

***Learning Analytics* para avaliar o desempenho discente em uma disciplina de cursos a distância**

Learning Analytics to assess student performance in a distance learning discipline

Enir da Silva Fonseca

Universidade Cruzeiro do Sul / Programa de Doutorado em Ensino de Ciências e Matemática
Universidade de Ribeirão Preto – Campus Guarujá / Professor Tecnologias
Centro Universitário Lusíada / Professor / Administração e Tecnologias
enir.fonseca@gmail.com

Carlos Fernando de Araújo Jr.

Universidade Cruzeiro do Sul / Professor Pesquisador do Programa de Mestrado e
Doutorado em Ensino de Ciências e Matemática
Universidade Cruzeiro do Sul / Universidade Cidade de São Paulo / Universidade de Franca
carlos.araujo@cruzeirosul.edu.br

Resumo

O *Learning Analytics* é visto por educadores como um caminho, uma solução para personalizar a jornada de cada estudante, identificando o seu percurso, as suas dificuldades e os seus hábitos. A análise da aprendizagem é realizada a partir dos dados gerados durante o processo de ensino e aprendizagem, e que a cada dia aumenta em proporções e complexidade superior a capacidade de análise humana, gerando um grande volume de dados que são conhecidos como *Big Data*. Para consecução desta pesquisa, que visa analisar o desempenho discente em cursos ofertados na modalidade a distância, aplicamos uma pesquisa quantitativa nos dados de 1809 alunos selecionados em um ambiente virtual de aprendizagem configurado no *Blackboard*, distribuídos em 5 turmas da disciplina de Cálculo Diferencial e Integral II, ofertadas no segundo semestre do ano de 2018. Com a coleta de dados, extraiu-se as informações brutas a partir dos relatórios de acessos e de notas, de toda a atividade do usuário nas áreas de conteúdo e do resumo geral da atividade do usuário. Após a mineração dos dados relevantes ao estudo, implementamos um algoritmo para construção das tabelas e gráficos, como o Boxplot e Histograma. A análise demonstrou que com 50 acessos, o discente obtém condições mínimas para absorver o conteúdo e alcançar média no intervalo de um a quatro pontos, no entanto, para os alunos com acessos superiores a 200 vezes, os resultados indicam maior probabilidade de manter-se no intervalo de três, tendendo ao valor máximo que é de quatro pontos. Assim confirmamos que a maior interatividade, proporciona ao discente um constante contato com conteúdo e conseqüentemente melhor avaliação.

Palavras-chave: Análise de Dados. Educação à distância. Mineração de Dados Educacionais. Tecnologia Educacional. Tomada de Decisão.

Abstract

The Learning Analytics is seen by educators as a way, a solution to customize each student's journey, identifying their path, their difficulties, and their habits. The analysis of the learning is carried out from the data generated during the process of teaching and learning, and that every day increases in proportions and complexity superior to the capacity of human analysis, generating a large volume of data that are known as Big Data. To achieve this research, which aims to analyze student performance in courses offered in the distance modality, we apply a quantitative research in the data of 1809 students selected in a virtual learning environment configured in Blackboard, distributed in 5 classes of the discipline of Differential and Integral Calculus II, offered in the second half of 2018. With the data collection, the raw information was extracted from the access and note reports, all the user activity in the content areas and the general summary of the activity of the user. After mining the data relevant to the study, we implemented an algorithm for the construction of tables and graphs, such as Boxplot and Histogram. The analysis showed that with 50 accesses, the student obtains minimum conditions to absorb the content and reach average in the range of one to four points, however, for students with accesses greater than 200 times, the results indicate a greater probability of remaining in the interval of three, tending to the maximum value that is four points. Thus we confirm that the greater interactivity, gives the student a constant contact with content and consequently better evaluation.

Keywords: Data analysis. Distance education. Educational Data Mining. Educational technology. Decision Making.

Introdução

O ser humano durante sua existência, mesmo que inconscientemente, sempre avaliou e reavaliou suas atividades, buscando assim evitar erros óbvios na tomada de decisão, para que elas ocorram de forma racional otimizando o seu tempo, princípio que se aplica as rotinas profissionais. Marras (2016), define que todo processo de avaliação humana apresenta falhas, e estas estão mais diretamente ligadas àquele que avalia do que ao instrumento de avaliação. Com a definição de Marras (2016), observa-se que mesmo uma empresa utilizando métodos e instrumentos adequados e adaptados à sua forma de gestão, é refém da falha humana, que é um fator desafiador na otimização dos processos. E em busca de alternativas para correções pontuais, novos mecanismos são criados para facilitar e ordenar o processo avaliativo, como o *Learning Analytics* (LA) ou análise de aprendizagem, que consiste em um sistema de avaliações que visa o aumento da qualidade do ensino à distância, coletando, analisando e distribuindo os dados de cada aluno.

A LA é vista por educadores como um caminho, uma solução para personalizar a jornada de cada estudante, identificando individualmente o seu percurso, as suas dificuldades e os seus hábitos. E associadas ao uso das modernas tecnologias, pode transformar dados educacionais em informações que possibilitem a melhor tomada de decisão, melhorando o processo de aprendizagem. Chatti et. al. (2012) defini que a LA é a

medição, coleção, análise e reportagem de dados sobre estudantes e seus contextos, para propósitos de entendimento e otimização da aprendizagem e dos ambientes em que ela ocorre. Para Dias (2017), a técnica relacionada à *big data* no âmbito educacional é denominada *Learning Analytics*, que permite avaliar o aluno de forma mais assertiva com base nas respostas inseridas na plataforma de ensino e os registros de sua vida acadêmica, pode ser transformado em análises possíveis de serem cruzadas e obter resultados.

A análise da aprendizagem é realizada a partir dos dados gerados durante o processo de ensino e aprendizagem, que a cada dia aumenta em proporções e complexidade superior a capacidade de análise humana, gerando um grande volume de dados que são conhecidos como *Big Data*, e representado por 5 V's, o Volume, a Velocidade, a Variedade, a Veracidade e o Valor das informações geradas. Campos (2014) afirma que a expressão *Big Data*, pode ser traduzida ao português como “Megadados” ou “Grande Quantidade de Informação”, que conquistou espaço nos debates sobre educação, para descrever um novo aspecto que não pode ser ignorado pelas instituições de formação que queiram crescer ou simplesmente sobreviver nas próximas décadas do século 21.

Associado ao *Big Data*, o uso das Tecnologias Digitais de Informação e Comunicação (TDICs) aumentou em escala. Kenski (2012) afirma que é possível processar qualquer informação gerada, o que tem provocado mudanças radicais na vida das pessoas. No ensino, estes dados são produzidos quando os alunos assistem videoaulas, respondem exercícios, realizam atividades online, entre outros. Com o uso das tecnologias associadas ao ensino, a interação humano-máquina-processo passa a ocorrer em alta velocidade, armazenando os resultados aos Banco de Dados (BD's) associados.

A partir dos dados armazenados, surge a necessidade de uma análise pontual para identificar e solucionar os problemas ocorridos durante o processo de ensino e aprendizagem, que se inicia com a coleta dos dados, quando é aplicadas técnicas mineração para identificar as relações descobertas e produzir informações relevantes que podem ser utilizadas em previsões futuras.

Conforme Silva Junior e Oliveira (2016), a mineração de dados educacionais (EDM, do inglês, *Educational Data Mining*), trata da aplicação de técnicas de mineração de dados para resolver problemas da área educacional. Com a interpretação desses dados, pode-se traçar o perfil do comportamento discente, de forma que aos gestores da instituição planeje quais ações corretivas devem proporcionar um melhor processo de aprendizagem, e consequentemente aumentar o desempenho e a redução da evasão acadêmica. Com a aplicação da EDM, é possível desenvolver uma alternativa educacional customizada, privilegiando os pontos fortes como, os conteúdos que geram o maior número de interações, e identificar as áreas de maior fragilidade, reforçando assim o ensino para que o aluno alcance uma formação mais uniforme e completa.

Para analisar o comportamento discente, optamos em utilizar o RStudio, que é um *software* livre de programação e desenvolvimento integrado ao R. O *software* R, é um ambiente computacional especializado na manipulação, análise e visualização gráfica de dados e cálculos estatísticos. De acordo com Machado e Becher (2016) e Souza, Peternelli e Mello (2014) o *software* R foi criado no departamento de Estatística da Universidade de Auckland na Nova Zelândia por Ross Ihaka e Robert Gentleman, e posteriormente

desenvolvido em esforço colaborativo de pessoas em vários locais do mundo. O nome R provém das iniciais dos criadores (Ross e Robert) e de um jogo figurado com a linguagem S (*Bell Laboratories*, antiga AT&T).

Learning Analytics e o processo de aprendizagem

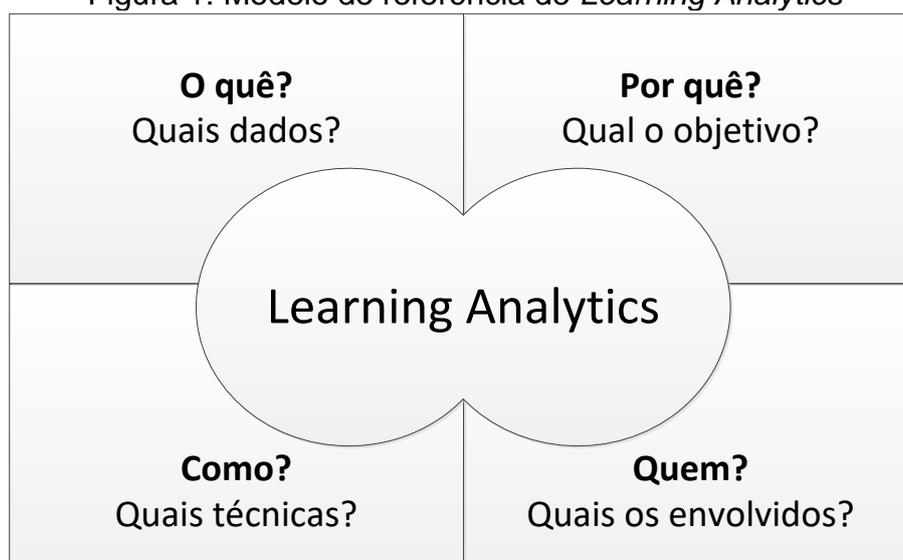
Moissa (2014), define que o processo de LA busca coletar, medir, analisar e relatar os dados e seus contextos com objetivo de otimizar o aprendizado e o ambiente em que este ocorre. Cooper (2012), afirma que o *Learning Analytics* é baseado no conceito de *Analytics* e interpretado como o processo de desenvolvimento de *insights* acionáveis por meio da definição de problemas, e da aplicação de modelos estatísticos e análise em relação a dados futuros existentes e / ou simulados. A partir do ponto de vista de Moissa (2014) e Cooper (2012), e sabendo que os dispositivos e recursos tecnológicos renovam-se constantemente, proporcionando ao mercado alternativas mais eficazes e que são incorporadas ao processo, o modelo educacional passa por uma frequente necessidade de reavaliar e adequar a ementa dos cursos a esta nova realidade, e que na Educação a Distância (EaD) a periodicidade destes ajustes é ainda mais importante, pois toda metodologia depende da tecnologia para o seu melhor funcionamento.

A análise de aprendizagem é a medição, coleta, análise e relato de dados sobre os alunos e seus contextos, com o objetivo de compreender e otimizar o aprendizado e os ambientes em que ele ocorre. Segundo Johnson et al. (2011), LA referem-se à interpretação de um grande volume de dados produzidos pelos alunos, a fim de avaliar o seu progresso acadêmico, e assim prever o desempenho e detectar possíveis problemas de aprendizagem. E para aplicar o LA é necessário definir quais os dados serão utilizados na análise, e assim identificar os alvos ou os interessados da análise (Alunos, Professores, Tutores), avaliar os objetivos dos tipos de cada análise (Monitoramento, Análise, *Feedback*, Intervenção) e a definição das técnicas que farão parte da análise dos dados (Técnicas Estatísticas, Mineração de Dados, Visualização de dados).

Nogueira et al. (2018) afirma que a LA tem capacidade de gerar informações importantes *just in time* para ser usada nas IES de maneira continuada, a fim de melhorar seu processo ensino-aprendizagem. O modelo de referência de LA trabalhado por Chatti et. al. 2012, é baseado em quatro dimensões citadas anteriormente, identificando os desafios e oportunidades apresentado na Figura 1, onde procura-se identificar: O que? Que tipo de dados o sistema coleta, gerencia e usa para a análise; Quem? Quem é alvo da análise; Por quê? Por que o sistema analisa os dados coletados; Como? Como o sistema realiza a análise dos dados coletados.

O modelo ilustrado na Figura 1, busca identificar os dados, os atores, os objetivos, as técnicas e as intervenções envolvidas no processo de LA. Os dados estão relacionados ao comportamento e as características dos alunos, já os objetivos definem o que será feito, auxiliando na seleção de dados, na escolha da melhor técnica, possibilitando então monitoramento, análise, predição, intervenção e avaliação pela tutoria envolvida, retornado ao discente um modelo adaptado e personalizado a sua necessidade com as recomendações e reflexões indicando os caminhos para o melhor aproveitamento.

Figura 1: Modelo de referência de *Learning Analytics*



Fonte: Os autores
Adaptado de: Chatti et. al. 2012.

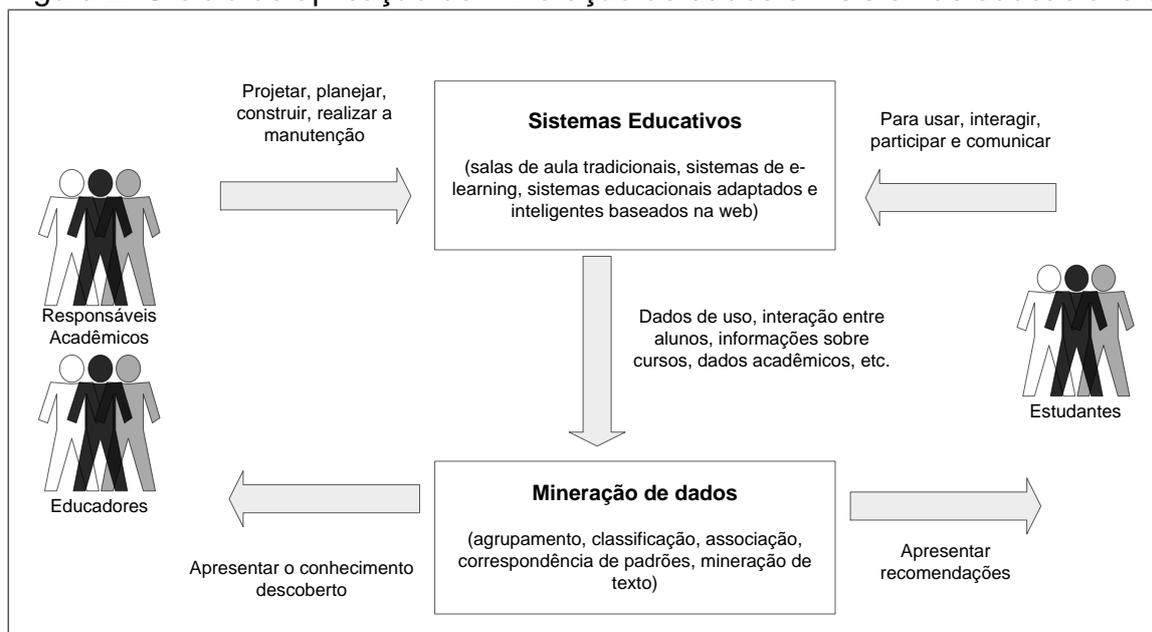
Com o uso do LA, é possível uma economia de tempo e recursos na obtenção de dados para o planejamento de um novo curso ou disciplina na EaD, realizando parte do estudo baseado em dados gerados por estudantes em curso e no histórico dos já formados. Silva et. al. (2016) afirma que as técnicas de LA visam avaliar os dados gerados por cursos, como participação em ambientes virtuais de aprendizagem, frequência e notas para identificar situações como o andamento da aprendizagem e comportamento dos alunos.

Mineração de dados e RStudio

A EDM descreve um campo de pesquisa relacionado à aplicação de mineração de dados, aprendizado de máquina e estatística a partir das informações geradas nos ambientes educacionais, como sistemas de tutoria inteligentes. Fonseca e Araújo Jr. (2018) afirmam que a mineração de dados, é uma área de pesquisa multidisciplinar, incluindo tecnologia de bancos de dados, inteligência artificial, aprendizado de máquina, redes neurais, estatística, reconhecimento de padrões, sistemas baseados em conhecimento, recuperação da informação, computação de alto desempenho e visualização de dados.

Com a EDM é possível buscar, desenvolver e aperfeiçoar métodos para explorar os dados, que geralmente têm múltiplos níveis de hierarquia significativa, e a fim de descobrir novos *insights* sobre como as pessoas aprendem no contexto de tais configurações. Ao fazer isso, a EDM contribuiu para teorias de aprendizagem investigadas por pesquisadores em psicologia educacional e ciências da aprendizagem. O campo está intimamente ligado ao da análise de aprendizagem. A Figura 2, apresenta o ciclo de aplicação de mineração de dados em sistemas educacionais.

Figura 2: O ciclo de aplicação de mineração de dados em sistemas educacionais.



Fonte: Os autores

Adaptado de: Martínez e Gómez. 2014. *Apud* Romero e Ventura. 2007.

Após a análise sobre a interação entre alunos e conteúdos no AVA, o conhecimento obtido sobre a quantidade de acessos e as notas alcançadas no processo de aprendizagem, é possível apresentar recomendações aos estudantes e indicar aos educadores alternativas e caminhos para correção pontual que melhore o modelo. Castro e Ferrari (2016) definem que.

O termo mineração de dados (MD) foi cunhado como alusão ao processo de mineração em uma mina, para extração de minerais preciosos, uma vez que se explora uma base de dados usando algoritmos adequados para obter conhecimento. Os dados são símbolos ou signos não estruturados, sem significado, como valores em uma tabela (Castro e Ferrari. 2016. p. 4).

Com a afirmação de Castro e Ferrari (2016), verifica-se que para mineração de dados, é necessário conhecimentos de banco de dados, estatística, aprendizagem de máquina, reconhecimento de padrões, linguagem de programação, entre outras, já que é necessário explorar grandes quantidades de dados à procura de padrões consistentes, afim de detectar relacionamentos sistemáticos entre as variáveis.

O RStudio é um *interface* amigável para o *software* R, e de acordo com a Fundação R (2019), o *software* R é um ambiente computacional livre para estatística e gráficos, é uma linguagem de programação que vem progressivamente se especializando em manipulação, análise e visualização gráfica de dados. Compila e executa os algoritmos desenvolvidos em uma ampla variedade de plataformas UNIX, Windows e MacOS.

O RStudio (2019) é um conjunto de ferramentas integradas projetadas para uso do *software* R, que inclui um console, editor de realce de sintaxe que suporta execução direta de código e uma variedade de ferramentas robustas para plotagem, visualização de histórico, depuração e gerenciamento de seu espaço de trabalho.

Materiais e Métodos

Para o estudo de caso aplicamos uma pesquisa quantitativa, que de acordo com Perovano (2016), a pesquisa quantitativa permite apontar a relação entre as variáveis contidas na hipótese de pesquisa, e já Fonseca (2002) define que:

A pesquisa quantitativa se centra na objetividade. Influenciada pelo positivismo, considera que a realidade só pode ser compreendida com base na análise de dados brutos, recolhidos com o auxílio de instrumentos padronizados e neutros. A pesquisa quantitativa recorre à linguagem matemática para descrever as causas de um fenômeno, as relações entre variáveis, etc. (Fonseca. 2002, p. 20).

Com as afirmações de Perovano (2016) e Fonseca (2002), a pesquisa quantitativa é centrada na objetividade e os resultados são considerados como um retrato real de toda a população alvo da pesquisa. E para a análise, foram selecionados dados de cinco turmas da disciplina de Cálculo Diferencial e Integral II, ofertada no ano de 2018 na modalidade a distância, conforme Tabela 1.

Tabela 1: Cursos/alunos por turma – Cálculo Diferencial e Integral II

01_09_2018 - 80		03_10_2018 - 80	
Análise e Desenvolvimento de Sistemas	1	Análise e Desenvolvimento de Sistemas	1
Ciências Contábeis	1	Engenharia Ambiental	21
Engenharia Ambiental	354	Engenharia de Produção	460
Engenharia de Produção	1		
Física	66		
Geografia	1		
Gestão Ambiental	1		
Química	105		
02_09_2018 - 80		04_10_2018 - 80	
Arquitetura e Urbanismo	1	Engenharia Ambiental	3
Engenharia Ambiental	24	Engenharia de Produção	25
Engenharia Civil	183	Engenharia Elétrica	1
Engenharia de Computação	43	Engenharia Mecânica	2
Engenharia de Produção	1		
Engenharia Elétrica	129		
Engenharia Mecânica	99		
		05_09_2018 - 80	
		Ciências Biológicas	1
		Engenharia Ambiental	224
		Física	19
		Gestão Ambiental	2
		Química	40

Fonte: Os autores

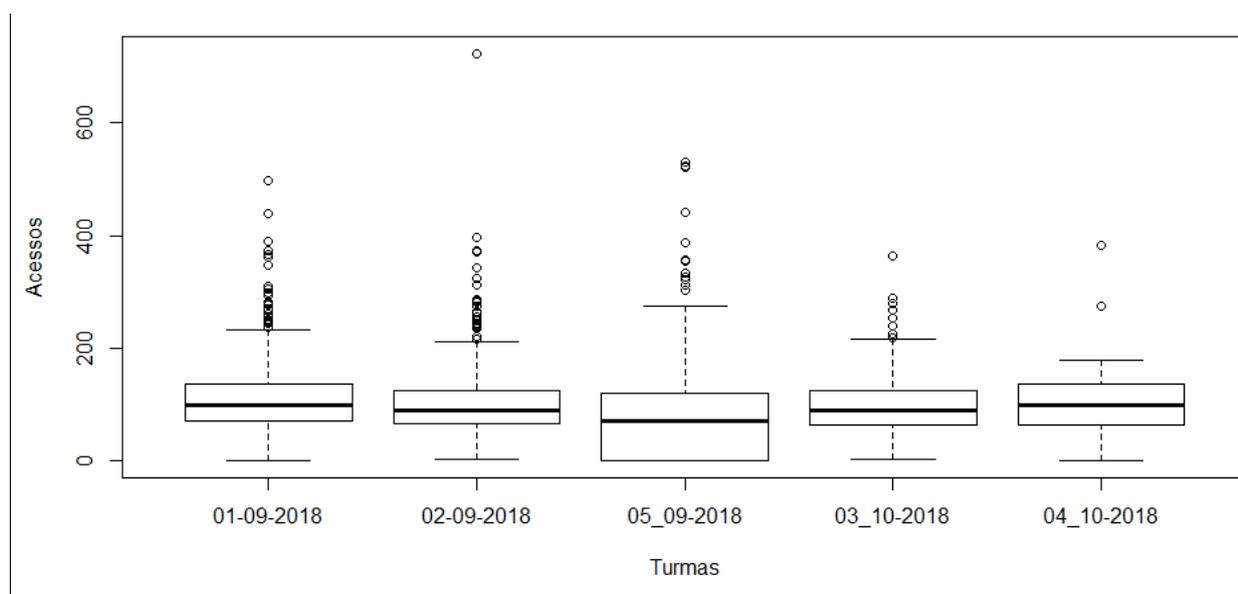
O intervalo de dados válidos utilizados é de 1809 alunos distribuídos em 530 para a turma 01_09_2018, 480 para a turma 02_09_2018, 482 para turma 03_10_2018, 31 para a turma 04_10_2018 e 286 para a turma 05_09_2018, com a aulas ofertadas pelo *Blackboard*, que é um Sistema de Gestão da Aprendizagem, do inglês *Learning Management System* (LMS).

A metodologia deste trabalho é baseada no processo de *Learning Analytics*, e seguiu os seguintes passos: capturar dados, reportar dados, predizer, adaptar e personalizar. A obtenção dos dados foi realizada a partir dos relatórios gerados no Ambiente Virtual de Aprendizagem (AVA), extraindo-se os relatórios de acessos contendo toda a atividade do usuário nas áreas de conteúdo, resumo geral da atividade do usuário, e do relatório de notas, disponibilizados no formato .CSV, que são arquivos de texto com valores separados por vírgulas ou ponto e vírgula. A Figura 3 exemplifica no bloco de notas o formato dos dados brutos.

docente ao discente, que é possível com a identificação do maior ou o menor interesse do aluno por determinadas áreas, e desta forma empregar tecnologias que possam gerar automaticamente *links* de estudos, ou permitir ao docente identificar rapidamente em quais pontos deve atuar para corrigir o aprendizado caso a caso. Entender os dados, possibilita consequentemente uma melhora no desempenho dos alunos com uma maior influência da instituição no resultado do processo de aprendizado.

O principal foco do trabalho é de se identificar a evolução discente em relação aos seus acessos ao material disponibilizado no AVA, e como estas ações podem contribuir para a melhoria do seu aprendizado. A partir da extração e análise dos dados, observou-se um desempenho semelhante entre quatro turmas em relação aos acessos, sendo que a quinta turma apresentou uma variabilidade maior, indicando que uma quantidade de 50 acessos é suficiente para obter-se o mínimo necessário do aprendizado visando a promoção. A Figura 5 demonstra um número expressivo de *outliers* em todas as turmas, o que a princípio pode ser interpretado como uma ocorrência inconsistente em relação ao afastamento às demais anotações da série.

Figura 5: Acesso das turmas

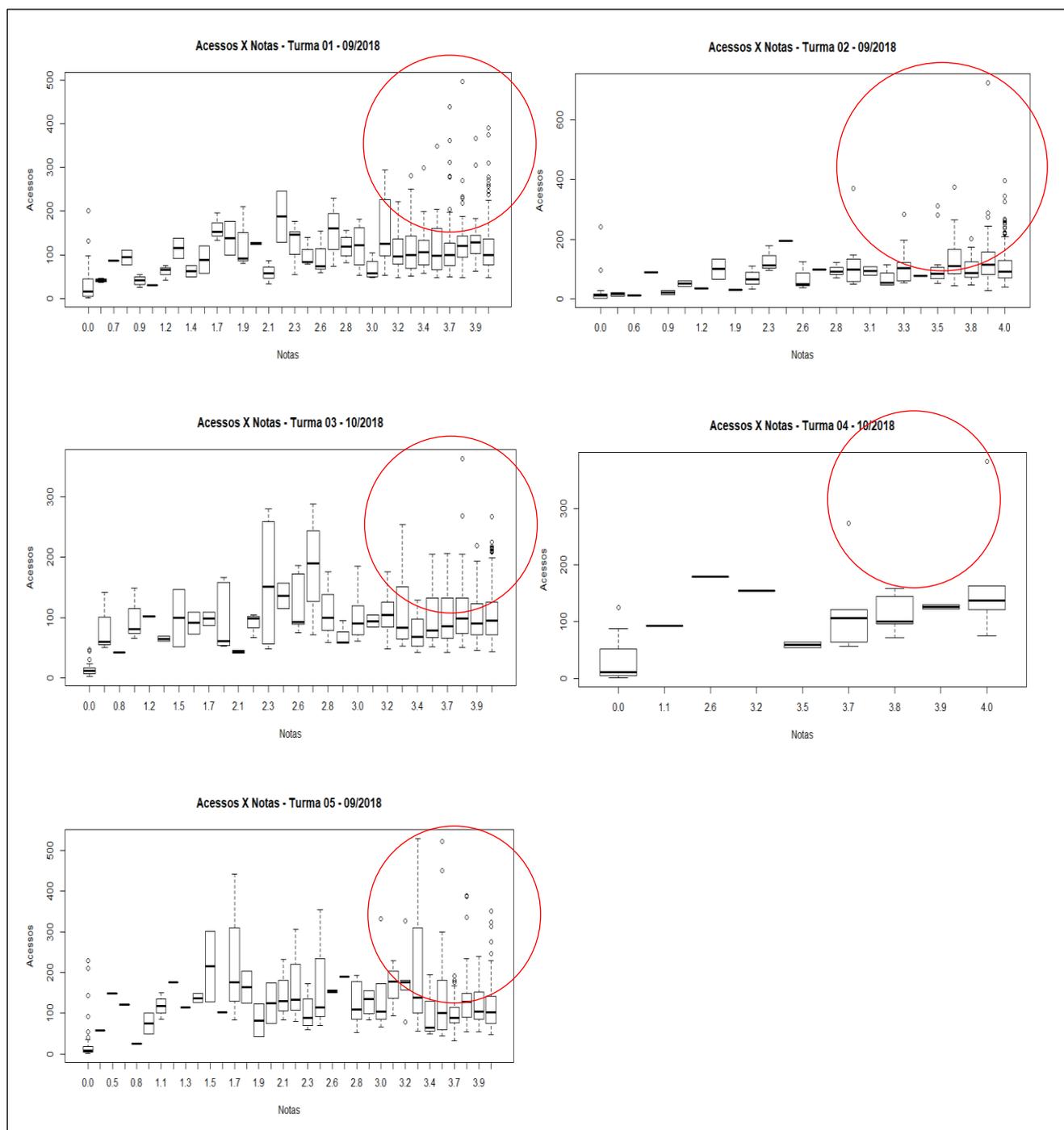


Fonte: Os autores

Dentre as cinco turmas estudadas, apenas o grupo 04_10_2018 apresentou dois pontos fora do padrão. E para melhor compreensão destes dados discrepantes, exploramos cada uma das turmas, agrupando as notas obtidas pelos aluno no AVA, que correspondem ao intervalo de um a quatro pontos.

Como os *outliers* são dados que se diferenciam de todos os outros, e que fogem da normalidade, conhecidos como pontos fora da curva, neste estudo identificou-se que o comportamento demonstra a importância na constância dos acessos para melhor aprendizagem. Quando plotados os gráficos por nota, o cenário indica que todos os alunos que alcançaram as melhores notas, do intervalo de três a quatro pontos, mantiveram o acessos aos conteúdos disponibilizados, acima de 200 vezes, conforme representado na Figura 6.

Figura 6: Concentração de *Outliers* acima da nota 3



Fonte: Os autores

A turma 04_10_2018 com 31 discentes, apresenta o menor número de alunos em relação as demais, no entanto foi mantida no grupo para demonstrar que a tendência de melhores notas, se caracteriza com acesso superiores a média encontrada. Em todas as turmas existem um número considerável de alunos que ficaram com nota zero e acessaram acima do valor médio e necessário para aprovação, no entanto, estes discentes não cumpriram o mínimo estipulado no curso para para obtenção de notas nas atividades.

A Figura 7 representa os valores apurados dos elementos dos BoxPlot, como o limite inferior, primeiro quartil, mediana, média, desvio padrão, terceiro quartil e limite superior.

Figura 7: Tabela resumo do conjunto de dados

Notas									
N_01_09_2018		N_02_09_2018		N_05_09_2018		N_03_10_2018		N_04_10_2018	
Min.	: 0.00								
1st Qu.	: 3.23	1st Qu.	: 3.70	1st Qu.	: 0.00	1st Qu.	: 3.50	1st Qu.	: 1.85
Median	: 3.80	Median	: 4.00	Median	: 2.50	Median	: 3.90	Median	: 3.70
Mean	: 3.25	Mean	: 3.57	Mean	: 1.97	Mean	: 3.43	Mean	: 2.80
Sd	: 1.21	Sd	: 0.98	Sd	: 1.73	Sd	: 1.06	Sd	: 1.63
3rd Qu.	: 4.00	3rd Qu.	: 4.00	3rd Qu.	: 3.70	3rd Qu.	: 4.00	3rd Qu.	: 3.85
Max.	: 4.00								

Acessos									
A_01_09_2018		A_02_09_2018		A_05_09_2018		A_03_10_2018		A_04_10_2018	
Min.	: 1.00	Min.	: 2.00	Min.	: 0.00	Min.	: 2.00	Min.	: 1.00
1st Qu.	: 71.00	1st Qu.	: 66.00	1st Qu.	: 1.00	1st Qu.	: 65.00	1st Qu.	: 65.00
Median	: 99.00	Median	: 90.50	Median	: 72.00	Median	: 89.00	Median	: 100.00
Mean	: 112.10	Mean	: 103.90	Mean	: 82.49	Mean	: 99.94	Mean	: 108.70
Sd	: 67.14	Sd	: 65.65	Sd	: 93.87	Sd	: 51.13	Sd	: 78.29
3rd Qu.	: 137.00	3rd Qu.	: 126.00	3rd Qu.	: 118.50	3rd Qu.	: 125.75	3rd Qu.	: 137.50
Max.	: 497.00	Max.	: 724.00	Max.	: 530.00	Max.	: 364.00	Max.	: 383.00

Fonte: Os autores

Para a comparação dos dois conjuntos de dados utilizados (Notas e Acessos), realizamos o cálculo do coeficiente de variação aplicando a Fórmula 1, visto que a grandeza das medidas são diferentes.

$$CV = \frac{\sigma}{\mu} \cdot 100 \quad (1)$$

Onde, σ é o desvio padrão, μ a média dos dados e **CV** o coeficiente de variação, que analisa a dispersão em termos relativos e o resultado apresentado em porcentagem. Quanto menor o valor, mais homogêneos serão os dados, e menor será a dispersão em torno da média.

O coeficiente de variação é usado para analisar a dispersão em termos relativos a seu valor médio quando duas ou mais séries de valores apresentam unidades de medida diferentes. Conforme Figura 8, demonstramos que a variabilidade dos dados são de alta dispersão, os dados são heterogêneos pois apresentam mais de 30% para o coeficiente de variação, com exceção da turma 02_09-2018, que o coeficiente de variação para nota é de 27,62%, sinalizando a média dispersão, no entanto, muito próximo ao limite para alta dispersão.

Figura 8: Média, Desvio Padrão e Coeficiente de Variação das turmas

Notas									
01_09-2018		02_09-2018		05_09-2018		03_10-2018		04_10-2018	
Média	3,25	Média	3,57	Média	1,97	Média	3,43	Média	2,80
Desvio Padrão	1,21	Desvio Padrão	0,98	Desvio Padrão	1,73	Desvio Padrão	1,06	Desvio Padrão	1,63
Coeficiente de variação	37,23%	Coeficiente de variação	27,56%	Coeficiente de variação	87,99%	Coeficiente de variação	30,96%	Coeficiente de variação	58,20%

Acessos									
01_09-2018		02_09-2018		05_09-2018		03_10-2018		04_10-2018	
Média	112,10	Média	103,90	Média	82,49	Média	99,94	Média	108,70
Desvio Padrão	67,14	Desvio Padrão	65,65	Desvio Padrão	93,87	Desvio Padrão	51,13	Desvio Padrão	78,29
Coeficiente de variação	59,89%	Coeficiente de variação	63,19%	Coeficiente de variação	113,80%	Coeficiente de variação	51,16%	Coeficiente de variação	72,02%

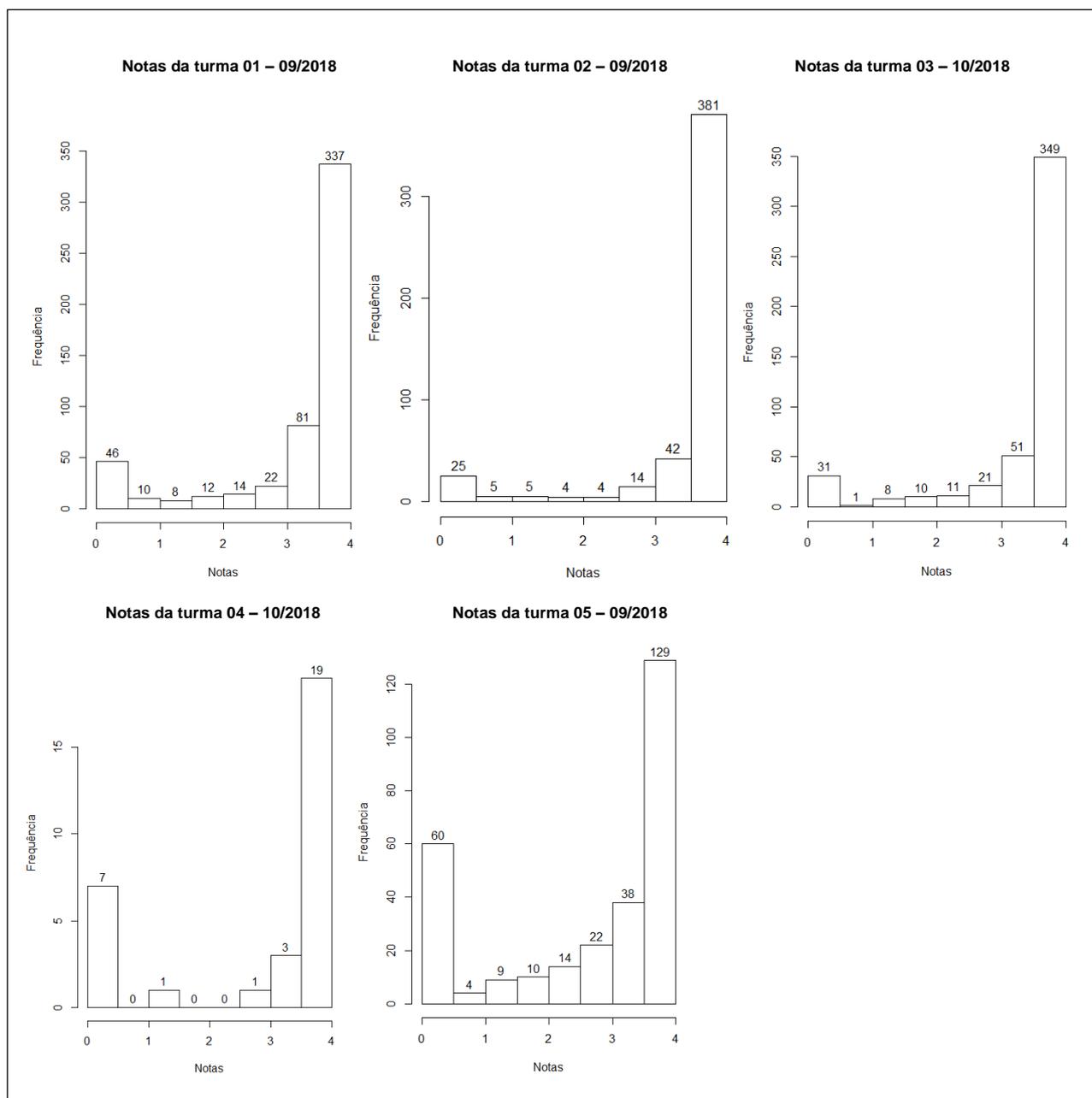
Fonte: Os autores

O desvio padrão é uma medida que expressa o grau de dispersão de um conjunto de dados. Ou seja, o desvio padrão indica o quanto um conjunto de dados é uniforme. Quanto mais próximo de 0 for o desvio padrão, mais homogêneo são os dados. E por conta

da diferença significativa do número de acesso em relação ao grupo, foi feita uma análise mais detalhada em relação a cada tipo de acesso efetuado pelo discente.

Identificou-se ocorrências de um número de acessos fora do conteúdo, e estes alunos que consultaram o ambiente, não realizaram as atividades necessárias para obtenção das notas conforme o histograma apresentado na figura 9.

Figura 9: Histograma com das notas das turmas



Fonte: Os autores

A razão dos acessos em relação ao número de notas baixas apresentou 13,65% para a turma 01-09_2018, 6,56% para a turma 02-09_2018, 8,88% para a turma 03-10_2018. Consultaram o tutor em 16,06%, 16,86% e 17,05%, na ordem.

As turmas 04-10_2018 e 05-09_2018 apresentou uma razão dos acessos em relação as notas zeros de 36,84% e 46,51%, o que pode estar associado ao elevado número de consultas ao tutor, com 26,56% e 23,17%, e não realização das atividades propostas no ambiente. O número de acesso aos tutores, indica que estes alunos enfrentaram maior dificuldade na compreensão das aulas.

Conclusões

Com o objetivo de verificar a evolução discente em relação aos seus acessos ao material disponibilizado no AVA, e como estas ações podem contribuir para a melhoria do seu aprendizado, o uso do RStudio possibilitou implementação adequada do *Learning Analytics* para identificar e analisar os dados apurados. Os resultados indicam que o maior acesso ao conteúdo disponibilizado no ambiente de estudo, possibilita ao discente melhor aproveitamento do conteúdo e conseqüentemente melhores notas em suas avaliações.

Após a análise, observou-se um desempenho semelhante entre todas as turmas, indicando que uma quantidade de 50 acessos é suficiente para obter-se o mínimo necessário do aprendizado visando a promoção. No entanto com a análise do expressivo número de *outliers*, comprovou-se no detalhamento das notas entre zero e quatro pontos, que com acesso superior a 200 vezes o discente é conduzido as melhores notas, que neste caso está no intervalo três e quatro pontos.

Em futuros trabalhos, pode-se pesquisar se as intervenções pontuais durante o aprendizado, contribuem efetivamente na real evolução discentes que se mantém no mínimo de acesso e notas.

Agradecimentos

O desenvolvimento deste trabalho foi possível graças ao auxílio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES/Brasil) e da Universidade Cruzeiro do Sul – São Paulo/SP-Brasil, através do programa de suporte à pós-graduação de instituições de ensino particulares (PROSUP-CAPES).

Referências

- CASTRO, L. N. FERRARI, D. G. **Introdução à mineração de dados**. Conceitos básicos, algoritmos e aplicações. São Paulo: Saraiva, 2016.
- CAMPOS, N. **Ensino adaptativo**: O big data na educação. 2014. Jornal Estadão. Disponível em <<https://educacao.estadao.com.br/blogs/a-educacao-no-seculo-21/ensino-adaptativo-o-big-data-na-educacao/>>. Acesso em 12/05/2019.

CHATTI, M. A. DYCKHOFF, A. L. SCHROEDER, U. THÜS, H. **A Reference Model for Learning Analytics**. 2012. International Journal of Technology Enhanced Learning (IJTEL). Disponível em: <https://www.thues.com/upload/pdf/2012/CDST12_IJTEL.pdf>. Acesso em : 10 de março de 2019.

COOPER, A. **What is Analytics?** Definition and Essential Characteristics. CETIS Analytics Series. Vol. No 5. Published by The University of Bolton. 2012. Disponível em <<http://publications.cetis.org.uk/wp-content/uploads/2012/11/What-is-Analytics-Vol1-No-5.pdf>>. Data de Acesso: 03 de março de 2019.

DIAS, R. S. **Caracterização do learning analytics na educação a distância**. SEPIT (Seminário de Pesquisa e Inovação Tecnológica). 2017. Disponível em <<http://editora.iftm.edu.br/index.php/sepit/article/download/312/145>>. Data de acesso em 02 de maio de 2019.

FONSECA, J. J. S. **Apostila de metodologia da pesquisa científica**. Fortaleza: Universidade Estadual do Ceará, 2002. Disponível em <<https://books.google.com.br/books?id=oB5x2SChpSEC&pg=PA20#v=onepage&q&f=false>>. Acesso em 14 de março de 2019.

FONSECA, E. S. ARAÚJO JR., C. F. **O envolvimento discente em um ambiente virtual de aprendizagem**. Análise realizada no curso de licenciatura em matemática. 2018. REnCiMa - Revista de Ensino de Ciências e Matemática. Disponível em <<http://revistapos.cruzeirosul.edu.br/index.php/rencima/article/view/1938>>. Acesso em 15 de abril de 2019.

FUNDAÇÃO R. **O Projeto R para Computação Estatística**. 2019. Disponível em <<https://www.r-project.org/>>. Acesso em 15 de março de 2019.

JOHNSON, L.; SMITH, R.; WILLIS, H.; LEVINE, A.; HAYWOOD, K. **The NMC Horizon Report: 2011 Higher Education Edition**. Relatório Técnico, New Media Consortium, 40p.

KENSKI, V. M. **Educação e tecnologias: o novo ritmo da informação**. 8. ed. Campinas, São Paulo: Papirus, 2012.

MACHADO, L. B. BECHER, E. L. **Aprendendo estatística com o software R**. Sociedade Brasileira de Educação Matemática. Encontro Nacional de Educação Matemática. 2016. Disponível em: <http://www.sbembrasil.org.br/enem2016/anais/pdf/6345_2685_ID.pdf>. Acesso em: 12 de abril de 2019.

MARRAS, J. P.; **Administração de Recursos Humanos: do operacional ao estratégico**. 15. ed. São Paulo: Saraiva, 2016.

MARTÍNEZ, D. L. R. GÓMES, C. E. P. **Contributions from Data Mining to Study Academic Performance of Students of a Tertiary Institute**. American Journal of Educational Research. 2014. Disponível em: <<http://www.sciepub.com/EDUCATION/abstract/2569>>. Acesso em: 02 de fevereiro de 2019.

MOISSA, B. GASPARINI, I. KEMCZINSKI, A. **Learning Analytics**: um mapeamento sistemático. TISE 2014. Nuevas Ideas en Informática Educativa. Disponível em <http://www.tise.cl/volumen10/TISE2014/tise2014_submission_262.pdf>. Data de Acesso em: 25 de março de 2019.

NOGUEIRA, S. P. FRANCISCO, R. E. SILVA, L. F. FERNANDES, M. A. PEREIRA JÚNIOR, C. X. **Big Data com Learning Analytics para apoiar o Planejamento Pedagógico Acadêmico**. 7º Workshop de Desafios da Computação aplicada à Educação (DesafIE). 2018. Disponível em: < <https://sol.sbc.org.br/index.php/desafie>>. Data de acesso: 10 de abril de 2019.

PEROVANO, D. G. **Manual de metodologia da pesquisa científica** [livro eletrônico]. Curitiba: InterSaberes, 2016.

ROMERO, C.; VENTURA, S. **Educational Data Mining: A Survey from 1995 to 2005**, Expert Systems with Applications (33), pp. 135-146. Elsevier. 2007.

RSTUDIO. **RStudio Desktop**. Open Source License. 2019. Disponível em <<https://www.rstudio.com/products/rstudio/download/#download>>. Acesso em 20 de março de 2019.

SILVA, W. R. M. SILVA, J. L. D. UGULINO, F. R. LUCENA, M. NUNES, I. D. **Learning Analytics como ferramenta para a análise do desempenho dos alunos em Cursos Semipresenciais**. V Congresso Brasileiro de Informática na Educação (CBIE 2016). Anais do XXVII Simpósio Brasileiro de Informática na Educação (SBIE 2016). Disponível em <<http://www.br-ie.org/pub/index.php/sbie/article/download/6708/4596>>. Data de acesso em 10 de abril de 2019.

SILVA JUNIOR, C. B. OLIVEIRA, Ivan Carlos Alcântara. **Learning Analytics** : revisão da literatura e o estado da arte. 2016. Disponível em: <<http://www.abed.org.br/congresso2016/trabalhos/329.pdf>>

SOUZA, E. F. M. PETERNELLI, L. A. MELLO, M. P. **Software Livre R: aplicação estatística**. 2014. Disponível em: <<http://www.de.ufpb.br/~tarciana/MPIE/ApostilaR.pdf>>. Data de Acesso: 15 de abril de 2019.