

Mineração de dados educacionais em uma disciplina ofertada na modalidade de ensino a distância

Educational data mining in a discipline offered in the distance learning modality

Minería de datos educativos en una disciplina ofrecida en la modalidad de educación a distancia

Recebido: 07/08/2020 | Revisado: 16/08/2020 | Aceito: 17/08/2020 | Publicado: 20/08/2020

Enir da Silva Fonseca

ORCID: <https://orcid.org/0000-0002-5308-6149>

Centro Universitário Lusíada, Brasil

E-mail: enir.fonseca@gmail.com

Carlos Fernando de Araújo Jr.

ORCID: <https://orcid.org/0000-0001-8188-161X>

Universidade Cruzeiro do Sul, Brasil

E-mail: carlos.araujo@cruzeirodosul.edu.br

Frederico Kauffmann Barbosa

ORCID: <https://orcid.org/0000-0002-8382-8991>

Centro Universitário Lusíada, Brasil

E-mail: barbosafk@outlook.com

Luiz Henrique Amaral

ORCID: <https://orcid.org/0000-0002-8606-4026>

Universidade Cruzeiro do Sul, Brasil

E-mail: luiz.amaral@unicid.edu.br

Resumo

A mineração de dados procura identificar relações entre os dados, para se criar informações que podem gerar novos conhecimentos ao desenvolvimento da ciência, assim como para tomada de decisões subsidiando novas ações que transforme a realidade atual. E o uso de técnicas de mineração de dados educacionais, que tem como foco o desenvolvimento de métodos para se explorar estes conjuntos de dados. Procurou-se analisar neste estudo, os acessos e notas discente a partir de um ambiente virtual de aprendizagem, aferindo a evolução de acordo com o número de acessos. Foram consideradas para a pesquisa três turmas da

disciplina matemática financeira ofertada no segundo semestre de 2019, com os dados extraídos de todas as atividades do usuário nas áreas de conteúdo, realizando-se uma análise de acordo com o processo CRISP-DM com implementação no *software* R e RStudio. Os resultados indicam que o maior período acessos ao ambiente de estudo, estão entre segunda e quarta-feira, e comprovou-se que com o mínimo de 55 vezes de interação no ambiente, os discentes tendem em obter a nota mínima para aprovação, ao passo que acesso acima de 100 vezes indicam notas próximas ao valor máximo.

Palavras-chave: Descoberta do conhecimento; Mineração de dados; Processo decisório.

Abstract

Data mining seeks to identify relationships between data, to create information that can generate new knowledge for the development of science, as well as for decision-making by subsidizing new actions that transform the current reality. And the use of educational data mining techniques, which focuses on the development of methods to explore these data sets. In this study, we tried to analyze the accesses and student grades from a virtual learning environment, measuring the evolution according to the number of accesses. Three classes of the mathematical financial discipline offered in the second semester of 2019 were considered for the research, with the data extracted from all the user activities in the content areas, carrying out an analysis according to the CRISP-DM process with implementation in the software. R and RStudio. The results indicate that the longest period of access to the study environment is between Monday and Wednesday, and it has been shown that with a minimum of 55 times of interaction in the environment, students tend to obtain the minimum grade for approval, while that access over 100 times indicate notes close to the maximum value.

Keywords: Knowledge discovery; Data mining; Decision-making process.

Resumen

La minería de datos busca identificar relaciones entre datos, generar información que pueda generar nuevos conocimientos para el desarrollo de la ciencia, así como para la toma de decisiones mediante el subsidio de nuevas acciones que transformen la realidad actual. Y el uso de técnicas de minería de datos educativos, que se centra en el desarrollo de métodos para explorar estos conjuntos de datos. En este estudio, se intentó analizar los accesos y las calificaciones de los alumnos desde un entorno de aprendizaje virtual, midiendo la evolución en función del número de accesos. Se consideraron para la investigación tres clases de la disciplina matemática financiera ofrecida en el segundo semestre de 2019, con los datos

extraídos de todas las actividades del usuario en las áreas de contenido, realizando un análisis según el proceso CRISP-DM con implementación en el software. R y RStudio. Los resultados indican que el período más largo de acceso al ambiente de estudio es entre el lunes y el miércoles, y se ha demostrado que con un mínimo de 55 veces de interacción en el ambiente, los estudiantes tienden a obtener la nota mínima de aprobación, mientras que que acceden más de 100 veces indican notas cercanas al valor máximo.

Palavras chave: Descubrimiento del conocimiento; Procesamiento de datos; Proceso de toma de decisiones.

1. Introdução

Diariamente e motivados principalmente pelas novas tecnologias de coleta e armazenamento de dados, diversos setores produzem dados em grande quantidade, e que atualmente são conhecidos como *Big Data*. Para (Caldas & Silva, 2016) o conceito de *Big Data* está focado principalmente em conjunto extremamente grandes, gerados a partir de processos e práticas tecnológicas, e (Erl, Khattak & Buhler, 2016, p. 4, tradução nossa) completam que o “*Big Data* é um campo dedicado à análise, processamento e armazenamento de grandes coleções de dados que frequentemente se originam de fontes diferentes”. Para (Righetto, 2020), a expressão *Big Data*, pode ser traduzida como megadados ou grande quantidade de dados. E de acordo com (Fonseca, 2020), refere-se à capacidade da humanidade de gerar a cada segundo um grande volume de dados estruturados e não estruturados, produzidos a exemplo na área da Saúde, Educação, Negócios, entre outros. Parte ponderável desses dados devem ser analisados pontualmente para adequada tomada de decisão.

Para (Patricio & Magnoni, 2018), a mineração de dados em contextos educacionais tem se tornado um campo emergente de exploração de dados em ambientes educacionais, e desta forma, muitas instituições educacionais que fazem uso de Ambientes Virtuais de Aprendizagem (AVA), ou outras tecnologias de apoio ao processo de ensino e aprendizagem, estão gerando dados que por vezes não são analisados prontamente. Para (Morais et al., 2018), as instituições estão investindo cada vez mais em *Big Data* melhorando a prestação de informações aos gestores e auxiliando, dessa forma na tomada de decisões. Considerando o volume gerado e atentos aos requisitos de qualidade e consistência, de obtenção rápida e de caráter oportuno, surge então um desafio macro que é o de como explorar adequadamente esses dados, visando obter a informação valiosa.

Utilizar-se de processos que auxiliem neste caminho é de fundamental importância na análise e decisões pontuais e tomadas de decisão, e o processo CRISP-DM (*Cross Industry Standard Process for Data Mining*) é utilizado para o desenvolvimento de projetos de mineração de dados, que de acordo com (Araújo, 2018) é formado por um conjunto de fases e processos padrão, flexíveis e independentes da área e das ferramentas utilizadas, concebida para funcionar em qualquer tipo de negócio e aceitar a aplicação de várias técnicas, porém de forma estruturada e sistemática.

Para (Santos, 2017) as possibilidades de interação e produção do conhecimento por meio dos recursos computacionais avançam significativamente com os novos cenários educativos proporcionados pelos ambientes virtuais de aprendizagem. A necessidade de análise dos dados, nos remete a questões importantes, tais como: O que especificamente fazer com esses dados educacionais? Como reverter estas informações extraídas em benefícios para as instituições, para estudantes e professores envolvidos no contexto de um curso? Como cada um destes atores podem identificar e utilizar as informações escondidas nos dados coletados de tal modo que consigam tirar proveitos delas?

Empregou-se neste estudo, a mineração de dados educacionais em uma disciplina ofertada no ensino a distância, com a análise a partir dos acessos e notas discente, aferindo desta forma a evolução de acordo com a quantidade de acessos.

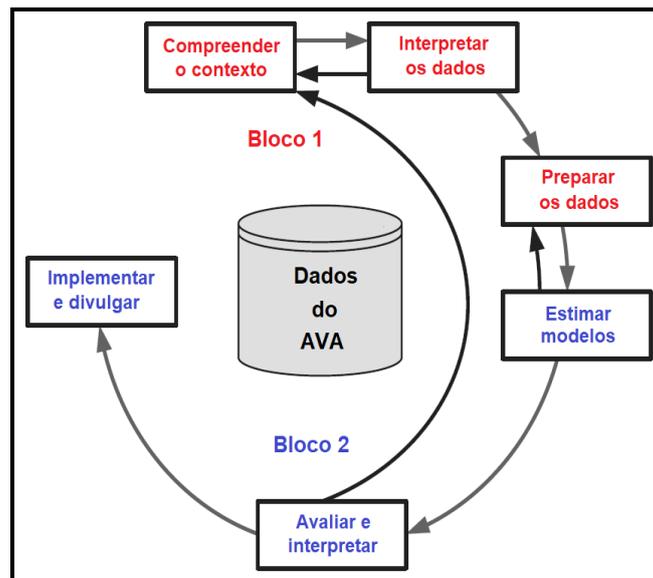
2. Mineração de Dados e Descoberta de Conhecimento

Mineração de Dados (MD), do inglês *Data Mining* (DM) e uma área interdisciplinar, mobilizando principalmente conhecimentos de análise estatística, aprendizagem de máquina, reconhecimento de padrões e visualização de dados, que de acordo com (Motta, 2016) é o uso das técnicas dentro do contexto educacional. A Mineração de Dados Educacionais (*Education Data Mining – EDM*), procura desenvolver ou adaptar métodos e algoritmos de mineração existentes, de tal modo que se prestem a compreender melhor os dados em contextos educacionais, produzidos principalmente por estudantes e professores, considerando os ambientes que interagem. Segundo (Silva Junior & Oliveira, 2016), a EDM trata da aplicação de técnicas de mineração de dados para resolver problemas da área educacional, e dentre as técnicas há diversas tarefas envolvidas em EDM, notadamente as que decorrem diretamente da análise de dados gerados nas interações dos estudantes com os ambientes de aprendizagem.

O processo para MD utilizado nesta pesquisa é o CRISP-DM apresentado na Figura 1, que de acordo com (Campos & Cazella, 2018) e (Chapman et al., 2000), consiste em seis

fases que se interligam e não possuem necessariamente uma ordenação rígida, pois trata-se de um processo cíclico onde o resultado de cada fase é que determina qual fase deverá ser realizada posteriormente. Para (Fonseca & Araújo Jr., 2018), a fim de se obter a confiabilidade nos modelos, testes e validações devem ser executados para extrair indicadores que auxiliem na análise dos resultados.

Figura 1: Modelo CRISP-DM em um AVA.



Fonte: Campos & Cazella (2018) e Chapman et al. (2000), adaptado por Fonseca (2020).

Conforme o modelo adaptado por (Fonseca, 2020) e apresentado na Figura 1, a coleta e organização dos dados definidas no primeiro bloco, emprega-se as etapas para compreensão do contexto, interpretação dos dados e preparação dos dados, e em resultados e discussões no bloco 2, as tarefas de estimar o modelo, avaliar e interpretar, implementar e divulgar.

3. Metodologia da Pesquisa

Pesquisas buscam trazer novos saberes para a sociedade como preconiza (Pereira et al., 2018), e neste estudo aplicamos uma pesquisa quantitativa e transversal temporal. A quantitativa ao classificar e analisar os dados em valores quantificável, traduzindo em números, identificando a relação entre as variáveis. (Zanella, 2013) afirma que a pesquisa quantitativa é aquela que se caracteriza pelo emprego de instrumentos estatísticos, tanto na coleta como no tratamento dos dados, e que tem como finalidade medir relações entre as variáveis.

Pesquisa transversal temporal, ao coletar dados de um intervalo específico de tempo, contexto e com análises de um momento particular. Para (Gomes & Gomes, 2020) destaca que o objetivo da pesquisa transversal é construir uma análise do objeto em um momento específico. Pode observar o mesmo fenômeno em diferentes grupos de sujeitos e analisar se há de fato alguma correlação entre eles, na tentativa de encontrar a existência ou não de relações de causa e efeito.

4. Materiais e Métodos

Para o desenvolvimento deste trabalho, empregou-se para coleta e organização dos dados com etapas definidas no bloco 1 da Figura 1, realizando-se a compreensão do contexto, interpretação dos dados e preparação dos dados. A coleta dos dados foi realizada em três turmas da disciplina de Matemática Financeira disponibilizada no ambiente *Blackboard*, com carga horária de 80 horas, ofertada no segundo semestre de 2019.

A primeira turma com 674 alunos matriculados apresentados na Tabela 1, nos cursos de Administração, Ciências Contábeis, Ciências Econômicas, Gestão Comercial, Gestão da Qualidade, Gestão da Segurança Privada, Gestão de Recursos Humanos, Gestão do Agronegócio, Gestão Financeira, Gestão Hospitalar, Processos Gerenciais, Secretariado, Matemática e Pedagogia.

A segunda turma com um total de 45 alunos matriculados nos cursos de Administração e Ciências Contábeis. A terceira turma com 614 discentes nas disciplinas de Administração, Ciências Contábeis, Ciências Contábeis, Ciências Econômicas, *Design* Gráfico, Gestão da Segurança Privada, Gestão do Agronegócio, Gestão Financeira, Gestão Hospitalar, Processos Gerenciais, Secretariado, Engenharia de Produção e Matemática, conforme Tabela 1.

Tabela 1: Quantidade de alunos por curso.

Turma 01_10		Turma 02_10		Turma 04_10	
Cursos	Alunos	Cursos	Alunos	Cursos	Alunos
Administração (Bacharelado)	63	Administração (Semipresencial)	43	Administração (Bacharelado)	102
Ciências Contábeis (Bacharelado)	5	Ciências Contábeis (Semipresencial)	1	Ciências Contábeis (Bacharelado)	7
Ciências Contábeis (P/ Bach Adm Econ Atuariais)	2	Ciências Contábeis (Bacharelado)	1	Ciências Contábeis (P/ Bach Adm Econ Atuariais)	3
Ciências Econômicas (Bacharelado)	51			Ciências Econômicas (Bacharelado)	30
Cst em Gestão Comercial	1			Cst em Design Gráfico	1
Cst em Gestão da Qualidade	1			Cst em Gestão da Segurança Privada	161
Cst em Gestão da Segurança Privada	208			Cst em Gestão do Agronegócio	36
Cst em Gestão de Recursos Humanos	2			Cst em Gestão Financeira	56
Cst em Gestão do Agronegócio	30			Cst em Gestão Hospitalar	39
Cst em Gestão Financeira	40			Cst em Processos Gerenciais	83
Cst em Gestão Hospitalar	57			Cst em Secretariado	63
Cst em Processos Gerenciais	152			Engenharia de Produção (Bacharelado)	1
Cst em Secretariado	45			Matemática (Licenciatura)	32
Matemática (Licenciatura)	16				
Pedagogia (Licenciatura)	1				
	674	Total	45	Total	614

Fonte: Os autores.

A seleção ocorreu a partir dos dados relativos as participações dos discentes nas atividades do usuário nos Fóruns, onde identificou-se um resumo da atividade do usuário nos fóruns de discussão; nas atividades do usuário nos Grupos onde é apresentado um resumo da atividade do usuário nos grupos; na base de dados com o desempenho discentes, extraindo-se as notas médias e as interrupções dos alunos; no resumo geral da atividade do usuário, quando é apontado a atividade do usuário para todas as áreas na disciplina; e nos registros que contem a visão geral de atividade, onde é apresentado a atividade geral dentro de um curso específico.

Empregou-se para análise, o *software* R que pode ser integrado ao RStudio que possui um conjunto de ferramentas integradas e projetadas para programação e desenvolvimento. O *software* R, é um ambiente computacional especializado na manipulação, análise e visualização gráfica de dados e cálculos estatísticos. De acordo com Landeiro (2011), o R é um *software* livre para computação estatística e construção de gráficos que pode ser baixado e distribuído gratuitamente de acordo com a licença GNU (*General Public License* - Licença Pública Geral). Em sua sua versão básica é instalada automaticamente alguns pacotes primordiais ao uso, com outros para download e fácil reconhecimento.

5. Resultados e Discussão

O processo CRISP-DM permite boas práticas para gerenciar as etapas de um trabalho, facilitando a tomada de decisões, pois a implantação seguindo o modelo não é o fim do projeto, de acordo com (Fonseca, 2020), todo o ciclo pode ser corrigido e melhorado durante

o processo, sem a necessidade de um resultado para os ajustes, com a retroalimentação de novos dados e correções pontuais.

Para a apresentação dos resultados e as discussões, é seguido a fases do bloco 2 apresentado anteriormente na Figura 1, que propõe as etapas de estimar modelos, avaliar e interpretar, implementar e divulgar. Na Tabela 2 é apresentado o total de acessos por dia da semana e turma, assim como o total de cada dia.

Tabela 2: Acessos por disciplinas *versus* turmas nos dias da semana.

Turma 01-10								
Curso	Dom	Seg	Ter	Qua	Qui	Sex	Sáb	Total
Administração (Bacharelado)	850	1432	1246	1188	565	653	435	6369
Ciências Contábeis (Bacharelado)	0	102	57	64	2	2	0	227
Ciências Contábeis (P/ Bach Adm Econ Atuariais)	0	22	9	114	4	72	0	221
Ciências Econômicas (Bacharelado)	446	929	759	979	413	177	331	4034
Cst em Gestão da Qualidade	0	0	0	48	0	0	0	48
Cst em Gestão da Segurança Privada	966	4990	3486	3538	2172	1938	1365	18455
Cst em Gestão do Agronegócio	252	273	426	500	247	256	69	2023
Cst em Gestão Financeira	226	1155	650	650	494	328	387	3890
Cst em Gestão Hospitalar	585	1106	972	1262	840	400	343	5508
Cst em Processos Gerenciais	1134	2822	2502	2551	1159	1013	648	11829
Cst em Secretariado	204	895	636	557	423	358	425	3498
Matemática (Licenciatura)	123	318	197	151	262	134	67	1252
	4786	14044	10940	11602	6581	5331	4070	57354

Turma 02 -10								
Curso	Dom	Seg	Ter	Qua	Qui	Sex	Sáb	Total
Administração (Semipresencial)	350	1099	1117	953	421	396	280	4616
Ciências Contábeis (Semipresencial)	0	0	0	0	0	0	0	0
Ciências Contábeis (Bacharelado)	0	0	0	0	0	0	0	0
	350	1099	1117	953	421	396	280	4616

Turma 04 -10								
Curso	Dom	Seg	Ter	Qua	Qui	Sex	Sáb	Total
Administração (Bacharelado)	903	2061	2016	2029	1080	536	389	9014
Ciências Contábeis (Bacharelado)	11	121	36	234	22	38	73	535
Ciências Contábeis (P/ Bach Adm Econ Atuariais)	109	0	56	0	22	24	0	211
Ciências Econômicas (Bacharelado)	114	489	471	616	416	353	126	2585
Cst em Design Gráfico	33	35	21	5	0	17	1	112
Cst em Gestão da Segurança Privada	1307	3421	3492	3454	2110	1955	1514	17253
Cst em Gestão do Agronegócio	260	657	611	893	409	357	256	3443
Cst em Gestão Financeira	342	1199	971	1090	635	478	525	5240
Cst em Gestão Hospitalar	427	993	539	572	439	284	166	3420
Cst em Processos Gerenciais	715	1851	1776	1887	905	629	459	8222
Cst em Secretariado	366	1793	1350	1199	885	600	383	6576
Matemática (Licenciatura)	234	1106	507	575	184	246	127	2979
	4821	13726	11846	12554	7107	5517	4019	59590

Fonte: Os autores.

A concentração de acessos nas três turmas, 01_10, 02_10 e 04_10 encontra-se na segunda terça e quarta-feira com índice superior a 20% nos três dias indicados, com soma de

valores neste período de 65,63% para turma 01_10, 70,87% para turma 02_10 e 65,53% para a turma 04_10 conforme apresentado na Tabela 3 e plotados no Gráfico 1.

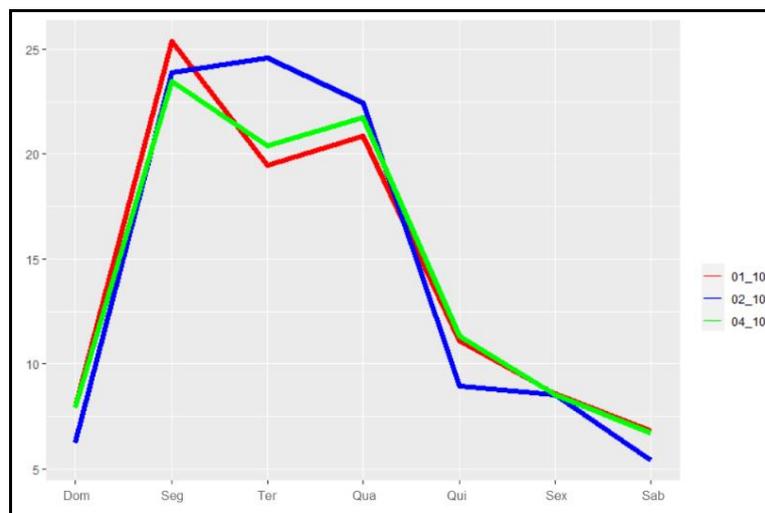
Tabela 3: Acessos dos alunos na semana.

Turma	Dom	Seg	Ter	Qua	Qui	Sex	Sáb
01_10	7,94%	25,35%	19,45%	20,83%	11,06%	8,57%	6,80%
02_10	6,25%	23,87%	24,57%	22,43%	8,94%	8,51%	5,43%
04_10	7,92%	23,47%	20,37%	21,72%	11,32%	8,51%	6,69%

Fonte: Os autores.

Com os índices identificados por dia da semana na Tabela 3, é possível analisando no Gráfico 1, como fica mais nítida a concentração de acessos para a turma 02_10 na terça-feira, diferente das turmas 01_10 e 04_10, com maior concentração na segunda e quarta-feira, o que pode ser justificado pelo menor número de discentes na turma 02_10.

Gráfico 1: Concentração de acessos durante a semana.



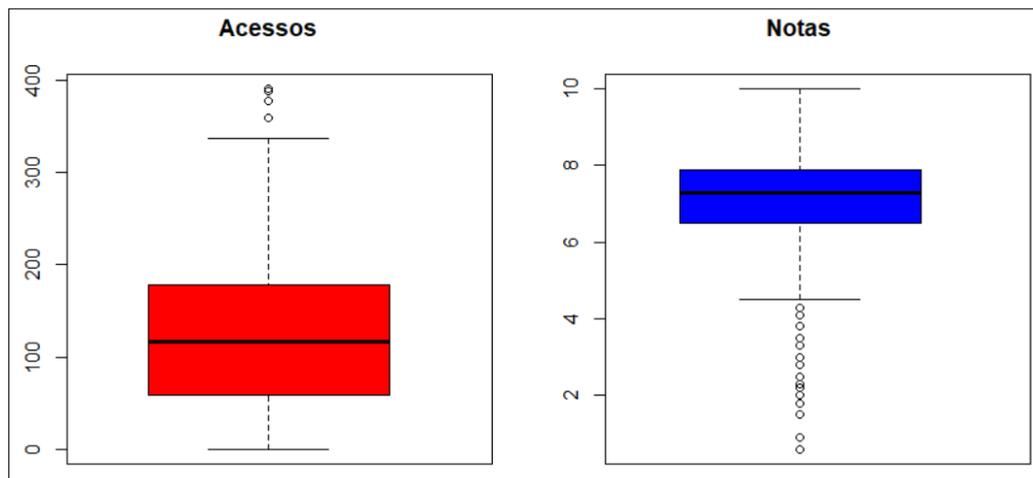
Fonte: Os autores.

A primeira turma representada pela linha vermelha no Gráfico 1 contou com 674 discentes, a segunda turma indicada com a linha azul com 45 alunos e o último grupo do estudo, representado pela linha verde com 614 discentes matriculados.

Para validar a correlação entre as variáveis acesso e notas, foi gerado um *boxplot*, com o objetivo de verificar a distribuição dos dados, e a dispersão é representada pela amplitude do gráfico, que pode ser calculada como máximo valor e mínimo valor. E de acordo com (Bussab & Morettin, 2010), a análise pode ser conduzida por meio de medidas-resumo,

histogramas ou *boxplot*. Ao analisar a relação dos acessos e nota final no Gráfico 2, observa-se no *Boxplot* acessos uma grande amplitude entre os valores máximo e mínimo do diagrama, no entanto o intervalo interquartílico, que é a diferença entre o terceiro quartil e o primeiro quartil, apresenta distribuição simétrica no conjunto de dados com linha da mediana no centro do retângulo.

Gráfico 2: Relação dos acessos e nota final.



Fonte: Os autores.

No *Boxplot* notas desenhadas no Gráfico 2, observa idêntica dispersão dos dados ente os valores máximo e mínimo, assim como distribuição simétrica no conjunto de dados. Os *outliers* em acessos, não interferem na análise dos dados, visto que apenas 4 alunos mantiveram acesso acima de 350 vezes. Para as notas, valida-se nesta análise que o baixo acesso ao sistema, representado posteriormente no Gráfico 3, indicam reprovação na disciplina. Pode-se observar na com a Tabela 4, o resumo dos acessos e notas.

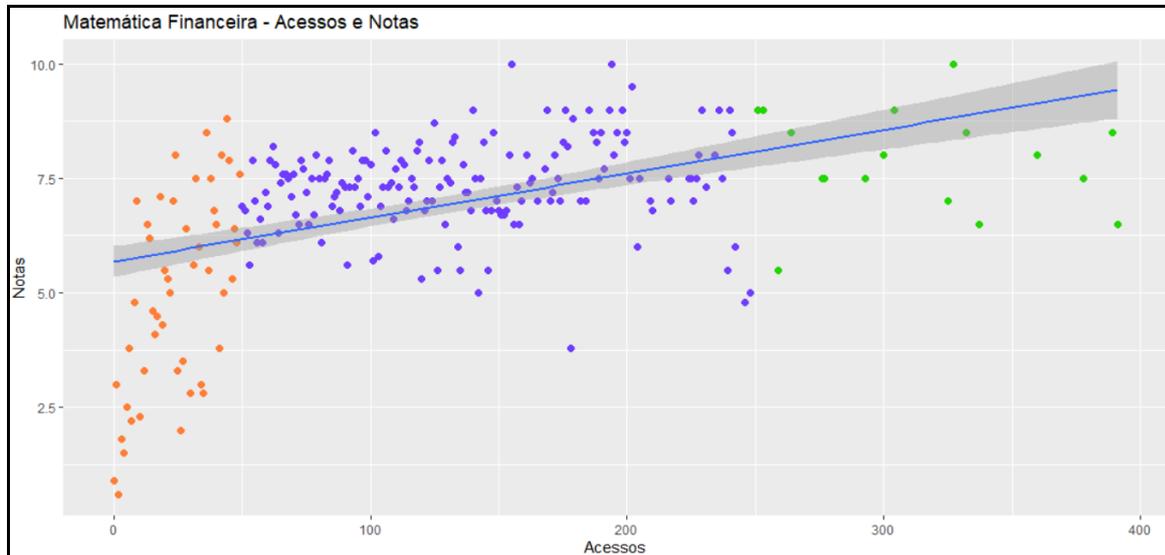
Tabela 4: Resumo dos acessos e notas.

Acessos		Notas	
Min.	: 0.00	Min.	: 0.60
1st Qu.	: 59.25	1st Qu.	: 6.50
Median	: 116.50	Median	: 7.30
Mean	: 126.16	Mean	: 6.89
3rd Qu.	: 177.75	3rd Qu.	: 7.90
Max.	: 391.00	Max.	: 10.00

Fonte: Os autores.

Com um total de 1333 discentes, ao agrupar os dados em um único banco de dados, verifica-se uma correlação entre as variáveis acessos e notas. No Gráfico 3 é apresentado dispersão em relação as notas com os acessos e a respectiva linha de tendência.

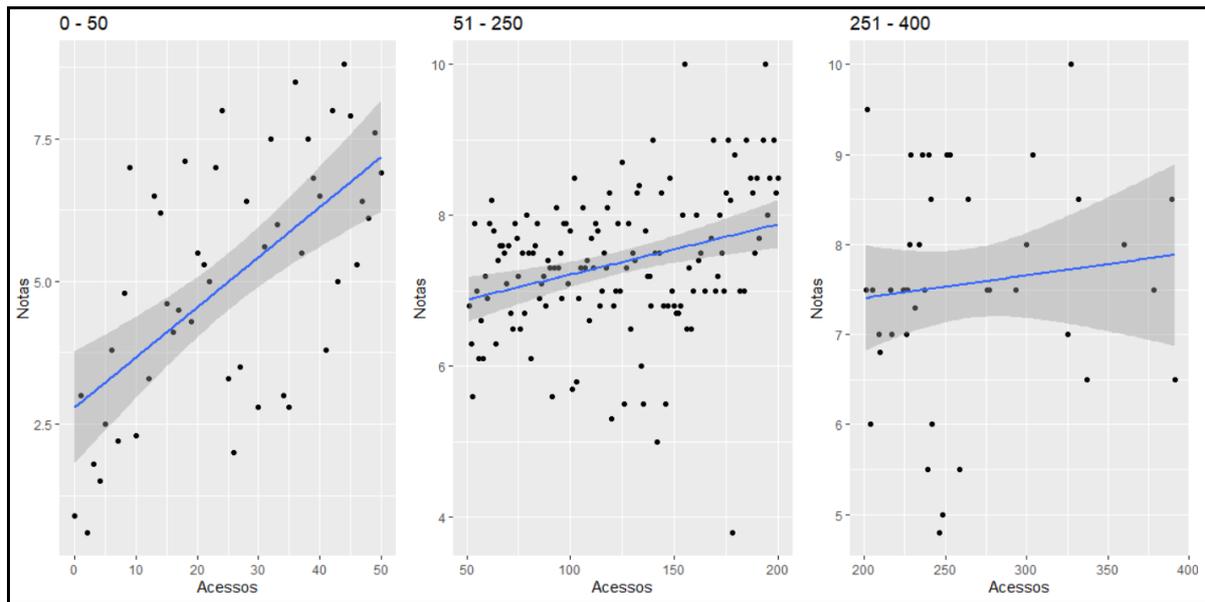
Gráfico 3: Concentração dos acessos e notas.



Fonte: Os autores.

Para análise da concentração dos acessos, os dados separados no Gráfico 3 e identificados pelas cores laranja, azul e verde, foram classificados nos intervalos de zero a sessenta, sessenta e um a duzentos, e duzentos e um a quatrocentos acessos, com intervalo e linha de tendência plotados no Gráfico 4.

Gráfico 4: Acesso e notas separadas por grupos.



Fonte: Os autores.

Ao calcular-se as medidas de tendências e analisar os dados plotados no Gráfico 3, observa-se uma relação linear positiva moderada entre as variáveis, pois alguns pontos estão perto da linha, mas outros pontos estão longe dela. Discentes que mantem acesso acima de 55 vezes, tendem em obter a nota mínima para aprovação, ao passo que acesso acima de 100 vezes ao material disponibilizado no ambiente virtual de aprendizagem, indicam a promoção com a notas próximo ao valor máximo.

6. Conclusões

Dessa análise surgem algumas demandas, dentre elas questões relacionadas a como melhorar a aprendizagem, como desenvolver ambientes educacionais mais eficazes que contribuam efetivamente para os estudantes aprenderem mais e em menos tempo. Em outra perspectiva, pretende-se saber quais processos se adequam para mineração de dados educacionais, quais ajustes devem ser feitos durante a análise, que são indicados no processo CRISP-DM.

Dentre as análises realizadas, identificou-se o período de maior acessos pelos discentes ao ambiente de estudo, que estão entre segunda e quarta-feira, e validou-se com a análise dos dados que com o mínimo de 55 vezes de acesso, os discentes tendem em obter a nota mínima para aprovação, ao passo que acesso acima de 100 vezes indicam notas próximas ao valor máximo. Cabe ressaltar e indicar a importância para investigações futuras, em quais

conteúdos disponibilizados o discente mantém o maior acesso, o tempo de permanência no ambiente, ou quais conteúdos complementares é acessado para o aprendizado.

Referências

Araújo, E. O. (2018) *Sistema de mineração de dados para apoiar a tomada de decisão em uma instituição de ensino superior - o problema da evasão escolar no IFTM*. Instituto Politécnico do Porto. Portugal. Disponível em:

<https://recipp.ipp.pt/bitstream/10400.22/13080/1/Disserta%C3%A7%C3%A3o%20-%20Eduardo%20de%20Oliveira%20Araujo.pdf>. Acesso em: 20 de fevereiro de 2020.

Bussab, W. O. & Morettin, P. A. (2010). *Estatística básica*. 6ª Edição. São Paulo: Saraiva.

Caldas, M. S. & Silva, E. C. C. (2016). *Fundamentos e aplicação do Big Data: como tratar informações em uma sociedade de yottabytes*. Bibliotecas Universitárias: pesquisas, experiências e perspectivas, Belo Horizonte, v. 3, n. 1, p. 65-83, jan./jun. 2016. Disponível em: <https://brapci.inf.br/index.php/res/v/17069>. Acesso em 15 de maio de 2020.

Campos, A. & Cazella, S. C. (2018). *Descoberta de conhecimento em base de dados sobre o perfil de estudantes brasileiros de tecnologia da informação*. VII Congresso Brasileiro de Informática na Educação (CBIE 2018). Anais dos Workshops do VII Congresso Brasileiro de Informática na Educação (WCBIE 2018). Disponível em:

<http://dx.doi.org/10.5753/cbie.wcbie.2018.449>. Acesso em: 10 de março de 2020.

Chapman, P. et al. (2000). *CRISP-DM 1.0. Step-by-step data mining guide*. Disponível em: <https://www.the-modeling-agency.com/crisp-dm.pdf>. Acesso em: 05 de agosto de 2020.

Erl, T., Khattak, W. & Buhler, P. (2016). *Big Data Fundamentals*. Concepts, Drivers e Techniques. Prentice Hall.

Fonseca, E. S. (2020). *Análise de desempenho discente em ambiente virtual de aprendizagem*. Mineração de dados educacionais através do processo CRISP-DM. Tese. Disponível em: <http://catalogodeteses.capes.gov.br/>. Acesso em 30 de maio de 2020.

Fonseca, E. S. & Araújo Jr., C. F. (2018). *O envolvimento discente em um ambiente virtual de aprendizagem*. Análise realizada no curso de licenciatura em matemática. REnCiMa, v. 9, n. 5, p. 189-204. Disponível em: <https://doi.org/10.26843/rencima.v9i5.1938>. Acesso em: 15 de maio de 2020.

Gomes, A. S. & Gomes, C. R. A. (2020). *Classificação dos tipos de pesquisa em informática na educação*. In: Jaques, P. A.; Pimentel, M.; Siqueira, S.; Bittencourt, I. (Org.) Metodologia de Pesquisa Científica em Informática na Educação: Concepção de Pesquisa. Porto Alegre: SBC. (Série Metodologia de Pesquisa em Informática na Educação, v. 1. Disponível em: <https://metodologia.ceie-br.org/livro-1/>. Acesso em 10 de março de 2020.

Landeiro, V. L. (2011). *Introdução ao uso do programa R*. Instituto Nacional de Pesquisas da Amazônia. 2011. Disponível em: <https://cran.r-project.org/doc/contrib/Landeiro-Introducao.pdf>. Acesso em 15 de fevereiro de 2020.

Morais, I. S. et al. (2018) *Introdução a big data e internet das coisas (IoT)*. Porto Alegre: SAGAH, 2018.

Motta, P. R. A. (2016). *Estudo exploratório do uso de classificadores para a predição de desempenho e abandono em universidades*. Disponível em: https://sucupira.capes.gov.br/sucupira/public/consultas/coleta/trabalhoConclusao/viewTrabalhoConclusao.jsf?popup=true&id_trabalho=3872977. Acesso em 01 de junho de 2020.

Patricio, T. S. & Magnoni, M. G. M. (2018). *Mineração de dados e big data na educação*. Revista GEMInIS, São Carlos, UFSCar, v. 9, n. 1, pp57-75. Disponível em <http://dx.doi.org/10.4322/2179-1465.0901004>. Acesso em 01 de junho de 2020.

Pereira, AS et al. (2018). *Metodologia da pesquisa científica*. [e-book]. Santa Maria. Ed. UAB/NTE/UFSM. Disponível em: https://repositorio.ufsm.br/bitstream/handle/1/15824/Lic_Computacao_Metodologia-Pesquisa-Cientifica.pdf?sequence=1.

Righetto, S. B. (2020). *Manutenção preditiva 4.0: conceito, arquitetura e estratégias de implementação*. Universidade Federal de Santa Catarina. 2020. Disponível em:

https://repositorio.ufsc.br/bitstream/handle/123456789/204345/TCC_Sophia%20Boing%20Ri ghetto.pdf?sequence=1. Acesso em: 05 de março de 2020.

Rstudio. (2020). *RStudio Desktop 1.3.959*. Open Source License. Disponível em: <https://www.rstudio.com/products/rstudio/download/#download>. Acesso em 20 de fevereiro de 2020.

Santos, W. T, (2017). *Integração de recursos computacionais para atividades docentes em sala de aula*. Educere. XIII Congresso Nacional de Educação. Disponível em: https://educere.bruc.com.br/arquivo/pdf2017/26092_14110.pdf. Acesso em 20 de maio de 2020.

Silva Junior, C. B. & Oliveira, I. C. A. (2016). *Learning Analytics: revisão da literatura e o estado da arte*. 2016. Disponível em: <http://www.abed.org.br/congresso2016/trabalhos/329.pdf>. Acesso em: 12 de abril de 2020.

Zanella, L. C. H. (2013). *Metodologia de pesquisa*. 2ª Edição. Florianópolis. Departamento de Ciências de Administração. UFSC. Disponível em: http://arquivos.eadadm.ufsc.br/EaDADM/UAB_2014_2/Modulo_1/Metodologia/material_didatico/Livro%20texto%20Metodologia%20da%20Pesquisa.pdf. Acesso em 10 de fevereiro de 2020.

Porcentagem de contribuição de cada autor no manuscrito

Enir da Silva Fonseca – 25%

Carlos Fernando de Araújo Jr. – 25%

Frederico Kauffmann Barbosa – 25%

Luiz Henrique Amaral – 25%