



Indicadores para avaliação de estudos envolvendo tecnologias de mineração de dados educacionais de apoio à avaliação do estudante na Educação Superior a Distância

Douglas de Jesus Costa¹ * (PQ), José Leonardo Oliveira Lima¹ ** (PQ)

¹Universidade Estadual de Goiás - Câmpus Anápolis de Ciências Exatas e Tecnológicas Henrique Santillo, Anápolis – Goiás

*douglas.jesuscosta@gmail.com, **jjleo@ueg.br

Resumo: A Educação a Distância na Educação Superior tem tido uma crescente importância no cenário brasileiro e mundial. Esse aspecto demanda que sejam discutidas possíveis melhorias que podem ser implementadas em face das necessidades de docentes e alunos no processo de ensino-aprendizagem para uma EaD calcada em princípios educacionais, científicos e tecnológicos mais sólidos. Considerando que uma das principais dificuldades levantadas no estudo de Lima (2016) envolve o grande volume de dados e a falta de informações necessárias para o complexo processo de avaliação realizado pelo docente *online*, essa pesquisa buscou investigar tecnologias do campo de estudos de Mineração de Dados Educacional (MDE) com o objetivo de identificar indicadores para composição de quadro de avaliação das tecnologias, métodos, técnicas e abordagens pedagógicas contemplados em estudos de MDE que podem nortear a implementação de futuras ferramentas de apoio ao docente na avaliação do estudante na Educação Superior a Distância. A abordagem da pesquisa foi qualitativa, apoiando-se nos processos de revisão sistemática e das técnicas de análise qualitativa de conteúdo para a identificação e composição dos indicadores.

Palavras-chave: Mineração de dados educacionais. Processo avaliativo. Educação Superior a Distância. Indicadores.

Introdução

A Educação a Distância (EaD) é uma temática em crescente discussão no cenário atual. Parte desse crescimento e popularidade podem ser relacionados ao potencial na democratização e flexibilização do ensino promovidos pela EaD.

A EaD, atualmente, contempla o processo de ensino-aprendizagem mediado por tecnologias, o que permite que aluno e professor não necessitem estar presentes em um mesmo espaço físico e em um mesmo horário para a realização



da atividade de ensino-aprendizagem (MOORE; KEARSLEY, 2008). Ressalta-se, também, que existem outras especificidades ligadas à EaD, como: os diferentes canais de comunicação, abordagens de gerenciamento específicas e diferentes abordagens e necessidades pedagógicas, o que remete à um processo que envolve grande complexidade e à necessidade de atuação e colaboração interdisciplinar (PRETI, 2009).

Face à expansão da EaD no cenário nacional, discussões em torno da qualidade na Educação Superior a Distância tornam-se cada vez mais importantes para que, com o argumento de satisfazer à necessidade de ampliação e democratização do acesso, não ocorra uma massificação precarizada, decorrentes de expansão demagógica para fins politiqueros ou de marketing de alguns segmentos ou para atender princípios de redução de custos indiscriminada, com vistas à mercantilização da educação superior que emerge da excessiva oferta de cursos superiores “baratos” que, muitas vezes, não estão alinhados à preconizada e necessária qualidade.

Segundo Santos e Nogueira (2017), é primordial para o desenvolvimento de uma EaD de qualidade a modificação de políticas públicas de regulamentação do controle da qualidade dos cursos superiores a distância no Brasil.

Dentre os questionamentos para uma Educação Superior a Distância mais efetiva, diferente do modelo massificado e pouco reflexivo que têm sido realizado em algumas instituições no Brasil, Lima (2016) discute o limitado entendimento das necessidades de docentes *online* no complexo processo de avaliação do estudante na EaD e, por intermédio de um estudo com especialistas de EaD do Brasil e da Europa, busca compreender algumas das necessidades dos docentes com relação a EaD. Lima (2016) propõe, dentre outras questões, a tecnologia como um instrumento de apoio ao docente *online* na tomada de decisão nos processos avaliativos, face ao grande volume de tarefas, dados que envolve os sistemas e as interações *online* e a sobrecarga de trabalho do docente *online*, quando envolvido em um projeto de EaD verdadeiramente interativo e de qualidade.

O uso das tecnologias em várias áreas contribui para a produção de um enorme volume de dados, que inviabiliza a análise desses pelos seres humanos.



Esses dados acabam sendo ignorados ou não aproveitados para benefício das organizações. Em face desse cenário, torna-se necessária a criação de técnicas e ferramentas para a análise e obtenção de informações² a partir desses dados (CASTRO, 2016).

O processo de mineração de dados envolve um conjunto de atividades relacionadas à extração e preparação de uma base de dados para análise seguida da aplicação de algoritmos nessa base de dados para que, finalmente, seja possível a extração de informação. (CASTRO, 2016).

Os ambientes de aprendizagem *online* registram as diversas ações e interações do estudante no momento de acesso ao ambiente *online*, o que implica na geração de uma grande quantidade de dados. Esse volume de dados favoreceu o surgimento do campo de estudos de Mineração de Dados Educacionais (MDE) ou *Educational Data Mining (EDM)* que, segundo Romero e Ventura (2013, p. 12), pode ser definida como “a aplicação de técnicas de mineração de dados para o específico tipo de dados provenientes do ambiente educacional [...]” (Tradução nossa).

Considerando a importância do campo e as necessidades levantadas no estudo de Lima (2016), principalmente com relação à falta de informação ao docente *online* em cursos de Educação a Distância, verificou-se o grande potencial de utilização de processos de mineração de dados de modo a produzir informações de auxílio ao professor no complexo processo de avaliação, a partir da aplicação de técnicas de mineração nos dados decorrentes do acesso e das diversas interações e atividades realizadas pelos alunos no Ambiente de Virtual de Aprendizagem (AVA).

O presente estudo tem por objetivo identificar e elaborar indicadores de avaliação de pesquisas envolvendo a compreensão das abordagens pedagógicas, processos, técnicas e tecnologias de mineração, redução e apresentação de dados existentes. A busca foi delimitada aos estudos que apresentam tecnologias com potencial para possível integração ao AVA Moodle e com foco na avaliação de estudantes de graduação a distância, para satisfazer as necessidades de informação dos docentes no processo.

² Para esse estudo, considera-se a definição de informação concebida no contexto da computação, em que uma informação pode ser compreendida como um dado ou conjunto de dados processados/analísados.

Material e Métodos

Para a obtenção dos indicadores, foram utilizados os princípios da metodologia de análise qualitativa de dados proposta por Lima e Manini (2016) para levantamento dos indicadores, juntamente com concepções inspiradas em processo de revisão sistemática (MACHI; MCEVOY, 2012; MOHER *et al.*, 2009).

Objetivou-se, através de uma revisão criteriosa da literatura relacionada às tecnologias de mineração de dados propostas para a educação superior a distância, chegar-se a uma concepção de grande abrangência sobre o cenário atual de pesquisa dessas ferramentas e, com base na análise dessas e a partir do olhar das necessidades de docentes *online* identificadas por Lima (2016), elaborar um quadro de indicadores para avaliação dos estudos. O estudo foi concebido em três fases.

A primeira fase compreendeu o estudo exploratório inicial e a formulação de *strings*³ de buscas. Foi feita a pesquisa exploratória de estudos nas bases de dados Google Scholar e Web of Science. Com isso, foram identificados termos que resultaram na elaboração e testes de 28 *strings* de buscas. Ao final, a partir de uma mesma concepção de *string* e operadores booleanos, foram elaboradas duas *strings* diferentes, em virtude das especificidades da sintaxe de busca de cada base utilizadas: Scopus e Web of Science. As duas *strings* utilizadas são apresentadas no Quadro 1.

Quadro 1 - Strings de busca consideradas

Base de dados	String elaborada	Total de estudos
Scopus	TITLE-ABS-KEY (("high education" OR "higher education" OR "college") AND ("learning management system" OR lms OR "virtual learning environment" OR vle OR "e-learning" OR "distance learning" OR "distance education") AND ("data mining" OR "learning analytics" OR analytics OR "educational data mining" OR edm) AND (tools OR technic OR software OR mechanism OR method OR technology OR dashboard))	168
Web of Science	TS=(("high education" OR "higher education" OR "college") AND ("learning management system" OR lms OR "virtual learning environment" OR vle OR "e-learning" OR "distance learning" OR "distance education") AND ("data mining" OR "learning analytics" OR analytics OR "educational data mining" OR edm) AND (tools OR technic OR software OR mechanism OR method OR technology OR dashboard))	70

Fonte: Elaborado pelos autores.

³ Nesse contexto, uma *string* de buscas compreende um conjunto de palavras utilizadas para a localização de estudos indexados através dessas palavras em bases de dados científicos.



Com a seleção dos estudos recuperados através das *strings* elaboradas, foi feita a segunda fase do estudo que teve como principal objetivo a realização de uma primeira filtragem dos estudos considerados e – mediante ao contato com os estudos – proporcionar uma melhor percepção sobre as pesquisas realizadas e, consequentemente, aos estudos de tecnologias de mineração para a educação. Para a realização da análise e filtragem, foi feita a leitura do título e resumo do artigo – se estendendo a leitura da introdução dos artigos caso os pesquisadores não conseguissem obter um entendimento considerável do estudo a partir do título e resumo – seguido da análise considerando os seguintes critérios: “O estudo utiliza ou introduz uma tecnologia de mineração de dados?” e “O estudo propõe/analisa a ferramenta objetivando contribuir ao processo de avaliação do aluno pelo docente *online*?”. Esses critérios foram utilizados junto a outro aspecto de percepção do pesquisador, que atribuiu uma nota de 0 a 10 (sendo que a nota 10 se refere ao maior grau em que o estudo analisado se adere ou responde ao objetivo proposto da presente investigação). Nessa fase, os pesquisadores começaram a fazer o *brainstorming* e o início da proposição dos indicadores.

A terceira fase, em fase de finalização, compreende a análise dos estudos que obtiveram uma nota superior a 6 na fase anterior. Essa fase objetiva a consolidação de uma percepção dos pesquisadores de maior profundidade das pesquisas e também revisar os indicadores inicialmente identificados na fase anterior, complementar e propor novos indicadores mediante a leitura completa dos estudos selecionados.

Portanto, a terceira fase compreende a consolidação dos indicadores para a avaliação dos estudos envolvendo as tecnologias de mineração de dados bem como o aprofundamento da visão dos pesquisadores com relação ao experimento e consequentemente um olhar mais aprofundado sobre possíveis indicadores avaliativos.

Para a terceira fase, da base de dados Scopus, foram selecionados 15 estudos e da base Web of Science, 4 estudos. Para isso, observou-se também como critério a disponibilidade desses estudos para *download* no portal de periódicos Capes.

Resultado parcial e Discussão

A presente pesquisa se encontra em realização e os resultados obtidos e discussão são ainda preliminares. Até o momento da escrita dessa comunicação, foram concluídos os estudos do material da Base de Dados Scopus, sendo que, preliminarmente, destacamos os resultados a seguir, de caráter ainda provisório e com possibilidades de acréscimo, supressão ou modificação.

A partir da análise, verificou-se que uma das principais utilizações da mineração de dados para proposição de ferramentas educacionais se dá na elaboração de ferramentas que objetivam a prevenção de desistência ou reprovação do aluno.

Os indicadores propostos são agrupados em diferentes aspectos da ferramenta, como a sua descrição, seu objetivo, a metodologia adotada e a apresentação de informações (Interface). Apresenta-se, no Quadro 2, os indicadores obtidos a partir da análise dos estudos da base de dados Scopus. Junto a cada indicador, são também citados os estudos que mais colaboraram para a identificação do indicador.

Quadro 2 - Indicadores propostos

Indicadores propostos até o momento da pesquisa	
Indicador	Descrição
Descrição da ferramenta	
Modelo ou conceito com descrição detalhada	Esse indicador baseia-se na existência de vários estudos que propõem técnicas interessantes, entretanto utilizam ferramentas pagas ou cujos códigos fonte não são disponibilizados ou não acessíveis a todos. Dessa forma, esse indicador preconiza o nível de detalhamento da ferramenta/metodologia adotada. O estudo, ao apresentar uma a descrição ou documentação detalhada, permite a reprodução da técnica em outros cenários com outras ferramentas, quando possível, além de proporcionar um bom entendimento sobre quais aspectos são abordados pela ferramenta, como se chegou a sua concepção e se tal ferramenta pode ser confiável levando-se em conta o contexto do curso, a instituição de ensino e as concepções do próprio professor sobre o processo de avaliação. (PARK; JO, 2015).
Objetivo da ferramenta	
Concepção de auxílio avaliação do professor	Com o estudo de Lima (2016), verifica-se que uma das necessidades levantadas por docentes <i>online</i> relaciona-se à complexidade do processo o avaliativo no contexto da Educação a Distância. Essa necessidade advém do limitado contato visual e físico com discentes (a exemplo do que

	ocorre no ensino presencial que permite ao docente captar reações de seus alunos face a um conteúdo) e, ao mesmo tempo, a grande quantidade de dados presentes no ambiente de aprendizagem <i>online</i>. Portanto, a partir dessa necessidade e mediante o cenário em que o docente <i>online</i> acaba por ser pouco considerado na proposição de ferramentas de mineração de dados, o indicador preconiza que a ferramenta proposta busque trazer como objetivo o auxílio ao professor no complexo processo de avaliação, verificando quais informações tornam-se mais úteis ao docente <i>online</i>. (LARA et al, 2015).
Metodologia do estudo	
Correlação significativa/variáveis embasadas em estudos anteriores	Muitas das ferramentas buscam, através da análise da grande quantidade de dados, trazer informações de forma sintetizada. Para se chegar a essas informações, são consideradas variáveis que norteiam quais dados serão considerados sobre o aluno, como por exemplo: número de visitas ao sistema, tempo de leitura, materiais baixados, entre outras. Com isso, torna-se necessário analisar nas ferramentas a correlação dessas variáveis ou dados extraídos com a informação objetivada, seja essa a possibilidade de reprovação do aluno, a indicação de que o aluno está ou não aprendendo, entre outras. Com isso, as variáveis consideradas devem passar por rigorosos testes de correlação para que seja possível atestar a correlação ou importância dessas informações levando-se em conta a informação que se deseja proporcionar. (PARK; JO, 2015).
Testes da ferramenta em diferentes contextos	A metodologia deve incluir, também, no contexto de novas tecnologias, mecanismos que comprovem a eficácia da ferramenta realizando testes em diferentes grupos, diferentes cursos, diferentes períodos de tempo, entre outros contextos. Nota-se a importância desses testes da ferramenta – assim como o das variáveis – em diferentes contextos em razão da grande quantidade de fatores que podem colaborar para o maior ou menor desempenho da ferramenta e dos dados elencados em um determinado contexto, que pode influenciar, como por exemplo, por meio do <i>design</i> do curso, de especificidades do material de estudo, do grupo de alunos da turma, da metodologia adotada pelo docente <i>online</i>, por especificidades da instituição de ensino, entre outros. (LARA et al, 2014; LIU et al, 2015)
Teorias educacionais e estudos empíricos.	Considerando a atuação da ferramenta no contexto educacional e a grande importância de teorias educacionais para o processo de ensino-aprendizagem como, por exemplo, a teoria sociointeracionista de Vygotsky e a teoria cognitivista de Jean Piaget, além de que, no estudo de Lima (2016), os especialistas explicitaram a necessidade de ferramentas embasadas em teorias educacionais, verifica-se como possível indicador o embasamento dessas tecnologias nas concepções teóricas educacionais.
Consideração de variáveis relacionadas ao curso	As ferramentas propostas também levam em consideração os materiais ou atividades feitas pelo professor na análise dos dados do ambiente. Isso se dá em razão de que – dependendo do curso, disciplina, atividades feitas, materiais, entre outros – o acesso do estudante ao ambiente pode ser influenciado. Torna-se necessário que as ferramentas propostas levem em consideração esses elementos que não dependem diretamente do estudante, mas sim do próprio <i>design</i> do curso. Com isso, as ferramentas devem ou considerar esses elementos ou buscar eliminar ou minimizar sua influência trazendo experimentos que em um determinado grupo de amostra, busque elencar diferentes grupos e cenários de estudantes. (MWALUMBWE; MTEBE, 2017; LIU et al, 2015; LARA et al, 2014).
Viés dos dados analisados para o teste da ferramenta (Amostragem)	A verificação realizada da ferramenta, com relação à seleção da base de dados de teste, é feita com bases de dados aleatórias ou turmas completas de um curso, averiguando a eficiência da ferramenta em determinadas unidades e a possível falha da ferramenta ou menor eficiência em unidades que, por diversos motivos, não sejam propícios para a análise



	feita pela ferramenta levando em conta as variáveis que são focadas por essa para determinar o sucesso ou a falha de estudantes. Recomenda-se, portanto, que não sejam analisados apenas dados que sejam propícios, mas também dados de módulos que não sejam propícios para que, assim, sejam delimitadas quais situações são propícias para o uso da ferramenta sem que seja atribuída eficiência apenas em base de dados que sejam propícias para a tecnologia apresentada. (Viés na base de dados). (LIU et al, 2015; AZEVEDO et al, 2017).
Pesquisas qualitativas e Identificação das necessidades dos usuários	Apesar desse estudo trazer uma análise das tecnologias sob o olhar de necessidades docentes identificadas no estudo de Lima (2016), verifica-se que cada curso e instituição possui suas próprias especificidades e necessidades. Torna-se necessário, portanto – para a implementação de uma ferramenta de grande valia e alinhada as necessidades locais – que métodos qualitativos de validação da ferramenta e identificação das necessidades dos docentes <i>online</i> e discentes de um determinado curso sejam utilizados. Dessa forma, para validação da utilidade de uma ferramenta, bem como sua concepção, além dos testes quantitativos, identifica-se a necessidade de pesquisas qualitativas para que percepções, necessidades e outras riquezas que geralmente não são agregadas em um experimento quantitativo possam ser agregadas ao desenvolvimento da ferramenta (MWALUMBWE; MTEBE, 2017; PARK; JO, 2015; LARA et al, 2014).
Apresentação das informações (Interface)	
Quantidade de informações disponibilizadas	Como os especialistas apontaram no estudo de Lima (2016), existe um grande volume de atividade e informações atribuídas ao docente <i>online</i> , fazendo com que esse não tenha disposição de tempo necessário para a realização de análises de dados complexos ou de uma grande quantidade de dados. Em razão disso, recomenda-se, em face da grande quantidade de informações disponíveis no ambiente virtual e volume de tarefas do professor, trazer uma quantidade de informações mínima, e simplificada ao docente, considerando também como norteador elementos presentes no campo de estudo de experiência do usuário (UX) e interação humano-computador. (PARK; JO, 2015).

Fonte: Elaborado pelos autores.

Considerações Finais

Considerando as ferramentas e pesquisas analisadas até o momento, a delimitação adotada por essa pesquisa com relação a tecnologias de mineração de dados propostas para a Educação Superior a Distância e o olhar trazendo as necessidades apontadas por especialistas do Brasil e da Europa no estudo de Lima (2016), foi possível identificar um conjunto de indicadores que podem ser levados em conta na avaliação de estudos para apoiar elaboração de novas tecnologias ou adequação de tecnologias existentes, objetivo do projeto ao qual este estudo está vinculado (vide detalhe no item Agradecimentos).

Foi possível identificar indicadores que correspondem a diferentes aspectos da ferramenta ou pesquisa, como sua descrição, sua metodologia, seu objetivo e o





modo com que as informações são apresentadas. Verifica-se que um dos principais aspectos que apresenta grande importância na elaboração das tecnologias, que nem sempre é considerado, envolve a influência do contexto na concepção da ferramenta e em sua eficácia.

Outro aspecto passível de observação está relacionado à quantidade de estudos que objetivam a produção de um indicador de maior abstração, trazendo parte do trabalho que antes seria realizado pelo docente através do complexo processo de avaliação para uma ferramenta, como, por exemplo, a constatação da possível desistência ou falha de um aluno. Face às necessidades apresentadas por Lima (2016), verifica-se que, na visão de docentes, existe uma demanda por ferramentas que ofereçam informações sintetizadas, favorecendo que as análises de maior complexidade e abstração decorrentes sejam realizadas pelo docente. Tal aspecto explicita uma discussão necessária que deve ser abordada em futuros estudos, contrastando a visão conflitante entre os docentes *online* e as concepções dos pesquisadores que propõem ferramentas de EaD.

De forma geral, observa-se que a análise de conteúdo trazendo elementos da revisão sistemática contribuiu para a seleção de estudos e para uma visão de maior completude do cenário de pesquisa.

Agradecimentos

O presente estudo é resultado parcial e integra o projeto de pesquisa aprovado na PrP-UEG sob o número 2651, intitulado “Tecnologias e recursos de mineração de dados para suporte ao docente na avaliação de estudante de graduação a distância no Ambiente Virtual de Aprendizagem (AVA) Moodle”, do Prof. Dr. José Leonardo Oliveira Lima. Os pesquisadores agradecem à Universidade Estadual de Goiás pela possibilidade de realização do projeto e oportunidade de divulgação do Estudo no CEPE.

Referências

AZEVEDO, J. M. et al. Learning Analytics: A Way to Monitoring and Improving Students' Learning. In: International Conference on Computer Supported Education, 9., 2017, Porto. **Proceedings...**Porto: Scitepress, p. 641-648.

CASTRO, L. N. D., FERRARI, D. G. **Introdução à Mineração de Dados: Conceitos Básicos, Algoritmos e Aplicações**. 1. ed. São Paulo: Saraiva, 2016.



LARA, J. A. et al. A system for knowledge discovery in e-learning environments within the European Higher Education Area – Application to student data from Open University of Madrid, UDIMA. **Computers & Education**, v. 72, p. 23–36, 2014.

LIMA, J. L. O. **Avaliação discente em cursos de graduação a distância mediados por ambientes virtuais de aprendizagem**: Necessidade de informações dos docentes na visão de especialistas europeus e brasileiros. 2016. Tese (Doutorado em Ciência da Informação) – Universidade de Brasília, Programa de Pós-Graduação em Ciência da Informação, Brasília, 2016.

LIMA, J. L. O.; MANINI, M. P. Metodologia para análise de conteúdo qualitativa integrada à técnica de mapas mentais com o uso dos softwares nvivo e freemind. **Inf. Inf.**, Londrina, v. 21, n. 3, p. 63 – 100, 2016.

LIU, Danny et al. Validating the Effectiveness of the Moodle Engagement Analytics Plugin to Predict Student Academic Performance. **Twenty-first Americas Conference on Information Systems**, Puerto Rico, 2015.

MACHI, L. A.; MCEVOY, B. T. **The literature review**: Six steps to success. Thousand Oaks, California: Corwin Press, 2012.

MOHER, D. *et al.* Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. **PLoS med**, v. 6, n. 7, p. e1000097, 2009.

MOORE, M. G.; KEARSLEY, G. **Educação a distância**: uma visão integrada. São Paulo: Thomson Learning, 2007.

MWALUMBWE, Imani; MTEBE Joel S. Using learning analytics to predict students' performance in Moodle learning management system: A case of Mbeya University of Science and Technology. **The Electronic Journal of Information Systems in Developing Countries**, v. 79, n. 1, p. 1-13, 2017.

PARK, Yeonjeong; JO, Il-Hyun. Development of the Learning Analytics Dashboard to Support Students' Learning Performance. **Journal of Universal Computer Science**, v. 21, n. 1, p. 110-133. 2015

PRETI, O. **Educação a distância**: Fundamentos e Políticas. Cuiabá: EdUFMT, 2009.

ROMERO, Cristobal; VENTURA, Sebastian. Data mining in education. **Wires Data Mining Knowl Discov**, v. 3, p. 12 – 27, 2013.

SANTOS, C. A.; NOGUEIRA, D. X. P. A expansão da educação superior a distância e os novos marcos regulatório: definições e desafios na perspectiva da qualidade. In: LIMA, D. C. B. P.; SANTOS, C. A.; TOSCHI, M. S. **Educação a distância (EaD)**: Realidades, Evolução e Contextos. Anápolis: Editora UEG, 2017.