

Mineração de Dados Educacionais Aplicada à Análise Preditiva em Fóruns no Moodle

Renê de Ávila Mendes¹, Leandro Augusto da Silva¹

¹Universidade Presbiteriana Mackenzie – São Paulo – SP – Brasil

renemendes@yahoo.com.br, prof.leandro.augusto@mackenzie.br

Abstract. *Educational Data Mining can contribute greatly to improving the teaching-learning process when the student involvement in the construction of knowledge is measured. This article discusses the learning model used by Moodle Learning Management System, identifies its activities that contribute to the collective construction of knowledge, maps log attributes related to forum, make exploratory analysis of these data and apply linear regression technique to them. Data of blended learning courses from an Educational unity of a private Higher Education Institution. The results do indicate dependence between the undergraduate period and the involvement in Moodle forums.*

Resumo. *A Mineração de Dados Educacionais (MDE) pode trazer importante contribuição para a melhoria do processo ensino-aprendizagem quando o envolvimento dos alunos na construção do conhecimento é medido. Este artigo discute o modelo de aprendizagem adotado no Ambiente Virtual de Aprendizagem Moodle, identifica as atividades desta ferramenta que mais contribuem para a construção coletiva do conhecimento, mapeia atributos de log referentes a participação em fóruns, faz análise exploratória destes dados e aplica a eles a técnica de regressão linear. São utilizados dados de cursos mistos de uma Unidade educacional de uma Instituição de Ensino Superior de caráter particular. Os resultados indicam dependência entre o semestre letivo e o envolvimento em fóruns no Moodle.*

1. Introdução

A presença dos computadores pessoais e a Internet continua afetando o modo de vida das pessoas, incluindo a maneira como elas aprendem. Termos como Ambiente Virtual de Aprendizagem (AVA), Aprendizado Misto e *Massive Open Online Courses* (MOOC) definem formas de uso de recursos computacionais para o apoio do processo ensino-aprendizado [Dillenbourg et al. 2002, Piccoli et al. 2001, Littlejohn et al. 2016, Park et al. 2016].

Moodle, *Modular Object-Oriented Dynamic Learning Environment*, é um exemplar dos softwares classificados como *Learning Management Systems* (LMS), do grupo de Ambientes Virtuais de Aprendizagem (AVA), e seus objetivos incluem adicionar tecnologia web ao aprendizado, suplementar os tradicionais cursos presenciais [Romero et al. 2008], complementar a experiência de aprendizado [Sin and Muthu 2015], entregar um ambiente de aprendizado online e até mesmo suportar uma universidade virtual [Coates et al. 2005]. A ampla adoção de AVAs por Instituições de Ensino é uma realidade [Coates et al. 2005] e, em particular, o Moodle está presente em 232 países,

contabiliza cerca de 70 mil sites registrados e atende mais de 91 milhões de alunos [Moodle 2016b].

O presente artigo propôs-se a coletar logs de utilização do AVA Moodle, tratá-los e submetê-los a técnica de Aprendizagem de Máquina visando a predição de utilização do recurso Fórum por alunos de oito semestres letivos de uma unidade educacional de uma Instituição de Ensino Superior de caráter privado. Foram analisados dados gerados nos semestres 2014/2, 2015/1, 2015/2 e 2016/1. Os resultados obtidos podem ser base de um estudo comparativo da participação de alunos de diferentes unidades educacionais em fóruns de discussões e oportunizarem estudos sobre a participação das discussões em fóruns nos resultados finais em cursos mistos.

2. Trabalhos Correlacionados

Logs do Moodle são analisados visando entender o comportamento de estudantes [Młynarska et al. 2016, Casey and Gibson 2010, You 2016], desenvolver padrões de persistência e abandono em cursos on-line [Hershkovitz and Nachmias 2011, Lara et al. 2014], medir a percepção de comunidade por parte de alunos on-line [Black et al. 2008] e relacionar o uso de recursos do Moodle ao resultado final do aluno [Romero et al. 2009], entre outros objetivos.

[Zacharis 2015] extrai e analisa dados relativos a atividades em um curso misto, identificando catorze variáveis com significativa associação com a nota final do aluno. Submetendo estas variáveis a uma análise de regressão multivariada são selecionadas as atividades que melhor predizem a nota final do aluno: acesso de recursos (*view*), envolvimento em questionários (*quiz*), leitura e postagem de mensagens (*forum*, *email* e *chat*) e criação de conteúdo (*wiki* e *blog*). [Nandi et al. 2011] identificam a interação aluno-aluno e aluno-tutor como o recurso mais importante em um curso online e encontram correlação entre a participação de alunos em fóruns e seu resultado final em cursos on-line. [Dougiamas and Taylor 2003] observam que uma maior utilização de fóruns no Moodle pode ser um indicativo de crescimento na interação cooperativa entre alunos e entre alunos e professores na construção coletiva do conhecimento em cursos à distância. Em cursos mistos é possível afirmar que a interação aluno-professor contribui para a manutenção da motivação e o sucesso do aluno [Jaggars and Xu 2011]. Por outro lado [Iglesias-Pradas et al. 2015] não identificaram influência das interações de alunos no Moodle sobre as competências inter-curriculares ‘comprometimento’ e ‘trabalho cooperativo’.

3. Aprendizado Cooperativo no Moodle

Um AVA é definido como mais que um repositório de recursos educacionais e sua efetividade é claramente condicionada ao modelo de aprendizado adotado pelo AVA [Piccoli et al. 2001] e pelo modelo de aprendizado utilizado no desenvolvimento dos recursos educacionais oferecidos [Ally 2004]. No Moodle o modelo de aprendizagem transparece quando seus idealizadores explicam que a motivação de sua criação é oferecer um sistema web que suporte estruturas web e interfaces que apoiem as epistemologias sócio construcionistas, que motivem a leitura aberta, a reflexão crítica e uma escrita construtiva [Dougiamas and Taylor 2003].

Dentre os modelos de aprendizagem que permeiam o uso de recursos computacionais no contexto ensino-aprendizagem e no desenvolvimento de AVAs podem ser citados o objetivismo, o construtivismo e o sócio-construcionismo. O primeiro crê que existe uma única realidade que pode ser descoberta pelos métodos objetivos da ciência. O estudo das estruturas e de seus relacionamentos leva a uma representação abstrata destas estruturas e relacionamentos, a qual é transmitida para os estudantes que, manipulando os símbolos abstratos, aprenderão à medida que a representação abstrata refletir a realidade [Vrasidas 2000]. O modelo de aprendizado objetivista entende que o conhecimento é transmitido, e está presente em recursos computacionais categorizados como *Computer Aided Instruction (CAI)* [Piccoli et al. 2001]. O construtivismo entende haver múltiplas perspectivas a respeito de uma realidade e que esta, embora imponha limites reais, pode ser interpretada de formas diversas [Vrasidas 2000] sendo o conhecimento não aquele que é transmitido mas o que é construído pela interpretação individual combinada aos estímulos sensoriais recebidos [Anderson 2008]. O modelo construtivista está presente em recursos de aprendizado categorizados como *Virtual Learning Environments* [Piccoli et al. 2001]. O sócio construcionismo também entende que há uma construção da realidade mas diferencia-se pela ênfase na construção social ao invés de uma construção individual atribuída ao construtivismo [Andrews 2012].

Embora os AVAs possam ser associados aos modelos de aprendizado construtivista e sócio construcionista a adoção de um ambiente de aprendizagem não condiciona a Instituição de Ensino ao uso de um ou de outro modelo [Piccoli et al. 2001, Ally 2004, Romero et al. 2008].

4. Mineração de Dados Educacionais

A utilização de AVA baseados em web pelas Universidades remonta da década de 1990 [Watson and Watson 2007], o desenvolvimento do Moodle e sua utilização pelas Instituições de Ensino data de 2002 [Moodle 2016a] e a adoção embrionária do Moodle pela Instituição de Ensino Superior analisada no presente artigo ocorreu em 2004, inicialmente como apoio em atividades de aprendizagem em Laboratórios, sendo completamente adotado pela Instituição de Ensino em 2008 (informação verbal)¹.

A utilização de um Ambiente Virtual de Aprendizagem gera uma grande quantidade de dados que podem ser analisados por aqueles interessados na melhoria do processo de aprendizado ou no aprimoramento de técnicas e algoritmos de mineração de dados para se obter conhecimento do fenômeno do aprendizado [Romero et al. 2008]. As áreas do conhecimento que se dedicam a este estudo são nomeadas de Mineração de Dados Educacionais (MDE) e *Learning Analytics (LA)*, a primeira definida como sendo a área que desenvolve e usa métodos de exploração de dados únicos e de grande escala obtidos do contexto educacional visando o entendimento dos estudantes e dos ambientes em que o aprendizado ocorre [IEDMS 2011]. A segunda área é definida como a área que mede, coleta, analisa e reporta dados de estudantes em seus contextos com propósitos de entendimento e otimização do aprendizado e dos ambientes em que ele ocorre [Siemens 2010]. Embora haja uma clara interseção entre estas duas áreas no que diz respeito à análise do contexto educacional visando seu entendimento e desenvolvimento, LA e MDE apresentam significativas diferenças [Siemens and d Baker 2012].

¹Informação obtida do Gerente de Tecnologia da Informação da Instituição de Ensino, presente no período de implantação do Moodle.

Dentre as subáreas da pesquisa de MDE encontra-se a de predição, subdividida em classificação, regressão e estimação de densidade, todas diretamente ligadas à mineração de dados. A predição, área cujo objetivo é prever aspectos dos dados a partir da análise de outros aspectos conhecidos [Baker et al. 2011], é o método adotado para análise dos dados do presente estudo.

Regressão pode ser definida como uma técnica de análise estatística de investigação e modelagem de relacionamento entre variáveis, aplicável à descrição de dados, estimação de parâmetros, predição e estimação e controle de processos. No modelo de regressão linear a variável independente é chamada de preditora ou regressora e a variável dependente é chamada de resposta. Na presença de apenas uma variável regressora o modelo é chamado de modelo de regressão linear simples e quando há mais de uma variável regressora o modelo passa a ser chamado de modelo de regressão múltipla. O modelo de regressão não indica a existência de relação causa-efeito entre as variáveis regressora e de resposta, relação esta que requer análise que vá além dos dados utilizados para ser identificada. A dependência entre as variáveis regressora e de resposta pode ser verificada por medidas de associação, tais como a correlação de Pearson, que mede a dependência linear entre as variáveis [Montgomery et al. 2001].

No Moodle, uma Atividade pode ser definida como a interação do aluno com outros alunos ou com o professor, sendo catorze os tipos de atividades disponíveis na versão 3.0 deste AVA. Atividades que permitem aos alunos um maior grau de interatividade são *chat*, *forum*, *glossary*, *wiki* e *workshop* [Romero et al. 2008].

A decisão por uma análise quantitativa do conteúdo do fórum é consubstanciada pelo fato de que uma análise qualitativa poderia requerer a interpretação das mensagens trocadas entre alunos e com os tutores (*text mining*).

5. Pré-processamento dos Dados

Os dados requeridos para o processo de predição por regressão foram definidos como dados de matrículas e dados de utilização do fórum. Para a obtenção destes dados foram identificadas as seguintes entidades (Moodle 3.0) (Tabela 1):

Tabela 1. Entidades do Moodle, versão 3.0, utilizadas na regressão.

TABELA	DESCRIÇÃO
mdl_course	Cadastro de cursos.
mdl_course_categories	Categoria do curso.
mdl_course_modules	Recurso ou atividade de um curso.
mdl_enrol	Tipo de vínculo de um usuário com um curso.
mdl_user_enrolments	Arrola um usuário a um curso com um tipo de vínculo.
mdl_context	Contextos para efeitos de gerenciamento de permissão.
mdl_role_assignments	Vínculos de usuários a contextos.
mdl_forum	Estrutura básica de um fórum.
mdl_forum_posts	Postagens por fórum.
mdl_logstore_standard_log	Registro de eventos do sistema.

Para a contabilização da quantidade de alunos por Unidade inicialmente foram identificadas as categorias correspondentes às Unidades educacionais da Instituição de

Ensino Superior. Em seguida foram identificadas as tabelas e campos que armazenam os dados de inscrição, seguindo o modelo apresentado por [Moniz 2015].

O atributo `mdl_role_assignments.timemodified` foi discretizado para o formato `AnoSemestre (AAAASS)` para permitir a contabilização de alunos por semestre. Como um mesmo aluno pode vincular-se a vários cursos, foram contabilizados valores distintos de `mdl_role_assignments.userid`. Os resultados encontrados na contagem de alunos por Unidade educacional pela tabela `mdl_role_assignments` apresentaram problemas de incompletude, pela ausência de dados em algumas Unidades e períodos, e inconsistência, pela discrepância entre a quantidade de alunos entre períodos diferentes. Uma contagem real de alunos por Unidade requereria a integração com dados de matrícula, dados estes não disponíveis na base de dados Moodle analisada. A identificação da causa real da inconsistência requereu intervenção de um especialista de negócio: há limpeza de dados de matrícula ao final de cada semestre letivo, o que justificaria a ausência parcial de dados na tabela `mdl_role_assignments` para semestres letivos anteriores ao atual.

A alternativa encontrada para a contagem de alunos por Unidade educacional foi a pesquisa pela tabela de logs do Moodle: `mdl_logstore_standard_log`. Esta tabela registra todas as ações de usuários em cada contexto. Para esta contagem foram considerados os usuários que tenham ao menos uma ação em um curso. As tabelas e campos utilizados foram [Moniz 2015]: `mdl_course.category`, `mdl_course_categories.name`, `mdl_context.contextlevel` (contexto de curso 50), `mdl_context.instanceid`, `mdl_logstore_standard_log.contextid`, `mdl_logstore_standard_log.timecreated`, `mdl_logstore_standard_log.userid`. O atributo `mdl_logstore_standard_log.timecreated` foi discretizado no formato `AnoSemestre (AAAASS)`.

Faz-se necessário ponderar que: a) um aluno poderá matricular-se em alguns cursos do Moodle que correspondem a disciplinas oferecidas por outra Unidade educacional diferente daquela à qual o aluno está formalmente vinculado, e neste caso este aluno será contabilizado em mais de uma Unidade educacional, o que introduz erro na contagem de alunos vinculados às Unidades, e b) a tabela `mdl_logstore_standard_log` registra ações de quaisquer usuários, não somente alunos. Para o primeiro caso (a), a alternativa encontrada foi calcular a margem de erro na contagem dos alunos, e para isto foram contados alunos que estão matriculados em cursos de mais de uma Unidade em relação ao total de alunos da Unidade (Tabela 2). Para o segundo caso (b) uma vez a identificação de alunos não é possível, assumiu-se que a quantidade de usuários que não são alunos não é significativa.

Tabela 2. Contagem de alunos incluindo margem de erro.

PERÍODO	ALUNOS	ERRO
201402	1632	1,7%
201501	1848	7,6%
201502	1894	15,6%
201601	1968	12,9%

Para a contabilização de participações em fórum por Unidade educacional foram utilizados as seguintes tabelas e campos [Moniz 2015]: `mdl_course.category`, `mdl_course.shortname`, `mdl_course_categories.name`, `mdl_context.contextlevel`,

mdl_context.instanceid, mdl_forum.course, mdl_forum.timemodified,
mdl_forum.type, mdl_forum_discussions.userid, mdl_forum_discussions.timemodified,
mdl_forum_posts.discussion, mdl_forum_posts.created.

O atributo mdl_forum.timemodified foi discretizado no formato AnoSemestre (AAAASS) e foram desprezados os dados de fóruns criados no período anterior a 201402 para haver correspondência com as quantidades de alunos do mesmo período. Não foram considerados na contagem os fóruns do tipo *News*, já que seu objetivo não é o de discutir o conhecimento do curso.

6. Seleção dos Dados

Para execução da análise preditiva foram selecionados os dados de uma Unidade educacional. A escolha baseou-se na hipótese de que alunos e professores desta Unidade fariam melhor uso de recursos tecnológicos disponibilizados pelo Moodle, tal como o fórum. Esta hipótese leva em consideração a afinidade das matrizes curriculares dos cursos oferecidos por esta Unidade com os conceitos tecnológicos subjacentes ao Moodle. O conjunto de dados analisado foi (Tabela 3):

Tabela 3. Quantidade de usuários, fóruns, discussões e postagens da Unidade por AnoSemestre.

ANO_SEMESTRE	USUÁRIOS	FÓRUNS	DISCUSSÕES	POSTAGENS
201402	1632	41	198	369
201501	1848	15	152	203
201502	1894	11	61	93
201601	1968	64	585	1027

As medidas resumo obtidas dos dados analisados estão disponíveis na Tabela 4.

Tabela 4. Medidas resumo dos dados analisados.

MEDIDA	USUÁRIOS	FÓRUNS	DISCUSSÕES	POSTAGENS
Média	1835,5	16,75	126,5	213,75
Erro padrão	72,1959	5,75	57,1114	98,4943
Mediana	1871	12	65	88
Moda	Não há	Não há	Não há	Não há
Desvio padrão	144,3918	16,2635	161,5355	278,5841
Variância da amostra	20849	264,5	26093,7143	77609,0714
Curtose	2,1048	3,8839	5,4955	4,7401
Assimetria	-1.3064	1,9009	2,2834	2,1432
Intervalo	336	49	486	821
Mínimo	1632	4	18	31
Máximo	1968	53	504	852
Soma	7342	134	1012	1710
Contagem	8	8	8	8

A seleção da Unidade educacional para a análise preditiva reduziu o número de medidas para quatro. Para permitir um melhor resultado na análise, o grau de granulari-

dade da pesquisa foi aumentado com o atributo `mdl_forum.timemodified` sendo discretizado no formato `AnoSemestreMês (AAAASSMM)`. As amostras foram então coletadas pelo agrupamento conjuntivo dos atributos `AnoSemestreMês` e `Turma`, este último sendo obtido a partir do nome do curso disponível em `mdl_course.shortname`. Esta alteração aumentou o número de amostras de quatro para oitenta e uma.

7. Resultados encontrados

Foram efetuadas três análises: no primeiro modelo (Figura 1) analisado considerou-se como variável dependente (y) a quantidade de Fóruns em relação ao semestre letivo (semestre 1 a 8), que representa a variável independente (x) do modelo. O segundo (Figura 2) e terceiro (Figura 3) modelos consideram a quantidade de discussões e a quantidade de postagens como variáveis dependentes (y), respectivamente.

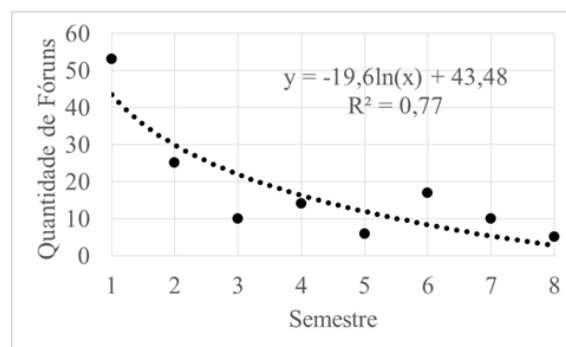


Figura 1. Quantidade de fóruns em relação ao semestre letivo para a Unidade educacional.

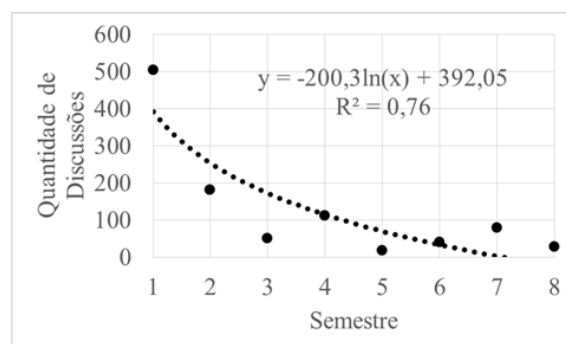


Figura 2. Quantidade de discussões em fóruns em relação ao semestre letivo na Unidade educacional.

Os coeficientes de determinação dos três modelos analisados (R^2) indicam forte correlação entre as variáveis analisadas: quantidade de fóruns e semestre letivo, quantidade de discussões em fóruns e semestre letivo, e quantidade de postagens em fóruns e semestre letivo.

A Tabela 5 apresenta os resíduos individuais e total por semestre e por quantidade de fóruns, de discussões e de postagens. O resíduo total foi calculado pelo método da Soma Quadrática.

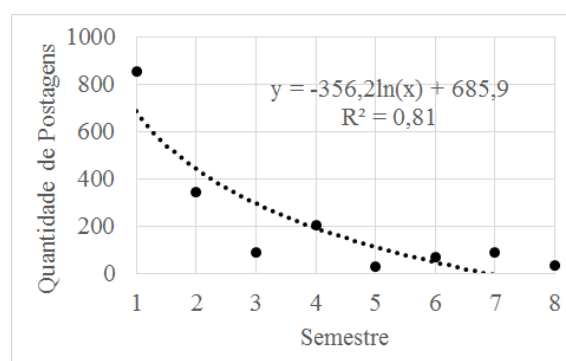


Figura 3. Quantidade de postagens nos fóruns em relação ao semestre letivo da Unidade educacional.

Tabela 5. Resíduos por semestre e por quantidade de fóruns, de discussões e de postagens. Resíduo total pelo método da Soma Quadrática.

SEMESTRE	RESÍDUO FÓRUNS	RESÍDUO DISCUSSÕES	RESÍDUO POSTAGENS
1	9,52	111,5	166,1
2	-4,894	-72,663	-94,001
3	-11,947	-121,448	-207,574
4	-2,308	-3,825	11,898
5	-5,935	-52,130	-81,618
6	8,638	7,389	20,325
7	4,660	76,266	96,233
8	2,277	51,012	88,797
TOTAL (SQ)	399,40	43667,154	103874,424

8. Conclusões

Dos resultados obtidos foi possível identificar a diminuição da utilização dos recursos de fóruns, discussões e postagens no Moodle no decorrer dos cursos na unidade educacional observada. Embora o uso de fóruns seja um indicador da interatividade entre alunos e entre alunos e tutores, este não é o único indicador [Dougiamas and Taylor 2003, Romero et al. 2008]. Uma análise do uso de outras atividades, tais como *view*, *quiz*, *email*, *chat*, *wiki* e, *blog* poderia contribuir na predição no resultado final [Zacharis 2015] e na manutenção da motivação do aluno [Jaggars and Xu 2011].

Oportunidades de aprofundamento do presente estudo podem ser a análise qualitativa das mensagens dos fóruns, a própria análise conjunta das atividades *view*, *quiz*, *email*, *chat*, *wiki* e, *blog* como indicativos mais seguros de interatividade, a ampliação desta análise para outras Unidades Educacionais da Instituição de Ensino Superior e a análise da contribuição das atividades do Moodle que contribuem para a nota final de alunos da modalidade de aprendizagem mista [Zacharis 2015].

Referências

Ally, M. (2004). Foundations of educational theory for online learning. *Theory and practice of online learning*, 2:15–44.

- Anderson, T. (2008). Towards a theory of online learning. *Theory and practice of online learning*, 2:15–44.
- Andrews, T. (2012). What is social constructionism. *Grounded theory review*, 11(1):39–46.
- Baker, R., Isotani, S., and Carvalho, A. (2011). Mineração de dados educacionais: Oportunidades para o brasil. *Revista Brasileira de Informática na Educação*, 19(02):03.
- Black, E. W., Dawson, K., and Priem, J. (2008). Data for free: Using lms activity logs to measure community in online courses. *The Internet and Higher Education*, 11(2):65–70.
- Casey, K. and Gibson, P. (2010). (m)oodles of data: Mining moodle to understand student behaviour. *International Conference on Engaging Pedagogy 2010 (ICEP10)*.
- Coates, H., James, R., and Baldwin, G. (2005). A critical examination of the effects of learning management systems on university teaching and learning. *Tertiary education and management*, 11:19–36.
- Dillenbourg, P., Schneider, D., and Synteta, P. (2002). Virtual learning environments. In *3rd Hellenic Conference "Information & Communication Technologies in Education"*, pages 3–18. Kastaniotis Editions, Greece.
- Dougiamas, M. and Taylor, P. (2003). Moodle: Using learning communities to create an open source course management system.
- Hershkovitz, A. and Nachmias, R. (2011). Online persistence in higher education web-supported courses. *The Internet and Higher Education*, 14(2):98–106.
- IEDMS (2011). International education data mining society. <http://www.educationaldatamining.org>.
- Iglesias-Pradas, S., Ruiz-de Azcárate, C., and Agudo-Peregrina, Á. F. (2015). Assessing the suitability of student interactions from moodle data logs as predictors of cross-curricular competencies. *Computers in Human Behavior*, 47:81–89.
- Jaggars, S. S. and Xu, D. (2011). Online and hybrid course enrollment and performance in washington state community and technical colleges.
- Lara, J. A., Lizcano, D., Martínez, M. A., Pazos, J., and Riera, T. (2014). A system for knowledge discovery in e-learning environments within the european higher education area—application to student data from open university of madrid, udim. *Computers & Education*, 72:23–36.
- Littlejohn, A., Hood, N., Milligan, C., and Mustain, P. (2016). Learning in moocs: Motivations and self-regulated learning in moocs. *The Internet and Higher Education*, 29:40–48.
- Młynarska, E., Greene, D., and Cunningham, P. (2016). Indicators of good student performance in moodle activity data. *arXiv preprint arXiv:1601.02975*.
- Moniz, L. (2015). Arquitetura do banco de dados do moodle 3: Análise sobre mudança da estrutura. in: Palestras ministradas no moodle moot. <http://comunidade.badiu.com.br/mod/forum/discuss.php?d=189>.

- Montgomery, D. C., Peck, E. A., and Vining, G. G. (2001). *Introduction to linear regression analysis*. John Wiley & Sons.
- Moodle (2016a). Moodle history. <https://docs.moodle.org/30/en/History>.
- Moodle (2016b). Moodle statistics. <https://moodle.net/stats/>.
- Nandi, D., Hamilton, M., Harland, J., and Warburton, G. (2011). How active are students in online discussion forums? In *Proceedings of the Thirteenth Australasian Computing Education Conference-Volume 114*, pages 125–134. Australian Computer Society, Inc.
- Park, Y., Yu, J. H., and Jo, I.-H. (2016). Clustering blended learning courses by online behavior data: A case study in a korean higher education institute. *The Internet and Higher Education*, 29:1–11.
- Piccoli, G., Ahmad, R., and Ives, B. (2001). Web-based virtual learning environments: A research framework and a preliminary assessment of effectiveness in basic it skills training. *MIS quarterly*, pages 401–426.
- Romero, C., González, P., Ventura, S., Del Jesús, M. J., and Herrera, F. (2009). Evolutionary algorithms for subgroup discovery in e-learning: A practical application using moodle data. *Expert Systems with Applications*, 36(2):1632–1644.
- Romero, C., Ventura, S., and García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education*, 51(1):368–384.
- Siemens, G. (2010). Call for papers of the 1st international conference on learning analytics & knowledge (lak 2011). <https://tekri.athabascau.ca/analytics>.
- Siemens, G. and Baker, R. S. (2012). Learning analytics and educational data mining: towards communication and collaboration. In *Proceedings of the 2nd international conference on learning analytics and knowledge*, pages 252–254. ACM.
- Sin, K. and Muthu, L. (2015). Application of big data in education data mining and learning analytics—a literature review. *ICTACT Journal on Soft Computing*, 5(4):1–035.
- Vrasidas, C. (2000). Constructivism versus objectivism: Implications for interaction, course design, and evaluation in distance education. *International Journal of Educational Telecommunications*, 6(4):339–362.
- Watson, W. R. and Watson, S. L. (2007). What are learning management systems, what are they not, and what should they become. *TechTrends*, 51(2):29.
- You, J. W. (2016). Identifying significant indicators using lms data to predict course achievement in online learning. *The Internet and Higher Education*, 29:23–30.
- Zacharis, N. Z. (2015). A multivariate approach to predicting student outcomes in web-enabled blended learning courses. *The Internet and Higher Education*, 27:44–53.