/1563601.1563681>]

CORBI, A.; BURGOS, D. Implementation of the Recommendation Model LIME in Cognitive and Visual Interactive Tutors from PSLC. *IEEE Latin America Transactions*, vol. 13, no. 2, p. 516-522, 2015. Disponível em [<https://ieeexplore.ieee.org/abstract/document/7055573>]

GOWDA, S. M.; BAKER, R. S. J. D.; PARDOS, Z.; HEFFERNAN, N. T. The Sum is Greater Than the Parts: Ensembling Models of Student Knowledge in Educational Software. *SIGKDD Explorations*, vol. 13, no. 2, p. 37-44, 2011. Disponível em [<https://dl.acm.org/doi/10.1145/2207243.2207249>]

KOEDINGER, K. R.; STAMPER, J. C.; LEBER, B. SKOGSHOLM, A. Learn Lab's Data Shop: A Data Repository and Analytics Tool Set for

Cognitive Science. *Topics in Cognitive Science*, vol. 5, no. 3, p. 668–669, 2013. Disponível em [<https://doi.org/10.1111/tops.12035>]

KOEDINGER, K. R.; BAKER, R. S. J. D.; CUNNINGHAM, K.; SKOGSHOLM, A. LEBER, B.; STAMPER, J. A *Data Repository for the EDM Community: The PSLC DataShop*. Handbook of Educational Data Mining, CRC Press, p. 43–56, 2010. Disponível em [<https://datashop.memphis.edu/about/pubs/KBCSL-v6.pdf>]

KOEDINGER, K. R.; STAMPER, J. C. *A Data Driven Approach to the Discovery of Better Cognitive Models*. In: International Conference on Educational Data Mining, 3rd, Pittsburgh, PA, USA, 2010, p. 325–326. Disponível em [[http://pact.cs.cmu.edu/pubs/Koedinger,%20Stamper%20edm2010\\_.pdf](http://pact.cs.cmu.edu/pubs/Koedinger,%20Stamper%20edm2010_.pdf)]

KOEDINGER, K. R.; STAMPER, J. C.; MCLAUGHLIN, E. A.; TRISTAN, N. *Using Data-Driven Discovery of Better Student Models to Improve Student Learning*. In: Conf. Artificial Intelligence in Education, 2013, p. 421–430. Disponível em [<https://www.learnlab.org/research/wiki/images/e/eb/Koedinger-et-al-aied2013.pdf>]

KRÜGER, A.; MERCERON, A.; WOLF, B. *A Data Model to Ease Analysis and Mining of Educational Data*. In: International Conference on Educational Data Mining, (EDM), 3, p. 131–140, Pittsburgh, PA, USA, 2010. Disponível em [[https://www.researchgate.net/publication/221570403\\_A\\_Data\\_Model\\_to\\_Ease\\_Analysis\\_and\\_Mining\\_of\\_Educational\\_Data\\_1](https://www.researchgate.net/publication/221570403_A_Data_Model_to_Ease_Analysis_and_Mining_of_Educational_Data_1)]

LESZCZENSKI, J. M.; BECK J. E. *What's in a word? Extending learning factors analysis to model reading transfer*. In: Educational Data Mining Workshop at the 14th International Conference on Artificial Intelligence in Education, Los Angeles, California, USA, 2007, p. 31–39. Disponível em [[https://www.cs.cmu.edu/~listen/pdfs/AIED2007\\_EDM\\_Leszczenski\\_lfa\\_transfer.pdf](https://www.cs.cmu.edu/~listen/pdfs/AIED2007_EDM_Leszczenski_lfa_transfer.pdf)]

MERCERON, A.; YACEF, K. *Interestingness Measures for Association Rules in Educational Data*. In: First International Conference on Educational Data Mining, 2008, p. 1–10. Disponível em [<https://pdfs.semanticscholar.org/604d/417ba6395fe183667f4496deb0f79579ac0d.pdf>]

NWAIGWE, A.; KOEDINGER, K. R.; VANLEHN, K.; HAUSMANN, R.; WEINSTEIN, A. Exploring alternative methods for error attribution in learning curves analyses in inteligente tutoring systems. In: LUCKIN and KOEDINGER (Eds.) *Proceedings of the 13th International Conference on Artificial Intelligence in Education*, Los Angeles, California, USA, 2007, p. 246-253. Disponível em [<https://dl.acm.org/doi/10.5555/1563601.1563643>]

PERRY, G. T.; PADOVANI, H.; SOUZA, N. S.; ROSSATTO, P. What affects permanence in a MOOC about Chemistry? *International Journal for Innovation education and Research*, v. 7, p. 19-29, 2019. Disponível em [<https://doi.org/10.31686/ijer.vol7.iss3.1315>]

RAFFERTY, A. N.; YUDELSON, M. *Applying learning factors analysis to build stereotypic student models*. In: 13th International Conference on Artificial Intelligence in Education, Los Angeles, California, USA, 2007, p. 253-257. Disponível em [<https://dl.acm.org/doi/10.5555/1563601.1563745>]

RAFTERY, A. Bayesian Model Selection in Social Science Research. *Sociological Methodology*, v. 28, p. 111-163, 1995. Disponível em [<https://www.stat.washington.edu/raftery/Research/PDF/socmeth1995.pdf>]

RAJNI, J.; MALAYA, D. B. *An approach to generate students' response on learning environment using Association Rule Mining*. In: 2014 International Conference on Data Mining and Intelligent Computing (ICDMIC), New Delhi, India, 2014, p. 15-20. Disponível em [<https://ieeexplore.ieee.org/abstract/document/6954225>]

RITTER, S.; KOEDINGER, K. R. An architecture for plug-in tutor agents. *Journal of Artificial Intelligence in Education*, v.7, n. (3-4), p. 315-347, 1998. Disponível em [<https://www.learntechlib.org/p/82400/>]

ROMERO, C.; VENTURA, S. *Educational Data Mining: A Survey from 1995 to 2005*. Expert Systems with Applications, p. 125-146, 2007. Disponível em [<https://www.sciencedirect.com/science/article/pii/S0957417406001266>]

SOUZA, N. S.; SANTIAGO, G.; LIMA, J. V.; CANTO FILHO, A. B.; PERRY, G. T. Análise das Trajetórias de Aprendizagem de Estudantes Inscritos em um MOOC. *Renote - Revista Novas Tecnologias na Educação*, v. 16, n. 1, p. 1-11, 2018. Disponível em [<https://doi.org/10.22456/1679-1916.86034>]

SOUZA, N. S.; WIVES, L. K.; PERRY, G. T. Tendências de Pesquisas que Utilizam Learning Analytics em MOOCs: um mapeamento sistemático. *Renote - Revista Novas Tecnologias Na Educação*, v. 17, n. 01, p. 82-92, 2019. Disponível em [<https://doi.org/10.22456/1679-1916.95710>]

SOUZA, V. F.; PERRY, G. Identifying student behavior in MOOCs using Machine Learning. *Jedm - International Journal for Innovation Education and Research*, v. 7, p. 30-39, 2019. Disponível em [<https://doi.org/10.31686/ijer.vol7.iss3.1318>]

SWEENEY, M.; RANGWALA, H.; LESTER, J. JOHRI, A. *Next-Term Student Performance Prediction: A Recommender Systems Approach*. *Journal of Educational Data Mining*. Memphis, Tennessee, p. 22-51, 2016. Disponível em [<https://arxiv.org/abs/1604.01840>]

#### COMO CITAR ESTE ARTIGO

**ABNT:** SOUZA, Vanessa Faria de; PERRY, Gabriela Trindade. *Proposta de melhoria dos dados de relatórios de uma plataforma de MOOCS brasileira*. *Revista Brasileira de Aprendizagem Aberta e a Distância*, V. 19, n. 1. 2020. doi:<http://dx.doi.org/10.17143/rbaad.v19i1.424>.

#### Autor Correspondente

Vanessa Faria de Souza  
E-mail: [vanessa.souza@ibiruba.ifrs.edu.br](mailto:vanessa.souza@ibiruba.ifrs.edu.br)

Recebido: 16/04/2020

Aceito: 24/07/2020

Publicado: 04/08/20