

Investigação acerca do Impacto dos Comportamentos dos Tutores em Turmas de Ensino a Distância

Rafael Castro de Souza¹, Francisco Milton Mendes Neto¹, Araken de Medeiros Santos¹, Laysa Mabel de Oliveira Fontes², Ricardo Alexsandro de Medeiros Valentim²

¹ Programa de Pós-Graduação em Ciência da Computação – UERN/UFERSA
Av. Francisco Mota, 572 - Bairro Costa e Silva - Mossoró, RN - Brasil

² Programa de Pós-Graduação em Engenharia Elétrica e de Computação – UFRN
Campus Universitário - BR 101 - Lagoa Nova - Natal, RN - Brasil

{rafaelcastro, miltonmendes, araken, mabel.fontes}@ufersa.edu.br,
ricardo.valentim@ufrnet.br

Abstract. *The integration of technology in education, combined with its wide use, boosted the stored educational data volume. Researchers realized, whereupon, the possibility of analyzing this large amount of information in order to enable knowledge assessment through it. Thus, this information can, for example, help support decision making or guide the identification of factors that might impact positively or negatively the teaching process. From this perspective, this paper presents a study concerning the influence of tutors' behavior in students participation rate at online education. With the information granted by this work, one can better understand the impact of tutors' behavior at online learning process, enabling pedagogical interventions guided by objective information in order to attenuate the limitations faced by this education mode.*

Resumo. *A inserção da tecnologia na educação, combinada com sua larga utilização, tem gerado um aumento no volume dos dados educacionais armazenados. Diante disso, pesquisadores perceberam a possibilidade de analisar esse grande volume de informações, a fim de extrair conhecimento, de modo que essas informações possam, por exemplo, auxiliar no suporte à tomada de decisão ou ainda a fim de identificar fatores que possam impactar positivamente ou negativamente o processo de ensino. Diante dessa perspectiva, o presente trabalho apresenta um estudo sobre a influência das ações comportamentais dos tutores, de disciplinas na modalidade de ensino a distância, na taxa de participação dos alunos em turmas dessa modalidade de ensino. Com as informações resultantes deste trabalho, pode-se compreender melhor o impacto dos comportamentos dos tutores nas turmas da modalidade de ensino a distância, além de possibilitar intervenções pedagógicas pautadas em informações objetivas, a fim de atenuar os problemas enfrentados por essa modalidade de ensino.*

1. Introdução

Com o surgimento dos sistemas computacionais e, posteriormente, da Internet, tornou-se possível o armazenamento e compartilhamento dos dados em larga escala, em mídias digitais, tais como áudios, vídeos, imagens e textos, que podem ser acessadas remotamente, utilizando o computador como ferramenta mediadora.

Uma vez que os sistemas computacionais provêm uma alta capacidade no processamento de dados, bem como um armazenamento seguro, confiável, barato e rápido da informação, pode-se perceber que isso resultou na inserção da tecnologia nos mais diferentes tipos de ambientes, tais como empresarial, educacional, industrial, médico, dentre outros.

Segundo McAfee e Brynjolfsson (2012), estima-se que 2,5 *exabytes* de dados são criados por dia e que esse número tende a convergir para o dobro a cada 40 meses. Todo esse grande volume de dados gerado recebe a nomenclatura de *Big Data*, que é um termo empregado para referir-se a grandes volumes de dados e que vem aumentando cada dia mais.

Diante da intensificação do volume de dados gerado, pesquisadores começaram a estudar formas de manusear esses dados, com o objetivo de, a partir de análises sobre estes, obter conhecimento, de modo que o conhecimento gerado pudesse auxiliar, por exemplo, a tomada de decisões, a realizar previsões etc. (Mayer-Schönberger e Cukier, 2013).

Quando aplicada no âmbito educacional, esse processo recebe a nomenclatura de *Learning Analytics* (LA). A LA objetiva extrair conhecimento, a partir da análise dos dados educacionais, para proporcionar às partes interessadas (alunos, educadores, administradores e financiadores) uma melhor informação e um profundo conhecimento sobre os fatores que contribuem para o sucesso do aprendiz (Siemens *et al.*, 2011).

Com a LA, portanto, muitas questões relacionadas ao ensino e a aprendizagem podem ser melhor compreendidas e, com isso, resultar na elaboração de melhores estratégias de ensino. Sendo assim, pode-se considerar que a matéria-prima para a aplicação da LA é o grande volume de dados educacionais, visto que o conhecimento só pode ser extraído uma vez que esses dados estejam disponíveis para análise.

Uma das principais formas de obter esses dados é por meio dos *Learning Management System* (LMS). Os LMS são sistemas de software concebidos para apoiar o ensino, por meio da administração de um ou mais cursos (Kumar e Samalia, 2015). Devido estes geralmente armazenarem a maioria das ações de seus usuários (alunos, professores, tutores e administradores), seus bancos de dados tornam-se ideais para a extração e a análise das informações referentes ao processo de ensino.

Diante do exposto, o presente trabalho apresenta resultados de uma análise sobre os comportamentos e ações dos tutores na modalidade de ensino a distância que possam estar positivamente ou negativamente associados aos índices de participação efetiva dos alunos nessas turmas.

Este artigo está organizado da seguinte forma: a Seção 2 descreve a fundamentação teórica da *Learning Analytics*; na Seção 3, é descrito o coeficiente de Correlação de Pearson; a Seção 4 apresenta trabalhos relacionados com a temática deste trabalho; na Seção 5 é apresentado os resultados obtidos, bem como uma discussão

sobre os resultados, e, por fim, a Seção 6 expõe as conclusões do trabalho, bem como os trabalhos futuros.

2. Learning Analytics

A ampla utilização das tecnologias no ensino provocou um aumento na demanda de recursos computacionais, pois, com o barateamento desses recursos, a quantidade de usuários cresceu e o volume de dados relacionado ao uso de sistemas acadêmicos acompanhou esse crescimento.

Com o armazenamento de informações referentes a *logs* de acessos, conversas entre alunos, professores e tutores, registros de notas, atividades, questionários, criação de fóruns, grupos, etc., pesquisadores perceberam a oportunidade de extrair informações que pudessem estar associadas ao processo de ensino e aprendizagem, com a finalidade de compreender melhor os fatores que podem estar associados, mas não somente limitado ao sucesso ou fracasso de cursos, turmas, alunos e práticas pedagógicas.

Diante disso, surgiu o campo de pesquisa denominado *Learning Analytics* (LA), que pode ser definido como sendo a medição, coleta, análise e comunicação de dados sobre os alunos e os seus contextos, para fins de compreensão e otimização da aprendizagem nos ambientes em que esse processo ocorre (Siemens *et al.*, 2011).

Outros autores, como Johnson *et al.* (2011), por exemplo, definem a LA como a interpretação de uma ampla gama de dados, produzidos pelos estudantes, e coletados pelo seu interesse, a fim de avaliar o progresso acadêmico, prever o desempenho futuro e identificar possíveis problemas.

Sendo assim, em concordância com as definições supracitadas, é possível definir a LA como sendo coleta, análise e compreensão das informações relacionadas ao processo de ensino, seja presencial, semipresencial ou a distância, e do ambiente em que este processo ocorre, a fim de proporcionar para as partes envolvidas, seja de forma direta (estudantes, professores, entre outros) ou indireta (instituições de ensino, comunidade científica, etc.), a percepção de fatores que possam influenciar positivamente ou negativamente o ensino, independentemente de práticas pedagógicas que possam estar sendo utilizadas.

Embora a LA seja considerada uma área de pesquisa recente, é possível perceber que seus objetivos e resultados trazem benefícios para a educação (Siemens e Long, 2011) e que, de fato, ainda existem muitas subáreas desse campo a serem exploradas.

Vale salientar ainda que não há restrições sobre as técnicas de análise de dados que podem ser empregadas, dessa forma, para a análise dos dados podem ser utilizadas técnicas estatísticas, modelos preditivos ou descritivos da área de Aprendizado de Máquina (AM), entre outros tipos de técnicas de análise de dados.

A fim de identificar quais comportamentos dos tutores à distância impactam no comportamento da turma, foi utilizado o método estatístico do coeficiente de correlação de variáveis que é exposto na seção a seguir.

3. Coeficiente de Correlação de Pearson

Dado X e Y como sendo duas amostras, como podemos estabelecer uma relação entre X e Y? O campo investigativo que objetiva responder essas questões são os métodos estatísticos da análise de correlação.

A análise de correlação, também conhecida como coeficiente de correlação, indica o grau de variação conjunta entre duas variáveis. Esse grau representa a intensidade e a direção da relação linear ou não-linear entre estas. Esse método atende à necessidade de se estabelecer a existência ou não de uma relação entre variáveis, sem a necessidade de aplicar uma função matemática, pois não existe a distinção entre a variável explicativa e a variável resposta. Em outras palavras, o grau de variação conjunta entre X e Y é igual ao grau de variação entre Y e X (Schumaker, 2014).

Conforme Meissner (2013), os coeficientes de correlação se dividem em três tipos: (i) o coeficiente de correlação de Pearson; (ii) o coeficiente de correlação de Spearman; e (iii) o coeficiente de correlação de Kendall.

Tendo em vista que a natureza dos dados analisados neste trabalho pertence à categoria das variáveis quantitativas, o coeficiente de correlação que mais se adéqua para essa análise é o coeficiente de correlação de Pearson (Rubin, 2012).

O coeficiente de correlação de Pearson tem como origem o trabalho em conjunto de Karl Pearson e seu professor Francis Galton. De forma sucinta, o coeficiente de correlação de Pearson (r) é uma medida de associação linear entre variáveis (Sharma, 2012), e é calculado da seguinte forma:

$$r = \frac{1}{n-1} \times \sum \left(\frac{x_i - \bar{x}}{s_x} \right) \times \left(\frac{y_i - \bar{y}}{s_y} \right) \quad (1)$$

Onde:

- i. n é o número das amostras;
- ii. x_i e y_i são os elementos de cada amostra;
- iii. \bar{x} e \bar{y} representam a média dos elementos das amostras x e y ;
- iv. S_x e S_y indicam o valor do desvio padrão das amostras x e y , respectivamente.

O resultado retornado deverá estar entre +1 e -1. O sinal indica a direção, ou seja, se a correlação é positiva ou negativa, e o valor numérico indica a força da correlação.

Segundo Cohen (1988) *apud* Weinberg e Abramowitz (2002), quando os objetos de estudo são fatores comportamentais, a interpretação para a correlação de Pearson deve ser feita da seguinte forma:

- I. Se $r \geq 0.5$, sendo positivo ou negativo, indica uma forte correlação;
- II. Se $0.3 \leq r < 0.5$, sendo positivo ou negativo, indica correlação moderada;
- III. Se $0.1 \leq r < 0.3$, sendo positivo ou negativo, indica fraca correlação;
- IV. Em último caso, para $0 \leq r < 0.1$, sendo positivo ou negativo, pode-se considerar a correlação nula.

4. Trabalhos Relacionados

Nesta seção são apresentados alguns trabalhos relacionados com este artigo.

Silva *et al.* (2015) apresentam um estudo sobre a análise dos dados das interações em um ambiente virtual de aprendizagem de alunos iniciantes na educação a

distância. Para a análise foram empregadas as técnicas de coeficiente de correlação de Pearson e por meio de técnicas de agrupamento. Ao todo foram analisados dados de 200 alunos de uma turma de Introdução à Educação a Distância (IEaD), do Bacharelado em Administração Pública, ofertado pela UNIVASF, onde foram coletadas e analisadas diversas informações contidas no banco de dados da disciplina dentro do ambiente Moodle.

Já Fidalgo-Blanco *et al.* (2015) propuseram a utilização da LA para analisar o progresso individual de um aluno, dentro do contexto do trabalho em equipe. Para que isso pudesse ser feito, foram coletados dados referentes à troca de mensagens entre estudantes no LMS, número de visualizações de mensagens, dentre outros. Além disso, o sistema identifica os estudantes autores de tópicos e o número de mensagens enviadas por estes, a fim de identificar quais seriam os estudantes ativos e passivos no grupo. O LMS utilizado foi o Moodle, e, ao todo, foram analisados dados de 110 estudantes da Universidade Técnica de Madri. Para a análise dos dados, foi utilizado o coeficiente de correlação de Pearson.

Embora a presente proposta esteja no mesmo campo de pesquisa que os trabalhos supracitados, que é a LA, esta apresenta uma perspectiva bem diferente. Neste trabalho, a aplicação da LA está voltada para a análise das ações dos tutores de cursos a distância, a fim de compreender quais as ações que podem, ou não, ter um grau de influência na participação efetiva dos alunos.

5. Análise e Resultados

Para a análise dos dados, foi construída uma ferramenta de análise na linguagem de programação Java (Oracle, 2016), e esta dispõe de basicamente dois módulos.

O primeiro módulo é responsável pela extração das informações da base de dados e pela normalização dos dados. O segundo módulo é responsável pela aplicação da correlação de variáveis sobre os dados extraídos.

Neste trabalho, os dados foram cedidos pela Secretaria de Educação a Distância (SEDIS) pertencente à Universidade Federal do Rio Grande do Norte (UFRN). Essa instituição usa o Moodle como LMS. Portanto, todas as informações que compõem o *dataset* foram extraídas desse LMS.

Ao todo, foram retornados dados de 62 turmas, pertencentes a 10 cursos de graduação, que são: (i) Bacharelado em Administração Pública; (ii) Licenciatura em Ciências Biológicas; (iii) Licenciatura em Educação Física; (iv) Licenciatura em Física; (v) Licenciatura em Geografia; (vi) Licenciatura em Letras; (vii) Licenciatura em Matemática; (viii) Licenciatura em Pedagogia; (ix) Licenciatura em Química; e (x) Licenciatura em História. A base de dados histórica cedida corresponde aos anos de 2012 a 2013. No total, o *dataset* foi composto por informações de 2.227 alunos e 38 tutores a distância.

As informações dos tutores extraídas e analisadas pela ferramenta foram: (i) Número de tópicos criados nos fóruns; (ii) Média de postagens em tópicos dos fóruns; (iii) Taxa de visualizações em fóruns; e (iv) Taxa de visualizações em tópicos dos fóruns.

Quanto aos comportamentos das turmas, foram analisados os seguintes comportamentos: (i) Taxa de participação da turma nos questionários; (ii) Tempo médio para finalização dos questionários; (iii) Média de postagens em tópicos da turma; (iv)

Média de acesso em fóruns da turma; (v) Média de visualização em tópicos dos fóruns da turma; (vi) Taxa de submissão em tarefas da turma; (vii) Média de postagens em *chats* por aluno da turma; (viii) Média de páginas visualizadas da turma; (ix) Média de arquivos visualizados da turma.

Primeiramente a ferramenta extrai as informações dos cursos cadastrados, dos tutores de cada curso e dos alunos matriculados em cada curso. Após a obtenção dessas informações, é iniciado o procedimento de extração das informações relacionadas aos comportamentos destes.

Após a extração dessas informações, o primeiro módulo as armazenam em um arquivo no formato *.arff* (WEKA, 2016), que servirá como arquivo de entrada para o segundo módulo, que é responsável pelo processamento dessas informações. Para o processamento dos dados foi utilizado o coeficiente de correlação de Pearson e para o nível de significância da análise foi utilizado o valor para $\alpha = 0,05$. Portanto, o nível de confiança é de 95%.

Visando a remoção de valores atípicos (*outliers*), foi utilizada a técnica de Z-Score (Gravetter e Wallnau, 2010), tendo a margem aceitável de Z definida para: $-2.1 < Z < 2.1$, sendo que pela tabela da distribuição normal esse intervalo abrange 96.42% da área. Dessa forma, se um dado tem 96.42% de chances de pertencer a margem de $-2.1 < Z < 2.1$, e ainda assim ficou de fora dessa margem, ao que tudo indica, esse dado é um *outlier* (Warner, 2012).

A análise realizada abrangeu quatro comportamentos dos tutores e dez comportamentos das turmas, gerando, portanto, 40 (4 x 10) resultados de correlações. Porém, a fim de simplificar a exposição dos resultados, as tabelas a seguir irão apresentar as informações das correlações cujas hipóteses alternativas (H_1) foram aceitas. Logo, as demais possibilidades que não constam nas tabelas a seguir indicam que a recomendação foi de aceitar a hipótese nula (H_0).

As tabelas a seguir apresentam os resultados da correlação para cada comportamento do tutor em relação com o comportamento da turma.

Tabela 1. Número de Tópicos Criados pelo Tutor.

Atributos da turma	Número de instâncias sem <i>outliers</i>	Resultado da correlação	Interpretação para a correlação
Média de tópicos criados pela turma	57	0.370 (Aceita H_1)	Correlação Moderada Positiva
Média de acessos em tópicos pela turma	57	0.276 (Aceita H_1)	Correlação Fraca Positiva
Média de <i>posts</i> em chats	57	0.251 (Aceita H_1)	Correlação Fraca Positiva

De acordo com a Tabela 1, podemos concluir que:

- i. Em relação à média de tópicos criados pela turma, a correlação moderada positiva evidencia que a criação de tópicos pelos tutores estimulam também aos alunos criarem seus tópicos de discussão.;
- ii. Em relação à média de visualização de tópicos pela turma, a correlação fraca positiva indica que o aumento da criação dos tópicos por parte dos tutores

influencia em uma melhora na média de acessos em tópicos nos fóruns pelos alunos;

- iii. Em relação à média de *posts* em chats, a correlação fraca positiva indica que a criação de tópicos em fóruns pelos tutores melhora, ainda que fracamente, os índices da média de *posts* em chats pelos alunos.

Tabela 2. Média de Postagens em Tópicos nos Fóruns.

Atributos da turma	Número de instâncias sem outliers	Resultado da correlação	Interpretação para a correlação
Taxa de participação em <i>quizz</i>	57	0.400 (Aceita H1)	Correlação Moderada Positiva
Média de tópicos criados por estudante	55	0.259 (Aceita H1)	Correlação Fraca Positiva
Média de <i>posts</i> em chats	55	0.333 (Aceita H1)	Correlação Moderada Positiva

De acordo com a Tabela 2 podemos concluir que:

- i. Em relação à taxa de participação de *quizzes* pela turma, a correlação moderada positiva permite inferir que a criação de tópicos de discussão por parte dos tutores refletem em uma melhora significativa na participação dos estudantes nos *quizzes* passados;
- ii. Em relação à média de criação de tópicos por parte da turma, a correlação fraca positiva indica que, à medida que os tutores fazem mais postagens em tópicos de fóruns, a média de criação de tópicos por parte da turma tende levemente a aumentar. Isto pode ser refletido devido à atuação dos tutores nos tópicos, no qual os alunos percebem a participação dos tutores nos tópicos dos fóruns, e criam mais tópicos de discussão, expondo suas dúvidas, com a convicção que seu tópico criado não será ignorado pelo tutor;
- iii. Em relação à média de postagens em *chats* pela turma, a correlação moderada positiva permite inferir que a criação de tópicos de discussão nos fóruns por parte dos tutores tendem a refletir positivamente no aumento moderado de postagens em *chats* pelo alunos.

Tabela 3. Taxa de Visualizações em Fóruns.

Atributos da turma	Número de instâncias sem outliers	Resultado da correlação	Interpretação para a correlação
Taxa de participação em <i>quizz</i>	60	0.318 (Aceita H1)	Correlação Moderada Positiva
Média de cliques dos estudantes	58	0.350 (Aceita H1)	Correlação Moderada Positiva

De acordo com a Tabela 3, podemos concluir que:

- i. Taxa de participação em *quizzes*, com $r = 0.318$, indica correlação moderada positiva;
- ii. Média de cliques dos estudantes, com $r = 0.350$, indica correlação moderada positiva.

A taxa de visualização de fóruns consiste em identificar o acompanhamento dos tutores nos fóruns de um curso. Por exemplo, se um curso tem 20 fóruns e o tutor visualizou apenas 12 fóruns, isso significa que a taxa de visualização em fóruns desse tutor foi de 60%, pois ele acessou apenas 12 dos 20 fóruns possíveis. Sendo assim, esse atributo reflete o acompanhamento dos tutores nos fóruns que são criados na turma.

O resultado da correlação indica que o acompanhamento dos tutores nos fóruns do curso implica em aumento moderado na taxa de submissão dos *quizzes*, e também aumenta moderadamente a média de cliques dos estudantes da turma. Vale salientar que a média dos cliques dos estudantes de uma turma, em um LMS, está diretamente relacionada ao grau de utilização do LMS. Em outras palavras, se a média de cliques de uma turma está bastante alta, significa que os alunos estão acessando bastante o ambiente, enquanto que se a média de cliques estiver muito baixa, significa que o acesso ao LMS acontece de forma casual.

Tabela 4. Taxa de Visualizações em Tópicos dos Fóruns.

Atributos da turma	Número de instâncias sem <i>outliers</i>	Resultado da correlação	Interpretação para a correlação
Taxa de participação em <i>quizz</i>	60	0.412 (Aceita H1)	Correlação Moderada Positiva
Taxa de submissão em atividades	62	0.262 (Aceita H1)	Correlação Fraca Positiva
Média de cliques dos estudantes	58	0.269 (Aceita H1)	Correlação Fraca Positiva

De acordo com a Tabela 4, podemos concluir que:

- i. Taxa de participação em *quizzes*, com $r = 0.412$, indica correlação moderada positiva;
- ii. Taxa de submissão de tarefas, com $r = 0.262$, indica correlação fraca positiva;
- iii. Média de cliques da turma, com $r = 0.269$, indica correlação fraca positiva;

A taxa de visualização de tópicos consiste em identificar o acompanhamento dos tutores nos tópicos que foram criados dentro dos fóruns de um curso. Por exemplo, se um curso tem 10 fóruns e cada fórum tem 10 tópicos, no total, esse curso possui 100 tópicos. Se tutor visualizou apenas 50 tópicos, isso significa que a taxa de visualização de tópicos desse tutor é de 50%, pois ele acessou apenas 50 dos 100 tópicos possíveis. Sendo assim, esse atributo reflete o acompanhamento dos tutores nos tópicos dos fóruns da turma.

O resultado da correlação indica que o acompanhamento dos tutores nos fóruns do curso implica em um moderado aumento na taxa de participação em *quizzes*, além disso, ainda indica um aumento leve na taxa de submissão de atividade e média de cliques dos alunos da turma.

6. Conclusões

O presente trabalho buscou investigar fatores comportamentais dos tutores que pudessem, ou não, estar associados com a participação efetiva dos alunos em turmas na modalidade de ensino a distância. Para a realização da análise, foram obtidos dados de

62 turmas do ensino a distância, com informações comportamentais de 2.227 alunos e 38 tutores a distância. Para a análise dos dados, foi utilizado o coeficiente de correlação de Pearson, com nível de significância de $\alpha = 0,05$ e nível de confiança de 95%. E para eliminação de *outliers* foi utilizada a técnica Z-Score.

Ao todo foram estudados quatro comportamentos dos tutores e dez comportamentos das turmas, gerando uma combinação total de 40 correlações, das quais 29 correlações não passaram no teste de confiança, ou seja, a indicação foi de aceitação da hipótese nula, e 11 correlações passaram no teste de confiança, sendo indicada a aceitação da hipótese alternativa, cujos resultados foram apresentados na Seção 5, bem como uma discussão sobre os resultados obtidos.

Espera-se que esta pesquisa possa proporcionar uma melhor compreensão sobre fatores comportamentais dos tutores que possam, ou não, impactar na participação efetiva dos alunos. Além disso, novas estratégias pedagógicas podem traçadas a partir dos resultados obtidos no que se refere a correlação entre o comportamento dos tutores e turmas do ensino a distância.

Como trabalhos futuros pretende-se coletar mais dados relacionados aos comportamentos dos tutores e das turmas, a fim de identificar novas correlações. Além de realizar uma análise com outros tipos de técnicas, de modo a obter novas informações sobre o processo de ensino e aprendizagem que ocorre nesse tipo de modalidade.

7. Referências

- Gravetter F. J., Wallnau L. B. *Essentials of Statistics for the Behavioral Sciences*. Cengage Learning, 2010.
- Fidalgo-Blanco, Á., Sein-Echaluce, M. L., García-Peñalvo, F. J., & Conde, M. Á. Using Learning Analytics to improve teamwork assessment. *Computers in Human Behavior*, 47, 149-156, 2015.
- Johnson, L., Smith, R., Willis, H., Levine, A., & Haywood, K. The 2011 Horizon Report. The New Media Consortium, (2011).
- Keller, G. *Statistics for Management and Economics*. Cengage Learning, 2011.
- Kumar, D., & Samalia, H. V. Learning Management System Adoption in Universities- Perception about Moodle as a Pedagogical Tool in a Private University in India. *International Journal of Management & Behavioural Sciences (IJMBS)*, 6, 169, 2015.
- Mayer-Schönberger, V., & Cukier, K. *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt, 2013.
- McAfee A., Brynjolfsson E. Big Data: The Management Revolution. *Harvard Business Review*, 1-9, 2012.
- Meissner, G. *Correlation Risk Modeling and Management: An Applied Guide including the Basel III Correlation Framework - With Interactive Models in Excel / VBA*. John Wiley & Sons, 2013.
- Oracle. Disponível em: <https://java.com/en/download/faq/whatis_java.xml> Acesso em: 18 de maio de 2016.

- Rubin, A. *Statistics for Evidence-Based Practice and Evaluation*. Cengage Learning, 2012.
- Schumacker, R. E. *Learning Statistics Using R*. SAGE Publications, 2014.
- Sharma, J.K. *Business Statistics*. Pearson Education India, 2012.
- Siemens, G., Gasevic, D., Haythornthwaite, C., Dawson, S., Shum, S. B., Ferguson, R., & Baker, R. S. J. D. *Open Learning Analytics: an integrated & modularized platform. Proposal to design, implement and evaluate an open platform to integrate heterogeneous learning analytics techniques*, 2011.
- Siemens, G., & Long, P. *Penetrating the Fog: Analytics in Learning and Education. EDUCAUSE review*, 46(5), 30, 2011.
- Silva, R., Ramos, J. L. C., Rodrigues, R., Gomes, A. S., & Fonseca, A.. *Mineração de dados educacionais na análise das interações dos alunos em um Ambiente Virtual de Aprendizagem*. In *Anais do Simpósio Brasileiro de Informática na Educação*, vol. 26, n. 1, p. 1197-1206, 2015.
- Warner, R. M. *Applied Statistics: From Bivariate Through Multivariate Techniques*, SAGE, 2012.
- Weinberg, S. L. and Abramowitz, S. K. *Data Analysis for the Behavioral Sciences Using SPSS*. Cambridge University Press, 2002.
- WEKA. Disponível em <<http://www.cs.waikato.ac.nz/ml/weka/arff.html>> Acesso em: 25 de maio de 2016.