



Análise do engajamento de estudantes com base na Distância Transacional a partir da Mineração de Dados Educacionais

João Carlos Sedraz Silva – UNIVASF, joao.sedraz@univasf.edu.br
Jorge Luis Cavalcanti Ramos – UNIVASF, jorge.cavalcanti@univasf.edu.br
Rodrigo Lins Rodrigues – UFRPE, rodrigo.linsrodrigues@ufrpe.br
Fernando da Fonseca de Souza – UFPE, fdfd@cin.ufpe.br
Alex Sandro Gomes – UFPE, asg@cin.ufpe.br

Resumo. Este estudo apresenta uma proposta de análise do engajamento de estudantes em um curso online. Foram usados dados de um curso de pós-graduação de uma universidade pública brasileira. O método seguiu o processo de Mineração de Dados Educacionais aplicada em dois ciclos, a partir da identificação dos construtos da distância transacional nos dados coletados, da utilização de métricas definidas pela análise de redes sociais e da aplicação de regressão logística para obtenção de modelos representativos do engajamento e dos fatores que o influenciam. Os resultados apontaram caminhos para intervenções que poderiam ser feitas para o aumento no nível de engajamento dos estudantes.

Palavras-chave: Engajamento, Mineração de Dados Educacionais, Distância Transacional

Analysis of engagement of students based on Transactional Distance from the Educational Data Mining

Abstract. This study presents an analysis of the proposal of engaging students in an online course. They used data from a postgraduate degree from a Brazilian public university. The method followed the process of Educational Data Mining applied in two stages, from the identification of the constructs of transactional distance on the data collected, using metrics defined by social network analysis and application of logistic regression to obtain representative models engagement and the factors that influence it. The results showed ways to interventions that could be made to increase the students' level of engagement.

Keywords: Engagement, Educational Data Mining, Learning Activities, Transactional Distance.

1. Introdução

Um dos conceitos oriundos da educação presencial adaptado e avaliado no contexto da educação a distância é o de engajamento dos alunos. A literatura aponta que o engajamento surge da interação do sujeito com o contexto educacional e que responde às mudanças no ambiente (Finn e Rock, 1997).

Diversos trabalhos vêm sendo desenvolvidos com o propósito de qualificar e avaliar o engajamento de alunos em atividades de aprendizagem. Há um consenso na literatura de que esse engajamento dos estudantes é necessário para promover resultados positivos na sua aprendizagem (Harris, 2008).

Assim, para contribuir com os estudos sobre o engajamento em ambientes de educação a distância, o objetivo deste trabalho é analisar os aspectos que afetam o grau de engajamento de estudantes em diálogos educacionais de um curso online, segundo o conceito de engajamento definido pelas métricas da análise de redes sociais (SNA) (Medeiros *et al.*, 2013).

Utilizando como elemento norteador a Teoria da Distância Transacional (Moore, 1993), o estudo é baseado em dados dos estudantes do curso de Pós-Graduação em Gestão Pública, oferecido por educação a distância em uma universidade pública no Brasil. A escolha desse curso deve-se a sua presença na maioria das universidades públicas brasileiras que, desde 2010, já ofertaram mais de 100.000 vagas nessa pós-graduação (Capes, 2016).

Além desta introdução, o trabalho está organizado com mais quatro seções, que apresentam a fundamentação deste estudo, o método utilizado, a discussão dos resultados e as considerações finais da pesquisa.

2. Teoria da Distância Transacional

Segundo Moore (1993), a transação em educação a distância é a interação de professores e estudantes em ambientes que têm a característica de serem espacialmente separados um do outro. Com essa separação, surge um espaço psicológico e comunicacional a ser transposto, um espaço de potenciais mal-entendidos na relação entre os professores e os estudantes. Este espaço psicológico e comunicacional é a distância transacional (TD).

A Teoria da Distância Transacional preconiza que os cursos a distância podem ser avaliados segundo uma medida da TD, a qual está fundamentada em três construtos: **diálogo, estrutura do curso e autonomia**.

O diálogo é o construto usado para descrever uma interação positiva entre os participantes de um curso, com o objetivo de aperfeiçoar a aprendizagem. A estrutura expressa o nível de flexibilidade das estratégias pedagógicas e quanto o curso pode ser personalizado aos interesses dos estudantes. A autonomia avalia a liberdade que os estudantes possuem para decidirem sobre o seu ritmo e processo de aprendizagem (Moore, 1993).

2.1. Medidas dos construtos da distância transacional

As decisões do professor sobre a estrutura do curso promovem impactos no comportamento de participantes de um programa educacional. Em razão dessas decisões, cada estudante pode apresentar um comportamento distinto, que pode ser analisado por meio de variáveis associadas aos construtos diálogo e autonomia.

As medidas do comportamento individual dos estudantes, obtidas por variáveis relacionadas ao diálogo e a autonomia, podem refletir um estado do curso durante a sua realização. Essas medidas permitem que professores percebam um distanciamento de alguns aprendizes e, a partir disso, possam intervir para reverter situações indesejadas no curso (Horzum, 2011).

2.2. Contribuições da Análise de Redes Sociais para a avaliação do diálogo

Os Sistemas Gerenciadores de Aprendizagem (LMS), que registram dados sobre as atividades dos estudantes em cursos online, favorecem a obtenção de medidas

quantitativas sobre as interações (Joksimovic *et al.*, 2015). Os registros dos LMS guardam informações que podem ser analisadas por métricas de Análise de Redes Sociais (SNA), que permitem a compreensão sobre a eficácia das estratégias educacionais e, especialmente, sobre os diálogos educacionais dos estudantes (Medeiros *et al.*, 2013; Dawson, 2008).

No trabalho de Medeiros *et al.* (2013), são descritas as métricas de SNA listadas em (1), (2) e (3) usadas neste estudo para avaliar o comportamento dos estudantes em diálogos estabelecidos por meio dos recursos disponíveis nos LMS.

$$C_D(n_i) = X_{i+} / (g-1) \quad (1)$$

$$dI_{ni} = Y_{i+} / (g-1) \quad (2)$$

$$dO_{ni} = Y_{+i} / (g-1) \quad (3)$$

O grau de visibilidade de cada estudante ($C_D(n_i)$) é medido pela razão entre o número de conexões estabelecidas (X_{i+}) e a quantidade de conexões possíveis, representada pelo número total de participantes do curso menos um ($g-1$). A métrica dI_{ni} fornece o grau de prestígio do estudante, que é a razão entre o número de mensagens recebidas pelo aprendiz (Y_{i+}) e a quantidade de conexões possíveis. Por fim, a métrica dO_{ni} avalia o grau de engajamento do estudante, calculado pela razão entre o número de diálogos iniciados (Y_{+i}) pelo estudante e a quantidade de conexões possíveis.

Na próxima seção, é apresentado o método utilizado neste trabalho para avaliar os aspectos que afetaram o grau de engajamento dos estudantes em um curso online.

3. Método

3.1. Contexto do Estudo

A pesquisa descrita neste artigo é um estudo de caso longitudinal que, segundo Yin (2013), justifica-se em situações contemporâneas onde é necessária a investigação de uma mesma unidade de análise em dois ou mais pontos diferentes do tempo. Foram analisados dados de estudantes que participaram de 12 disciplinas de um curso de Pós-Graduação em Gestão Pública, oferecido no formato online por uma universidade pública no Brasil, no período de 2014 a 2015.

As disciplinas do curso foram realizadas de maneira sequencial e cada uma teve uma duração de três semanas. Os conteúdos e as atividades de aprendizagem das disciplinas foram fornecidos por meio do LMS Moodle.

3.2. Objetivo

Analisar os aspectos que afetam o grau de engajamento de estudantes em diálogos educacionais de um curso on-line, a partir de construtos da distância transacional combinados com algumas métricas de SNA.

3.3. Participantes

No período em que a pesquisa foi realizada, o curso analisado possuía 150 estudantes matriculados. Dentre esses, foram analisados os dados de 113 estudantes, que realizaram atividades em todas as disciplinas oferecidas pelo curso. Os dados do perfil dos participantes desta pesquisa estão indicados na Tabela 1.

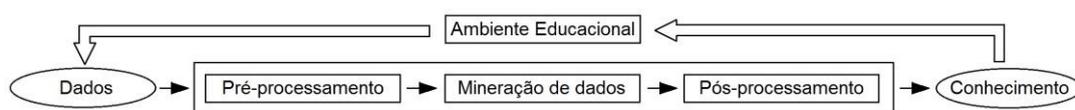
Tabela 1. Dados do perfil dos participantes da pesquisa

Variáveis do perfil	Alternativas	N	%
Experiência anterior em curso on-line.	Sim	51	45.13
	Não	62	54.87
Faixa etária do estudante.	Entre 20 e 30 anos	57	50.44
	Entre 31 e 40 anos	41	36.28
	Entre 41 e 50 anos	11	9.74
	Mais de 50 anos	04	3.54
Sexo do estudante.	Feminino	62	54.87
	Masculino	51	45.13

* N = Número de estudantes para cada alternativa.

3.4. Procedimentos

Para atender o objetivo da pesquisa, os procedimentos metodológicos adotados foram baseados nas etapas da Mineração de Dados Educacionais (EDM, do inglês *Educational Data Mining*) (García *et al.*, 2011), apresentadas na Figura 1.

**Figura 1. Processo de Mineração de Dados Educacionais (García *et al.*, 2011)**

A EDM é definida como a área responsável pelo desenvolvimento de métodos de extração de conhecimento a partir de bases de dados de ambientes educacionais (Baker *et al.*, 2011). Romero e Ventura (2013) definiram a EDM como a aplicação de técnicas de *Data Mining* em dados provenientes de plataformas ou ambientes de educação *online*. Dentro do contexto deste estudo, as etapas da EDM estão descritas a seguir.

3.4.1. Dados coletados

Os dados analisados neste estudo foram extraídos do LMS Moodle adotado nas disciplinas e do cadastro com o perfil dos estudantes obtido na matrícula. No LMS, foram identificados 1.566.141 registros de interações dos estudantes.

3.4.2. Pré-processamento

Os dados coletados foram integrados e filtrados, reunindo em uma mesma base apenas os registros de estudantes que participaram de todas as disciplinas do curso. Assim, para cada uma das 12 disciplinas do curso, foram selecionadas informações de 113 estudantes, o que gerou 1.356 casos de análise.

Tabela 2. Descrição das variáveis avaliadas no estudo

Construto	Variável	Descrição
Diálogo	DE	Grau de engajamento para o estudante, por disciplina.
	DP	Grau de prestígio para o estudante, por disciplina. DV
		Grau de visibilidade para o estudante, por disciplina.
Autonomia	AA	Quantidade de acessos do estudante ao ambiente virtual, por disciplina.
	AT	Tempo total utilizado pelo estudante para visualização de recursos do ambiente virtual, por disciplina.
Dados do perfil	EXP	Experiência anterior em curso on-line.
	IDADE	Faixa etária do estudante.
	SEXO	Sexo do estudante.

Os registros das interações dos estudantes no LMS foram codificados em variáveis que representaram os construtos da Teoria da Distância Transacional e aos dados do perfil, conforme estão descritas na Tabela 2.

Segundo Moore (1993), o diálogo educacional se refere às interações positivas, em que os estudantes estabelecem relações com o objetivo de aprimorar a

aprendizagem. No curso analisado, os diálogos dos estudantes foram materializados por meio de discussões em fóruns. Dessa forma, a unidade de referência adotada nesta pesquisa para o diálogo foi um par de mensagens em um fórum, em que um estudante respondeu ou comentou a uma mensagem de outro estudante. A partir dessa referência, para calcular os valores das variáveis de diálogo indicadas na Tabela 2, foram aplicadas as métricas de Análise de Redes Sociais sugeridas por Medeiros *et al.*(2013).

Com relação ao construto autonomia, em todas as disciplinas, os estudantes tiveram a liberdade para estabelecerem com que frequência e por quanto tempo iriam explorar os objetos de aprendizagem disponíveis. Assim, semelhante à abordagem realizada por Joksimovic *et al.*(2015), para as variáveis de autonomia, foram considerados o tempo de visualização (AT) dos recursos e o número de acessos ao ambiente virtual (AA). Na variável AT, foram consideradas as ações praticadas pelos estudantes no acesso ao conteúdo ou material de apoio das disciplinas. Dessa maneira, foram contabilizadas as ações listadas na Tabela 3, registradas pelo LMS.

Tabela 3. Registros do LMS considerados para a determinação do tempo de visualização

Módulo	Ação
Assignment	View
Forum	View discussion
Page	View
URL	View.
Resource	View

A etapa de pré-processamento foi concluída com a transformação dos dados, que permitiu a aplicação da técnica de mineração selecionada na pesquisa e, além disso, simplificou a interpretação dos modelos obtidos. Com a transformação, as medidas das variáveis foram convertidas em valores binários. Na variável SEXO, o valor 1 representa o sexo masculino e 0 o sexo feminino. Já para DE, DP, DV, AA, AT, EXP e IDADE, 1 indica que o valor da instância é maior do que a mediana da variável na disciplina, enquanto 0 indica que o valor da instância é menor ou igual do que a mediana da variável na disciplina. A mediana foi escolhida como medida de tendência central das variáveis, pois não sofre grande influência de *outliers*.

3.4.3. Mineração de dados

O propósito desta fase foi obter modelos de classificação do grau de engajamento dos estudantes (DE), com base nas outras variáveis, que estão associadas ao diálogo, autonomia e dados do perfil. Para geração e testes dos modelos, os casos de análise foram divididos em duas amostras, com 30% dos casos reservados para o treinamento e 70% para os testes.

Para a determinação dos modelos, foi utilizada a técnica de Regressão Logística, que estuda o comportamento de uma variável dependente binária (binomial) com base em um conjunto de variáveis independentes numéricas ou categóricas (Kleinbaum e Klein, 2010).

$$\text{logit}(p_i) = \ln \left(\frac{p_i}{1 - p_i} \right) = \beta_0 + \beta_1 x_{1,i} + \dots + \beta_k x_{k,i} \quad (4)$$

Os modelos de Regressão Logística são formulados por (4), onde $\beta_0, \beta_1, \dots, \beta_k$, são os coeficientes das variáveis que explicam a ocorrência de um determinado evento. Existem algumas abordagens para auxiliar na seleção de variáveis de modelos obtidos por essa técnica. Neste trabalho, foi utilizada a abordagem *stepwise*. O método *stepwise* é o processo de estimativa de modelos estatísticos em que as variáveis independentes

são adicionadas ou retiradas do modelo de acordo com o poder de discriminação que agregam ao grupo de variáveis explicativas (Hair *et al.*, 2009).

3.4.4. Pós-processamento

Nesta etapa os modelos obtidos pelos algoritmos de EDM devem ser compreensíveis e úteis para o processo de tomada de decisão. A validação dos modelos desenvolvidos foi realizada a partir dos valores de acurácia e da área sob a Curva ROC. A Curva ROC é uma representação gráfica que esclarece o desempenho de um sistema classificador binário e, além disso, permite estudar a variação da sensibilidade e especificidade para diferentes valores de corte (Fawcett, 2006).

3.4.5. Conhecimento

A cada ciclo, a Mineração de Dados Educacionais filtra e transforma dados em conhecimento, que pode apoiar a tomada de decisão durante o planejamento de novas estratégias pedagógicas para melhoria da aprendizagem, mesmo durante a execução do curso. Para explorar e avaliar a vantagem desse processo nesta pesquisa, os dados foram analisados em dois ciclos.

No primeiro ciclo, foi realizada a análise de dados da primeira à décima primeira disciplina. Ao final desse ciclo, com base no modelo gerado pela Regressão Logística, foram avaliados os impactos das variáveis independentes sobre o grau de engajamento dos estudantes.

O segundo ciclo ficou restrito aos dados da décima segunda disciplina, que teve o seu planejamento orientado pelo conhecimento gerado pelo primeiro ciclo da EDM. Nessa disciplina, as estratégias pedagógicas foram ajustadas para potencializar a participação dos estudantes em diálogos educacionais. A análise dos dados do segundo ciclo gerou outro modelo, com uma nova configuração das variáveis independentes que explicaram o grau de engajamento dos estudantes.

4. Resultados e Discussões

As análises estatísticas deste trabalho foram realizadas por meio do *software* estatístico R¹, com o suporte do pacote DMwR², a partir dos dois ciclos da Mineração de Dados Educacionais apresentados na Seção 3.

4.1. Resultados do primeiro ciclo de EDM

A Tabela 4 apresenta o modelo de Regressão Logística obtido no primeiro ciclo de EDM conduzido na pesquisa, que utilizou os dados coletados das onze primeiras disciplinas para explicar os aspectos que afetavam o grau de engajamento dos estudantes.

Tabela 4. Modelo para a classificação do grau de engajamento dos estudantes (primeiro ciclo)

Variável	β	Exp(β)	Std. Err.	Z	p> z
(Intercept)	-3.0297	---	0.2185	-13.863	< 2e-16*
DP	0.7265	2.0678	0.2221	3.271	1.07e-03*
DV	3.4160	30.4474	0.2167	15.764	< 2e-16*
AA	0.2699	1.3098	0.1909	1.414	0.15737
AT	0.9070	2.4769	0.1919	4.726	2.29e-06*
EXP	0.4948	1.6402	0.1941	2.549	0.01082*
IDADE	0.4495	1.5675	0.1947	2.309	0.02096*
SEXO	-0.1108	0.8951	0.1849	-0.599	0.54903

* Significância < 0,05, β = coef. logístico, Exp(β) coef. logístico exponencial.

¹ <https://cran.r-project.org>

² <https://cran.r-project.org/web/packages/DMwR/>

O modelo desenvolvido, referente ao primeiro ciclo, obteve uma taxa de acurácia de 87.53%. Embora este modelo tenha tido boa taxa de acurácia, apenas as variáveis DP, DV, AT, EXP e IDADE foram incluídas no modelo, pois apresentaram uma maior significância ($p < 0.05$).

Para o modelo do primeiro ciclo, a área abaixo da Curva ROC é de 0.912 (Figura 2), valor que demonstra um poder de discriminação considerado excelente para estimar o grau de engajamento dos estudantes (Fawcett, 2006).

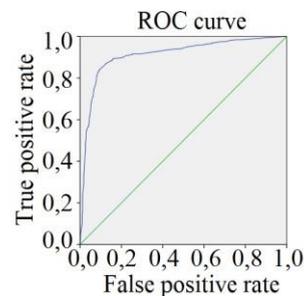


Figura 2. Curva ROC para o primeiro ciclo

O sinal dos coeficientes logísticos (β) indica o sentido da relação (Kleinbaum e Klein, 2010). No modelo da Tabela IV, a relação positiva de DP, DV, AT, EXP e IDADE significa que um aumento nessas variáveis é associado com o aumento na probabilidade do estudante pertencer ao grupo de maior grau de engajamento.

Os coeficientes logísticos exponenciais ($\text{Exp}(\beta)$) refletem diretamente a magnitude da variação no valor da probabilidade de um evento acontecer [Hair *et al.* 2009]. No modelo obtido com os dados das onze primeiras disciplinas, DV representa a variável que mais contribuiu para explicar a ocorrência de estudantes com maior grau de engajamento.

Pelo valor de $\text{Exp}(\beta)$ apresentado na Tabela 4, observando o grau de visibilidade e mantendo constante as demais variáveis, estudantes com $DV=1$ (maior visibilidade) tem uma probabilidade de pertencer ao grupo com maior grau de engajamento cerca de 30 vezes maior que os estudantes com $DV=0$ (menor visibilidade).

4.2. Resultados do segundo ciclo de EDM

O conhecimento gerado por meio do modelo obtido no primeiro ciclo de EDM realizado permitiu inferir que, para as estratégias adotadas pelo curso e os estudantes observados, o grau de visibilidade tem uma relação positiva com o grau de engajamento em diálogos educacionais. Em razão disso, as estratégias pedagógicas da última disciplina ofertada (D12) foram ajustadas para promover uma maior visibilidade dos estudantes. Assim, conforme recomendação de Medeiros (2013), o professor passou a adotar um número maior de fóruns e a encorajar os estudantes, especialmente os de menor grau de visibilidade, a criarem novos tópicos de discussão ao invés de somente comentar os tópicos existentes.

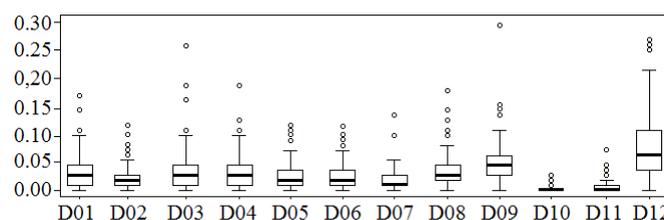


Figura 3. Box Plot do grau de engajamento por disciplina

Com os ajustes promovidos para aumentar a visibilidade dos estudantes, foi possível perceber uma melhoria no grau de engajamento (Figura 3), o que reforçou as evidências apresentadas no modelo encontrado no primeiro ciclo de EDM.

Embora graficamente seja possível perceber diferenças no grau de engajamento, foram feitos testes estatísticos para verificar se a última disciplina (D12) apresentou diferenças significativas em relação às demais disciplinas. Nessa análise, foi utilizado o teste de comparação múltipla de Tukey (Kleinbaum e Klein, 2010) que, para um nível de significância de 5%, confirmou a diferença significativa do grau de engajamento dos estudantes na última disciplina em relação a todas as outras disciplinas (Tabela 5).

Tabela 5. Valores obtidos pelo teste de Tukey

Disciplina	Média*	n-valor	Disciplina	Média*	n-valor
D01	0.0337	0.0000	D07	0.0188	0.0000
D02	0.0234	0.0000	D08	0.0376	0.0000
D03	0.0345	0.0000	D09	0.0526	1.6196E-10
D04	0.0341	0.0000	D10	0.0022	0.0000
D05	0.0272	0.0000	D11	0.0079	0.0000
D06	0.0286	0.0000	D12	0.0808	1.0000

* Média do grau de engajamento dos estudantes na disciplina.

** Valores obtidos após comparação com os dados da disciplina D12.

O segundo ciclo de EDM foi realizado a partir dos dados da disciplina D12. O modelo obtido nesse novo ciclo de mineração teve um elevado nível de discriminação da variável dependente, com uma taxa de acurácia de 89,38% e uma área sob a Curva ROC igual a 0.971.

Os coeficientes das variáveis que explicam o grau de engajamento dos estudantes na disciplina D12 estão representados na Tabela 6.

Tabela 6. Modelo para a classificação do grau de engajamento dos estudantes (segundo ciclo)

Variável	β	Exp(β)	Std. Err.	Z	n> z
(Intercept)	-3.6760	---	0.9705	-3.788	1,52e-04*
DP	-0.8724	0.4179	1.2274	-0.711	0.477199
DV	6.0493	423.8163	1.3865	4.363	1.28e-05*
AA	-0.7645	0.4656	0.8071	-0.947	0.343547
AT	2.6885	14.7096	0.8510	3.159	0.001582*
EXP	1.9527	7.0477	0.9635	2.027	0.042686*
IDADE	-1.3458	0.2603	0.9347	-1.440	0.149934
SEXO	0.7926	2.2091	0.8760	0.905	0.365569

* Significância < 0,05, β = coef. logístico, Exp(β) coef. logístico exponencial.

De acordo com os valores de Exp(β) da Tabela 6, para uma significância de 0.05, na configuração da disciplina D12, as variáveis DV, AT e EXP possuem uma relação positiva com o grau de engajamento. Com esse conhecimento, em uma outra oferta da disciplina, as estratégias pedagógicas poderão ser ajustadas para promover um novo incremento no grau de engajamento.

5. Considerações Finais

Os procedimentos apresentados neste trabalho, com informações sobre o processo de mineração de dados, representam uma contribuição para a tomada de decisão de professores em cursos online, principalmente durante o planejamento de disciplinas.

Uma abordagem para aplicar um conceito da educação presencial na modalidade a distância foi apresentada, apontando caminhos para sua quantificação e avaliação.

A partir da descoberta das variáveis que afetam o grau de engajamento em diálogos educacionais, os professores podem dispor de informações importantes para decidir sobre as melhores estratégias pedagógicas que poderão aumentar as chances de participação dos estudantes.

Ressaltamos que, por se tratar de um estudo de caso, os resultados deste trabalho estão limitados aos dados e a metodologia do curso analisado, mas não invalida sua utilização para outros cenários educacionais, com as devidas adaptações.

Para trabalhos futuros, propõe-se o desenvolvimento de um *plugin* para o LMS que automatize a obtenção e visualização das contribuições de cada variável para o engajamento dos alunos, assim como a aplicação do método utilizado na pesquisa em outras instituições, cursos e ambientes para validação dos seus resultados.

Agradecimentos

Esta pesquisa é apoiada pela FACEPE-APQ No. 0525-1.03/14. Os autores também agradecem ao apoio da PRPPG/UFRPE (006999/2015-28) e PRPPGI/UNIVASF (Edital 18/2013). Alex Sandro Gomes é bolsista DT Nível 2/CNPq, processos nº 310466/2012-1 e nº 475634/2013-6.

Referências bibliográficas

- BAKER, R. *et al.* **Mineração de dados educacionais: oportunidades para o Brasil.** Revista Brasileira de Informática na Educação, 19(2). 2011.
- CAPES. **Dados da educação a distância no Brasil.** Disponível em: <http://sisrel.capes.gov.br/index.php/rel>. Acessado em: 10/04/2016.
- DAWSON, S. **A Study of the Relationship between Student Social Networks and Sense of Community.** Educational Technology and Society, 11(3), 224-238. 2008.
- FAWCETT, T. **An introduction to ROC analysis.** Pattern recognition letters, 27(8), 861-874. 2006.
- FINN, J., ROCK, D. **Academic success among students at risk for school failure.** Journal of applied psychology, 82(2), 221. 1997.
- MEDEIROS, F. **Uma abordagem de monitoramento abrangente das interações sociais em ambientes colaborativos virtuais de aprendizagem como suporte a presença docente.** Recife: UFPE, 2013. Tese de Doutorado.
- GARCÍA, E. *et al.* **A collaborative educational association rule mining tool.** The Internet and Higher Education, 14(2), 77-88. 2011.
- HAIR, J. *et al.* **Multivariate data analysis.** Vol. 6. Upper Saddle River, NJ: Pearson Prentice Hall. 2009.
- HARRIS, L. R. **A phenomenographic investigation of teacher conceptions of student engagement in learning.** The Australian Educational Researcher, 35(1), 57-79. 2008.
- HORZUM, M. **Developing Transactional Distance Scale and Examining Transactional Distance Perception of Blended Learning Students in Terms of Different Variables.** Educational sciences: Theory and practice, 11(3), 1582-1587. 2011.
- JOKSIMOVIC, S. *et al.* **Learning at distance: Effects of interaction traces on academic achievement.** Computers and Education, 87, 204-217. 2015.
- KLEINBAUM, D., KLEIN, M. **Analysis of matched data using logistic regression.** In Logistic Regression (pp. 389-428). Springer New York. 2010.



MEDEIROS, F. *et al.* **Architecture for Social Interactions Monitoring in Collaborative Learning Environments as a Support for the Teacher's Awareness.** In IEEE 13th International Conference on Advanced Learning Technologies (pp. 123-127). IEEE. 2013.

MOORE, M. **Theory of transactional distance.** In Theoretical principles of distance education. 1993.

ROMERO, C., VENTURA, S. **Data mining in education.** Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 3(1), 12-27. 2013.

YIN, R. **Case study research: Design and methods.** Sage publications. 2013.