Uma Experiência do Uso de Ferramentas de *Learning Analytics* para Análise de Interações

Italo Dantas de Souza, Thereza Patrícia Pereira Padilha

Departamento de Ciências Exatas – Universidade Federal da Paraíba (UFPB) Campus IV – Rio Tinto – PB – Brazil

{italo.souza, thereza}@dcx.ufpb.br

Abstract. Problems such as retention and evasion have been studied for several years and the Learning Analytics research area has been seen as an alternative to minimize such problems because analyses data such as grades, attendance and interaction with peers, teachers, and educational resources. This paper has a goal to present an experience of the use of two tools using data from a Computer Science course stored in the Management Integrated System of Academic Activities with the purpose to analyze the interactions made by students. It was possible to perceive that the extracted information was important warning for teacher and academic manager.

Resumo. Problemas como retenção e evasão têm sido objeto de estudo a vários anos e a área de pesquisa Learning Analytics tem sido vista como uma alternativa para minimizar tais problemas, pois analisa dados como notas, frequência e interação com colegas, professores e recursos educacionais. Este artigo tem como objetivo apresentar uma experiência do uso de duas ferramentas utilizando dados de uma disciplina do curso de Licenciatura em Ciência da Computação armazenados no Sistema Integrado de Gestão de Atividades Acadêmicas com o intuito de analisar as interações realizadas pelos alunos. Diante dos resultados, pôde-se perceber que as informações extraídas foram importantes sinalizadores para professor e gestor acadêmico.

1. Introdução

A evasão escolar tem sido objeto de estudo há vários anos e, constantemente, pesquisas são realizadas, buscando suas causas e maneiras para diminuição de tal problema. Segundo [Filho et al. 2007], perdas de estudantes que iniciam, mas não terminam seus cursos são desperdícios sociais, acadêmicos e econômicos. Isso se dá pelo fato de que o dinheiro investido não tem retorno, devido à desistência do estudante, além de não haver formação de novos profissionais qualificados. Além da evasão estudantil, existem outros problemas no âmbito educacional, tais como: frustração das expectativas com o curso, dificuldades de aprendizado, altos índices de reprovação, baixa participação dos alunos nos ambientes virtuais de aprendizagem (AVA) e falta de práticas metodológicas e motivadoras [Dias et al. 2010].

Diante desses problemas, o presente artigo tem como objetivo apresentar uma experiência do uso de duas ferramentas de *Learning Analytics* (LA), Gephi e RapidMiner, para análise de interações entre alunos, interações aluno – professor e interações aluno – sistema, armazenados no Sistema Integrado de Gestão de Atividades Acadêmicas (SIGAA) a fim de compreender melhor os seus desempenhos. Os dados utilizados nos experimentos foram de alunos matriculados na disciplina presencial de

DOI: 10.5753/cbie.wcbie.2017.644

Lógica Aplicada à Computação (LAC), semestre 2016.2, do curso Licenciatura em Ciência da Computação, da Universidade Federal da Paraíba.

Este artigo está estruturado da seguinte forma: na seção 2 são apresentados os conceitos sobre *Analytics* e *Learning Analytics*. Na seção 3 são descritas as ferramentas de LA Gephi e RapidMiner Studio usadas nos experimentos. Na seção 4 são apresentados resultados da experiência da aplicação dessas ferramentas com dados colhidos na disciplina de LAC. Na seção 5 são descritas as considerações finais do trabalho e, por fim, são listadas as referências bibliográficas.

2. Learning Analytics

De acordo com a 1ª Conferência Internacional *Learning Analytics and Knowledge* [*LAK* 2011], LA são a medida, coleta, análise e relato dos dados de estudantes e seus contextos com o propósito de entender e otimizar o aprendizado e o ambiente em que este ocorre. Já segundo [Johnson et al. 2012], LA refere-se à interpretação de uma ampla gama de dados produzidos por e recolhidos em nome dos estudantes, a fim de avaliar o progresso acadêmico, prever o desempenho futuro e identificar possíveis problemas. Para [Barneveld et al. 2012], em uma das definições, refere-se a LA como sendo a utilização de dados e modelos para prever o progresso e desempenho do aluno, e a capacidade de agir sobre essa informação.

A área de *LA* tem como objetivo fazer a análise dos dados que são produzidos de uma maneira eficiente, buscando melhoria no sistema de aprendizagem. Segundo [Lockyer et al. 2013], LA oferece um método de obter informações de como os estudantes estão interagindo com os recursos de aprendizagem, ou de como está sendo a interação entre estudantes e seus professores. Informações como quanto tempo o estudante passa para realizar *login* no AVA, tempo que fica *on-line*, tempo que ficou realizando alguma atividade, se está interagindo com colegas através de *chats* e fóruns disponibilizados no AVA, qual mídia (áudio, vídeo e texto) o estudante mais se interessa. Enfim, uma quantidade de dados, muitas vezes inexplorados, que podem ser utilizados pelos professores em busca de uma maneira mais adequada para personalizar o ensino para seus estudantes a fim de fazê-los alcancar o melhor desempenho possível.

2.1 Modelo de Referência de *Learning Analytics*

Vários autores explicam e demonstram o processo de *Learning Analytics*, tais como: [Greller e Drachsler 2012] e [Chatti et al. 2012 apud Moissa et al. 2014]. Para esse artigo, será utilizado como modelo de referência o trabalho de [Chatti et al. 2012 apud Moissa et al. 2014] por ser mais citado, explicativo e documentado. O modelo de referência de LA, proposto por Chatti et al. 2012, pode ser visualizado na Figura 1.

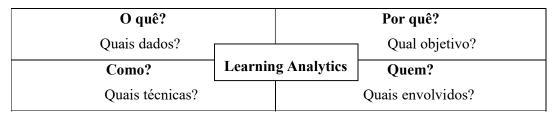


Figura 1. Modelo de Referência de *Learning Analytics*. Fonte: [Chatti et al. 2012 apud Moissa et al. 2014].

As quatro dimensões deste modelo são:

- O que? Esta dimensão refere-se aos tipos de dados coletados. Esses dados podem ser provenientes do ambiente de aprendizagem utilizado, de fontes institucionais, de redes sociais, etc.
- Quem? Esta dimensão refere-se a quem os resultados da análise serão direcionados, podendo ser a alunos, professores, instituições educacionais (administradores), pesquisadores, projetistas de sistema, etc. Cada um dos envolvidos tem suas próprias perspectivas, objetivos e expectativas em relação aos resultados.
- **Por quê?** Esta dimensão está relacionada com os resultados/objetivos da análise realizada, por exemplo, monitoramento e análise, predição e intervenção, tutoria e monitoria, avaliação e *feedback*, adaptação, personalização e recomendação, reflexão.
- Como? Esta dimensão refere-se às diferentes técnicas que podem ser empregadas para detectar padrões contidos nos dados e atingir os objetivos contidos na dimensão *Por quê?*

Após a identificação do que é necessário para construir a análise de dados, é preciso antes entender as etapas do processo de LA, que serão demonstradas a seguir.

2.2 Processo de LA

O processo de LA, segundo [Chatti et al. 2012 apud Cambruzzi 2014], está dividido em três grandes etapas:

- coleta de dados e pré-processamento: responsável pela coleta de dados nos diversos sistemas e ambientes educacionais e pela transformação dos dados em um formato que possa ser usado de entrada para as técnicas de LA.
- análises e ações: com base nos dados obtidos na etapa anterior, deve ser escolhida uma técnica de LA que permita descobrir padrões escondidos nos dados com o objetivo de proporcionar uma experiência de aprendizado mais efetiva. Além das análises e visualizações das informações, essa etapa pode incluir ações de monitoramento, predição, intervenção, adaptação, personalização, recomendação e outras.
- pós-processamento: responsável pela melhoria contínua no processo de LA, pois envolve a inclusão e exclusão de dados e atributos, redefinição de indicadores e identificação de novos métodos de análises.

3. Ferramentas de LA

Existem diversas ferramentas de LA na literatura, tais como: Gephi (Bastian et al. 2009), RapidMiner Studio (Hofmann e Klinkenberg 2014), Tableau (Murray 2013), LightSIDE (Mayfield e Rosé 2013) e D2L Brightspace (Heiman 2016). Essas ferramentas podem, por exemplo, proporcionar: predição de sucesso de um estudante/turma; identificar os estudantes aos quais pode haver intervenção; favorecer a adaptação do conteúdo para se tornar mais adequado a cada aluno (personalização); verificar o nível de interação aluno-aluno, aluno-professor e fornecer dados que possibilitem identificar falhas no processo de ensino aprendizagem. Neste artigo serão apresentadas as ferramentas gratuitas Gephi e RapidMiner Studio. Essas ferramentas

foram escolhidas de acordo com a simplicidade de instalação, por terem documentação mais acessível, e por atenderem além do propósito empresarial, o propósito educacional.

3.1 Gephi

Gephi é uma plataforma de código aberto para a visualização e manipulação de grafos dinâmicos e hierárquicos, incluindo todos os tipos de redes e sistemas complexos. O usuário é livre para modificar o grafo e interagir com a estrutura, posteriormente adicionando filtros para ressaltar os aspectos que deseja e, se necessário, exportar o resultado para SVG, PNG ou PDF [Marquez et al. 2013]. A ferramenta usa grafos para a representação dos dados, onde cada nó (no caso educacional) representa um estudante e cada aresta representa suas relações de interação entre eles, e também a espessura da aresta define o peso das relações. Portanto, relações com menos peso têm arestas mais finas e relações com mais peso são representadas com arestas mais grossas.

Sua principal função é servir como método de análise de dados, elaboração de hipóteses ("Como?", "De que modo?" e o "Por quê? de algo") e descoberta de padrões. Também é largamente utilizado na visualização de redes de relações entre indivíduos e dos conteúdos que (re)produzem tornando-se uma ferramenta para pesquisas de opinião e no campo da cibercultura [Marquez et al. 2013]. O Gephi permite a entrada e a saída de dados nos formatos: GEXT, .GDF, .GraphML (grafos com nós e arestas definidos), .Gephi (grafos com nós, arestas, estrutura e visual definidos) e .CSV.

3.2 RapidMiner Studio

O RapidMiner é uma plataforma de software e fornece um ambiente integrado para a aprendizagem de máquina, mineração de dados, mineração de texto, análise preditiva e análise de negócios [Jalloule et al. 2014]. RapidMiner é usado para aplicações comerciais e industriais, bem como para a pesquisa, a educação, a formação a prototipagem rápida e desenvolvimento de aplicativos, suportando todas as etapas do processo de mineração de dados, incluindo resultados de visualização, validação e otimização [Jalloule et al. 2014].

RapidMiner Studio contém mais de 1500 operações para todas as tarefas de análise de dados profissionais, desde o particionamento de dados, até a análise baseada no mercado. Mas também métodos de mineração de texto, mineração de web, análise de fóruns de discussão na Internet, bem como a análise de séries temporais e previsão, estão disponíveis. RapidMiner Studio permite usar visualizações em gráficos 3-D, matrizes de dispersão e mapas auto organizados. Além de permitir que seus dados sejam totalmente personalizáveis, gráficos exportáveis com suporte para zoom e reescalonamento para impacto visual [RapidMiner Studio Manual 2014]. A entrada de dados pode ser realizada através de banco de dados ou planilhas do Excel, e a saída dos resultados pode ser feita na forma de tabelas, grafos ou árvores de decisão.

4. Estudo de Caso

O estudo de caso foi realizado com o intuito de demonstrar o potencial da área de LA em questões como o monitoramento e análise de interações. A Figura 2 apresenta as quatro dimensões do modelo de referência de LA para o estudo de caso.

O quê?	Por quê?				
	Monitoramento/Análise/Intervenção				
	P1: Como os alunos estão utilizando o SIGAA como suporte ao processo de ensino-aprendizagem?				
SIGAA	P2: O tempo em que o estudante demora em acessar os materiais da disciplina influencia a sua nota?				
	P3: É possível predizer o desempenho do estudante na disciplina através das suas interações?				
	P4: Qual período do dia em que os alunos mais entram no SIGAA?				
Como?	Quem?				
Visualização de dados	Professores				
Estatística	Coordenadores				

Figura 2. Dimensões do Modelo de Referência de LA para o Estudo de Caso.

A dimensão "O quê?" define a fonte dos dados que serão analisados e, neste caso, foi do SIGAA. Na Tabela 1, é possível observar a lista de dados coletados do SIGAA durante o período de 19/01/2017 a 21/04/2017. A dimensão "Por quê" está associada com os objetivos a serem atingidos. Para isso, foram elaboradas quatro perguntas pontuais (P1, P2, P3 e P4), que serão respondidas a seguir. A dimensão "Quem?" define que os resultados deste estudo de caso se destinam a professores e coordenadores. A dimensão "Como?" indica as técnicas empregadas para atingir os objetivos propostos.

Tabela 1. Dados utilizados para o estudo de caso.

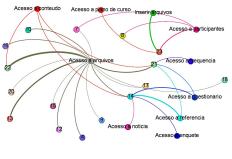
ID (1 a 22 correspondem aos alunos e 23 ao professor)				
Acesso aos arquivos (384 ocorrências)				
Acesso ao SIGAA (365 ocorrências)				
Inserção de arquivos (20 ocorrências)				
Acesso à frequência (13 ocorrências)				
Acesso às notícias (11 ocorrências)				
Acesso aos participantes (9 ocorrências)				
Acesso às notas (9 ocorrências)				
Acesso a conteúdo/plano de curso (8 ocorrências)				
Acesso aos questionários (7 ocorrências)				
Acesso às avaliações (6 ocorrências)				
Acesso à enquete (4 ocorrências)				
Acesso às referências (3 ocorrências)				
Acesso ao fórum (1 ocorrência)				
Frequência				
Nota da unidade I				
Nota da unidade II				
Data				

Para responder as perguntas P1, P2 e P4, foi utilizada a ferramenta Gephi e foi necessário a produção de duas tabelas (node e edge). Na tabela node encontram-se todos os nós com seus IDs (estudantes e funcionalidades do sistema). Na tabela edge encontram-se o nó de saída, o nó de destino, o tipo de ligação entre os nós (direta e indireta) e o peso das ligações entre os nós. No grafo, estudantes e professora serão

representados por nós numéricos (1, 2, 3, ..., 23) e as funcionalidades do sistema serão representadas pelos nós nominais (Acesso a arquivos, acesso a notas, etc.). A seguir são apresentadas as perguntas e as respostas obtidas após a utilização de ferramentas de LA.

P1: Como os alunos estão utilizando o SIGAA como suporte ao processo de ensinoaprendizagem?

O SIGAA apresenta diversas funcionalidades (visualizar arquivos, fazer o *download* de arquivos, ter acesso a fóruns de discussão, visualizar notas, frequência, realizar enquetes, entre outros), todas com o intuito de fornecer um ambiente de ensino-aprendizagem diversificado. O estudo de caso foi realizado com dados entre os dias 19/01/17 a 21/04/17, e, para cada mês (janeiro, fevereiro, março e abril), foi produzido um grafo, demonstrando às interações dos estudantes e do professor para com o SIGAA, conforme podem ser vistos nas Figuras 3, 4, 5 e 6.

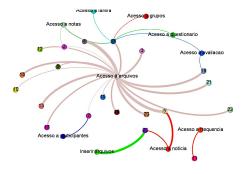


Acesso a graphics

Acesso a grap

Figura 3. Interação Janeiro

Figura 4. Interação Fevereiro



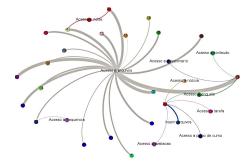


Figura 5. Interação Março

Figura 6. Interação Abril

Foi possível notar que, ao longo dos meses, o número de interações dos estudantes com os recursos disponíveis no SIGAA aumenta. No mês de janeiro e fevereiro, alguns alunos acessaram o plano de curso da disciplina e a referência utilizada pela professora. Já nos meses de março e abril os alunos acessaram os arquivos disponibilizados no sistema, avaliações e notas, o que pode ser um reflexo do período de provas da disciplina.

P2: O tempo em que o estudante demora em acessar os materiais da disciplina influencia a sua nota?

Para tentar responder a essa pergunta, foram utilizados os dados de interação com os materiais inseridos pela professora da disciplina, apenas entre o período que corresponde à unidade II da disciplina (10/03/17 a 21/04/17), por se tratar de uma

análise mais detalhada, o que demandaria uma grande quantidade de tempo se fosse realizada também para a unidade I. A diferença da espessura das arestas nesse caso significa o tempo que foi decorrido para acessar o arquivo, sendo arestas com menor espessura para os estudantes que acessaram mais rapidamente e arestas com maior espessura para maiores tempos até o acesso.

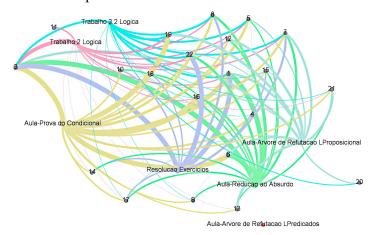


Figura 7. Acompanhamento dos materiais.

Foi possível identificar que o grafo apresenta diferença em relação ao tempo de acesso dos arquivos por parte dos alunos, onde é notável que os alunos alternem entre acessar rapidamente alguns arquivos, já em outros nem tanto. Vale destacar que o nó "Aula-Árvore de Refutação LPredicados" não foi acessado nenhuma vez durante o período em que os dados foram coletados pois, apesar de ter sido adicionado no período da unidade II, é material associado à unidade III e que o nó "2" não é apresentado no grafo, pois não acessou nenhum dos arquivos deste período. Após a produção do grafo foi realizado um estudo para detectar se há uma relação com as notas dos alunos na disciplina. Na Figura 8 é apresentada uma relação com a identificação e as notas dos alunos.

ID	Nota unidade II	ID	Nota unidade II
1	2	12	10
2	9,5	13	4
3	0	14	4,5
4	8,8	15	6,8
5	5	16	8,5
6	3	17	6,8
7	5,5	18	9,5
8	2	19	2
9	8,9	20	6,3
10	6,8	21	9,5
11	0	22	2,5

Figura 8. Notas unidade II

Analisando a tabela de notas e com o grafo gerado, foi possível identificar que dos 07 alunos que tiveram sucesso na disciplina (tiraram nota maior ou igual a 7,0) na unidade II, 5 (71,42%) acessaram os materiais disponíveis em até 5 dias, 1 (14,59%) acessou os materiais em até 10 dias e 1 (14,59%) não acessou o sistema acadêmico no período correspondente a unidade II (aluno 2). Neste último caso, possivelmente, o aluno obteve os materiais da disciplina por outro meio.

P3: É possível predizer o desempenho do estudante na disciplina através das suas interações?

Para responder a essa pergunta, foi utilizada a ferramenta RapidMiner Studio em que é possível realizar a predição de sucesso dos estudantes. Para realizar essa predição foram utilizados os dados de acesso dos alunos no SIGAA, número de interação, frequência e notas da unidade I. A ideia foi realizar a predição de notas da unidade II, mesmo que essas notas já estivessem disponíveis, com o intuito de verificar se a ferramenta seria capaz de realizar a predição.

Os dados foram manuseados para serem comportados pela ferramenta. Foram produzidas duas tabelas no Excel, a primeira e a tabela de controle, com os atributos: "ID"; "acesso ao sigaa 1°" (correspondente a janeiro); "interação 1° mês"; "faltas 1° mês"; acesso ao sigaa 2°" (correspondente a fevereiro); "interação 2° mês"; "faltas 2° mês"; "total de acessos" e "notas da 1° unidade". Já a segunda conta com os mesmos atributos da primeira, com a diferença que o 1° e 2° mês correspondem a março e abril respectivamente, além do acréscimo da coluna "notas da 2° unidade maior ou igual a 7" que é a coluna na qual será realizada a predição. Uma amostra dos resultados é apresentada na Figura 9.

ID	prediction(Notas 2° unidade >= 7)	confidence(Sim)	confidence(Não)	Acessos ao sigaa 1°	Interação 1º mês	Faltas 1º mes	Acessos ao sigaa 2°	Interação 2º mês	Faltas 2º mes	Total de Acessos	Notas 1º Unidade
1	Sim	1	0	0	0	2	2	6	0	2	7.600
2	Sim	1	0	4	1	0	1	1	0	5	7.900
3	Não	0	1	4	7	0	3	9	0	7	0
4	Sim	1	0	9	12	0	7	13	0	16	8.200
5	Não	0	1	16	16	2	9	11	0	25	6.800
6	Não	0	1	0	0	0	1	3	0	1	5.500
7	Sim	1	0	5	2	0	4	7	0	9	9
8	Sim	1	0	0	0	2	4	8	0	4	7.800
9	Sim	1	0	3	4	0	5	9	0	8	8.500
10	Sim	1	0	2	2	0	7	15	0	9	8.800

Figura 9. Amostra da predição de notas da unidade II.

De acordo com a tabela disponibilizada pela ferramenta após a realização da análise dos dados, foi possível ver que a predição foi realizada de forma correta e para este caso houve (100%) de acerto na predição das notas dos 22 alunos.

P4: Qual período do dia em que os alunos mais acessam o SIGAA?

É importante para o professor ter conhecimento do melhor período do dia para disponibilizar os materiais para os seus alunos, ou seja, horários que os alunos possam ser mais facilmente observar os materiais postados. Para que fosse possível responder à pergunta foram elaborados quatro (04) grafos, com os períodos do dia manhã (das 6:00 até 11:59), tarde (das 12:00 até 17:59), noite (das 18:00 até 23:59) e madrugada (das 00:00 até 05:59) em que os alunos acessaram o sistema acadêmico para analisar qual deles, de fato, é o período do dia mais acessado. Para a produção dos grafos foram necessários os dados de *log* de acesso com os horários em que cada estudante logou no sistema durante o período em que foi realizado o estudo de caso.

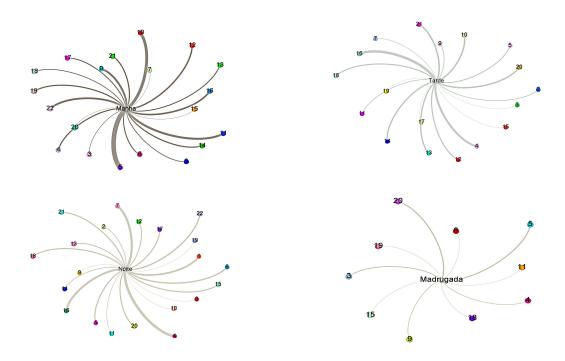


Figura 10. Acesso dos alunos por turno.

Foi detectado que os estudantes costumam entrar em horários alternados durante o dia, tendo deles que entraram em todos os períodos do dia. Dos 22 alunos da disciplina, 20 (90,91%) costumam entrar no sistema durante o período da manhã (aluno 1 e 8, não entraram nenhuma vez), 18 (81,82%) entram à tarde (alunos 1, 6, 8 e 22 não entraram nenhuma vez), 21 (95,45%) entram pela noite (aluno 6 não entrou nenhuma vez), e 10 (45,45%) entram de madrugada (alunos 1, 2, 6, 7, 10, 12, 13, 14, 16, 17, 21, 22 não entraram nenhuma vez). Portanto, mostrou-se que os alunos acessam mais no sistema durante à noite.

5. Considerações Finais

Através da revisão da literatura, foi possível identificar que LA ainda é uma área pouco estudada, mas que vem crescendo em termos de pesquisas e de estudos relacionados. Foi possível ainda verificar que é uma área a ser explorada e que pode trazer beneficios para todas as partes envolvidas na gestão acadêmica e no processo de ensino-aprendizagem. Os sistemas acadêmicos, em geral, possuem muitas informações para serem exploradas, embora as vezes haja necessidade de refinar os dados armazenados a fim de que possam ser utilizados por ferramentas de LA.

Este artigo teve como objetivo demonstrar o potencial da área de LA através de um estudo de caso contendo dados obtidos da turma de LAC com o intuito de analisar as interações entre os envolvidos no processo de ensino-aprendizagem. Para isso, foram utilizadas as ferramentas Gephi e RapidMiner Studio. A partir dos resultados alcançados, foi possível constatar horários mais e menos acessados pelos alunos, os materiais mais e menos acessados, predizer notas e acompanhar as diversas interações dos alunos com o sistema acadêmico. Diante dessas informações, tanto o professor quanto o gestor acadêmico podem-se beneficiar do seu uso para melhorar o processo de ensino-aprendizagem na disciplina.

Referências Bibliográficas

- Barneveld, A. V. B., Arnold, K. E. and Campbell, J. P. Analytics in Higher Education: Establishing a Common Language. EDUCAUSE Library. Jan. 2012.
- Bastian, M., Heymann, S. and Jacomy, M. Gephi: Na Open Source Software for Exploring and Manipulating Networks. Paris, França, 2009.
- Cambruzzi, W. L. GVWISE: Uma aplicação de learning analytics para a redução da evasão na educação a distância. São Leopoldo, RS, 2014.
- Chatti, M. A., Dyckhoff, A. L., Schroeder, U. and Thus, H. A reference Model for Learning Analytics. Int. J. Technol. Enhanc. Learn., Inderscience Publishers, Geneva, SWITZERLAND, v. 4, n. 5/6, p. 318-331, Jan. 2012.
- Dias, E. C. M., Theóphilo, C. R. and Lopes, M. A. S. Evasão no ensino superior: estudo dos fatores causadores da evasão no curso de ciências contábeis da Universidade Estadual de Mones Claros Unimontes MG. São Paulo SP. 2010.
- Filho, R. L. L. S., Motejunas, P. R., Hipólito, O. and Lobo, M. B. C. M. A evasão no ensino superior brasileiro. Cadernos de Pesquisa, v. 37, n. 132, p. 641-659, 2007.
- Heiman, M. M. QR4U: Using Gamification, D2L Brightspace, and QR Codes to Influence Completion Rates in Fall English Classes. Minnesota, EUA, 2016.
- Hofmann, M. and Klinkenberg, R. RapidMiner Data mining Use Cases and Business Analytics Applications. Londres, Inglaterra, 2014.
- LAK 1st International Conference on Learning Analytics and Knowledge, Banff, Alberta, 27 fev-01 mar, 2011.
- Lockyer, L., Heathcote, E. and Dawson, S. Informing pedagogical action: Aligning learning analytics with learning design. American Behavioral Scientist, SAGE, 2013.
- Jalloule, J., Sallé, J. and Bittencourt, R. RapidMiner Aprenda a Usar. Disponível em: https://prezi.com/-yo8qjamdbbq/rapidminer-aprenda-a-usar/.
- Johnson, L., Adams, S. and Cummins, M. The NMC Horizon Report: 2012 Higher Education Edition. 2012.
- Marquez, A. C., Gonçalves, B. B., Medeiros, J. M. R. and Reis, N. A. Gephi: um software open source de manipulação e visualização de grafos. Labic, 2013.
- Mayfield, E. and Rosé, C. P. LightSIDE Open Source Machine Learning for Text. 2013.
- Moissa, B., Gasparini I. and Kemczinski A. Learning Analytics: um mapeamento sistemático. TISE 2014, Fortaleza CE, Dez. 2014.
- Murray, D. G. Tableau Your Data!: Fast and Easy Visual Analysis with Tableau Software. Indianapolis, EUA, 2013.
- RapidMiner Studio. RapidMiner Studio Manual. Boston, 2014.