

Análise dos Perfis de Alunos do Ensino Superior sobre a Realização de Aulas na Modalidade a Distância Durante Pandemia da Covid-19 Usando Algoritmos de Aprendizagem de Máquina

Laertes Pereira Junior¹, Simone Nasser Matos¹, Helyane Bronoski Borges¹

¹Depto de Informática – Universidade Tecnológica Federal do Paraná (UTFPR-PG)
Caixa Postal 15.064 – 91.501-970 – Ponta Grossa – PR – Brazil

Resumo. Este artigo propõe analisar perfis de alunos do ensino superior sobre o ensino na modalidade a distância durante a pandemia da Covid-19 usando mineração de dados. A metodologia de pesquisa utilizada foi exploratória composta por 5 fases: entendimento do problema, construção de um formulário para coleta de dados, pré-processamento dos dados, aplicação de algoritmos de aprendizagem de máquina e avaliação dos resultados. Os resultados encontrados identificaram grupos de alunos com características distintas contendo diferentes opiniões sobre a aplicação do ensino a distância. Foi possível classificar os perfis dos alunos por meio de uma análise dos atributos mais relevantes em cada um dos grupos.

Palavras-chaves: Educação. Ensino a distância. Aprendizagem de Máquina. Covid-19.

Abstract. This paper proposes to analyze the profiles of higher education students on the distance learning method (EaD) during the Covid-19 pandemic period using data mining. The research methodology used for this work was exploratory and consisted of 5 phases: understanding the problem, development of a form to collect the data, data pre-processing, application of machine learning algorithms and evaluation of the results. The results found showed different characteristics for groups of students with different opinions on the application of the distance learning, and was possible to classify the students through an analysis of the most relevant attributes in each of the groups.

Keywords: Education. Distance Learning. Machine Learning. Covid-19.

1. Introdução

A Inteligência Artificial (IA) aplicada a educação é uma área de pesquisa multi e interdisciplinar, pois considera o uso das tecnologias da IA em softwares que possuem como objetivo o ensino e a aprendizagem (VICARI, 2018). Segundo Luckin *et al.* (2016, p. 18), a IA educacional possui como pilar principal: “tornar formas computacionalmente precisas e explícitas de conhecimento educacional, psicológico e social que muitas vezes são deixadas implícitas”.

A IA trabalha com a investigação, pois examina a forma que o ser humano raciocina. Ou seja, transforma o pensamento em tecnologia e dedica seus esforços com o objetivo de construir esclarecimentos algorítmicos dos processos mentais humanos. Tais processos dividem-se em quatro grupos: a) área conexionista, capacidade dos computadores de identificarem e aprenderem modelos; b) segmento ligado à biologia molecular na tentativa de gerar vida artificial; c) campo referente à robótica em conjunto com a biologia, com o intuito de criar máquinas que hospedem IA; d) área compatível com a psicologia, epistemologia e sociologia que procura mostrar à máquina formas de raciocínio e procura (FAVA, 2016).

De acordo com Vicari (2018), os sistemas educacionais com maiores aplicações que utilizam da tecnologia da IA são os: Sistemas Tutores Inteligentes Afetivos (STIs), os *Learning Management Systems* (LMSs), a Robótica Educacional Inteligente e os *Massive Open Online Courses* (MOOCs), sendo estes todos referentes a *Learning Analytics* (LA). Além disso, cada uma dessas aplicações utilizam as técnicas da IA de forma diferente, uma vez que se tratam de aplicações distintas.

O aprendizado de máquina e mineração de dados estão ligados à área de *Learning Analytics* e Mineração de Dados Educacionais. Para Isotani e Pinto (CIEB, 2019), as duas áreas, apesar de distintas são fortemente correlatas e possuem como objetivo melhorar a educação a partir da análise e do uso de grandes quantidades de dados. Isto provê ferramentas para melhor entender os problemas educacionais, tanto do ponto de vista dos estudantes quanto dos professores e gestores e permite o desenvolvimento de modelos e estratégias que ajudam no processo de ensino e aprendizagem.

Baker *et al.* (2012) relatam que a mineração de dados educacionais é uma área que adapta algoritmos e métodos de mineração de dados para a melhor compreensão dos dados em um contexto educacional. Esta área tem desenvolvido trabalhos focados em prever o risco de evasão de alunos (DO NASCIMENTO *et al.*, 2018) ou visa compreender melhor a maneira como eles aprendem (LORENZE *et al.*, 2018).

Com o propósito de identificar características relevantes sobre o perfil de estudantes universitários da região de Ponta Grossa – PR, no que tange as aulas na modalidade de ensino a distância (EAD), este trabalho aplicou algoritmos de aprendizagem de máquina *Self-organizing map* (SOM) (KOHONEN, 1990) e *k-means* (PIMENTEL *et al.*, 2003) para realizar uma análise comparativa sobre os perfis.

2 Trabalhos Relacionados

Hijazi e Naqvi (2006) conduziram um estudo sobre o desempenho de estudantes, com base numa hipótese, enquadrada por eles, declarada como “Ações do aluno em relação à frequência às aulas, horas gastas no estudo diariamente após a faculdade, renda familiar dos alunos, idade da mãe e educação da mãe está significativamente relacionada ao desempenho do aluno”. A partir do uso do modelo de regressão linear simples, verificou que fatores como a educação da mãe e a renda familiar do aluno estão altamente correlacionados com o desempenho acadêmico do aluno.

Kampff *et al.* (2008) propuseram aplicar mineração de dados na construção de alertas em ambientes virtuais de aprendizagem para identificar perfis de alunos com risco de evasão ou reprovação, de forma a promover alterações nos AVA que facilitem a implementação de alertas como apoio a prática docente.

Shih, Koedinger e Scheines (2008) desenvolveram uma pesquisa para descobrir táticas de estudos de alunos utilizando descoberta não-supervisionada, em que o algoritmo incorpora medidas educacionais no nível do aluno diretamente no processo de aprendizagem. Os dados para este estudo derivam da plataforma tutora de ensino *Geometry Cognitive Tutor*. Os algoritmos de comportamento do aluno, além de prever o ganho de aprendizagem, sugerem que os alunos que aprendem melhor tendem a fazer tentativas mais persistentes ao invés de utilizar a ajuda do software.

Lopez *et al.* (2012) compararam técnicas de agrupamento para prever notas finais baseadas nas respostas de participação de alunos em fóruns, buscando a existência de uma correlação entre as notas finais dos estudantes com a participação dos mesmos nos fóruns educativos.

Bazaldua, Baker e San Pedro (2014) utilizaram dados de respostas de atividades de alunos, submetidos em um ambiente de aprendizagem online na área da matemática, objetivando encontrar regras de associação para avaliar as relações entre os comportamentos afetados e desmembrados dos alunos durante o uso do sistema.

Rau, Mason e Nowak (2016) utilizaram algoritmos de Aprendizagem e Máquina para analisar a similaridade de representações visuais de moléculas químicas apresentadas a alunos de graduação em Química de uma universidade dos EUA. A partir das respostas,

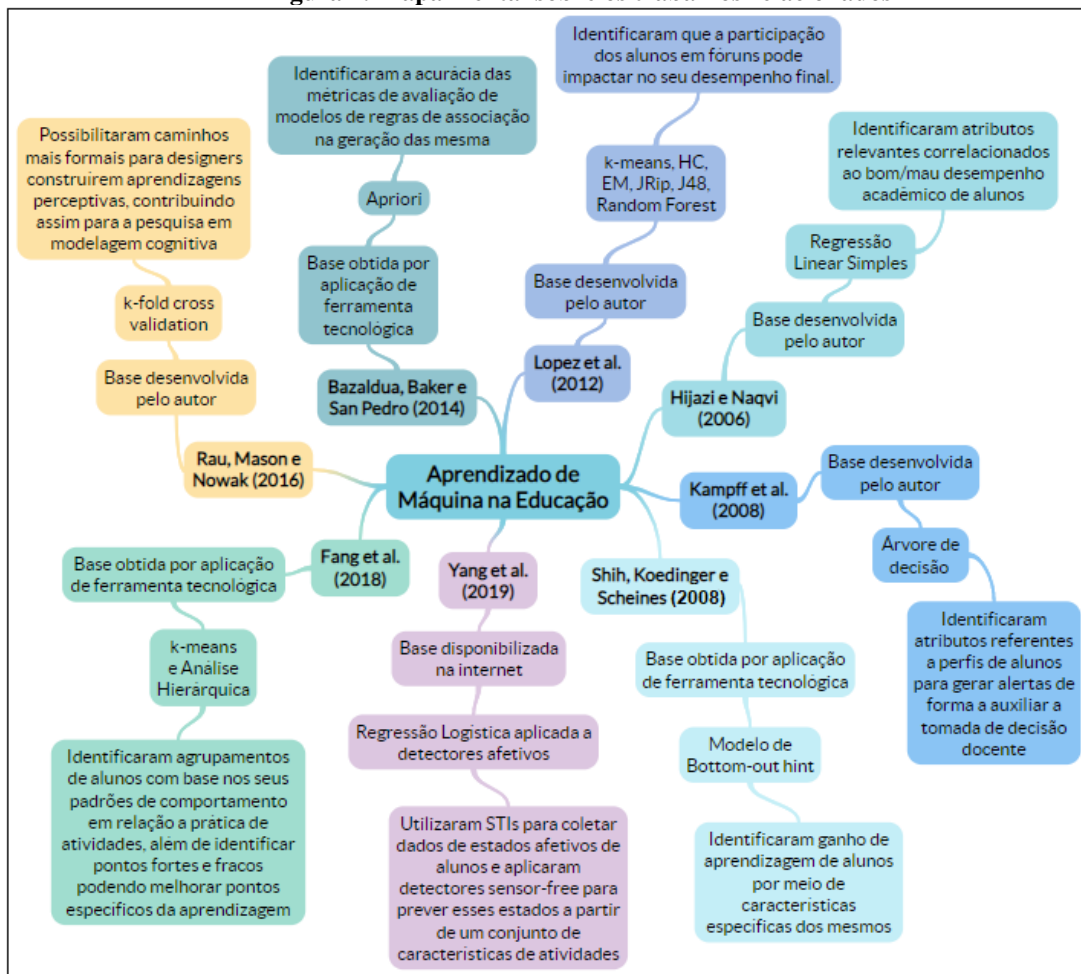
eles aplicaram duas abordagens para medir a similaridade descritas como: aprendizado de similaridade por *ranking*, e escala multidimensional não métrica. Eles não conseguiram identificar uma grande diferença entre os dois modelos, porém recomendaram o uso de uma das abordagens por otimizar o erro de previsão, cuja característica mede objetivamente a qualidade entre os modelos.

Fang *et al.* (2018) utilizaram as técnicas de agrupamento *k-means* e Análise Hierárquica para identificar padrões na aprendizagem de adultos com baixa alfabetização a partir da interação com um sistema tutor inteligente. Eles conseguiram separar os alunos em grupos, entre: leitores proficientes, leitores com dificuldades, leitores conscientes e leitores desengajados, baseando-se nos seus padrões de comportamento em relação as atividades praticadas, além de identificarem pontos fortes e fracos dos alunos podendo melhorar materiais específicos da aprendizagem para cada indivíduo.

Yang *et al.* (2019) utilizaram sistemas tutores inteligentes para alunos em salas de aula coletando dados afetivos e aplicando estes a detectores afetivos para prever estados afetivos dos alunos a partir de um conjunto de características das atividades executadas.

A Figura 1 apresenta as pesquisas discutidas nesta seção por meio de um mapa mental, identificando os autores, a obtenção da base de dados, os algoritmos trabalhados e os principais objetivos alcançados.

Figura 1. Mapa mental sobre os trabalhos relacionados



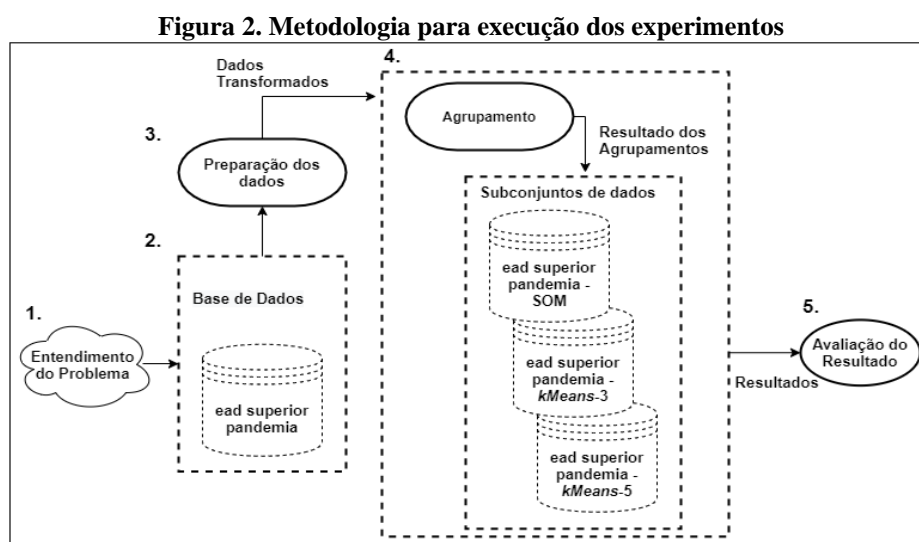
Observa-se que os trabalhos apresentados buscam seus dados para análise, no

geral, a partir de informações e respostas de alunos. Os autores utilizaram métodos como pesquisa ao público (desde que o público esteja inserido no ambiente educacional), dados de alunos de universidades, e a maioria das pesquisas extraíram as informações de ferramentas educacionais, pois essas plataformas produzem grande quantidade de dados e fornecem a possibilidade de encontrar informações importantes.

Também se pode concluir que os trabalhos implementam diferentes algoritmos, pois existem diversas implementações para um mesmo segmento da AM (classificação, agrupamento, associação), o que possibilita diversificar a escolha dos algoritmos para utilização. Os algoritmos de classificação e agrupamento aparecem em maior quantidade de aplicações e alguns fazem uso de algoritmos de associação. Ainda, como resultados, as pesquisas buscam identificar critérios para avaliação e melhorias no ensino, recuperação de informações e obtenção de critérios para aprimoramento de ferramentas.

3. Metodologia

A Figura 2 ilustra a metodologia utilizada para os experimentos. O processo foi baseado no método CRISP DM (*Cross Industry Standard Process for Data Mining*) (WIRTH, HIPPEL, 2000), abordando as fases: 1. Entendimento do Problema, 2. Compreensão dos Dados, 3. Preparação dos Dados, 4. Modelagem e 5. Avaliação.



O conjunto de dados para este trabalho foi obtido por meio de uma pesquisa com alunos do ensino superior presencial (graduação e pós-graduação). Os dados foram coletados a partir de um formulário desenvolvido na plataforma *Google Forms*, e foi divulgado publicamente via redes sociais, em páginas e grupos de universidades no *Facebook*, além de *Whatsapp* e *E-mail*. O período de coleta dos dados se deu no mês de abril de 2020 até o fim do mesmo, sendo as datas de 19/04/2020 até 27/04/2020. Como o formulário foi divulgado na forma *online*, foi possível analisar os registros da base e informações incompletas, erradas, que não pareciam ser verdadeiras ou que não condiziam com as esperadas foram removidas dos dados finais.

A base foi nomeada como *ead_superior_pandemia*, formada por 30 questões, sendo estas os respectivos atributos. Ao fim, um total de 483 respostas foram coletadas dos alunos do ensino superior presencial, constituindo os registros da base. A Figura 3 apresenta uma descrição dos atributos do conjunto de dados desenvolvido por categorias.

Figura 3 – Descrição do Conjunto de Dados *ead_superior_pandemia*

Categoria	Atributo	Descrição	Valores Possíveis
Informações pessoais	Idade	Idade dos estudantes	Entre 16 e 61 anos
	Sexo	Sexo/gênero dos estudantes	Masculino, Feminino
Informações acadêmicas	Nível Acadêmico	Grau de instrução	Graduação, Especialização, Mestrado, Doutorado
	Curso	Curso que está matriculado	Curso ficou para preenchimento pelo aluno
	Aluno Bolsista	Se é aluno bolsista	Sim, Não
Informações socioeconômicas e de acesso	Classe Social	Padrão de vida que melhor se encaixa	Classe baixa, Classe média baixa, Classe média, Classe média alta
	Situação de Moradia	Situação atual de moradia	Sozinho, Em Família, Em República, Divide Apartamento
	Lugar Calmo	Se possui lugar calmo para as atividades ou não	Sim, Não
	Situação de Emprego	Se está trabalhando	Trabalhando dirigindo-se ao trabalho, Trabalhando em Home Office, Dispensado das Atividades, Não trabalhando
	Filhos	Se possui filhos	Possuo filhos e teria as atividades EAD afetadas, Possuo filhos mas não seria afetado pelo EAD, Não possuo filhos
	Situação de internet	Se possui internet e qual o tipo de provedor	Banda larga fixa, Banda larga móvel, Ambas, Não possuo internet
	Possui computador/notebook	Se possui computador ou notebook	Sim, Não
	Possui Dispositivo Virtual	Se possui dispositivo virtual	Celular, Tablet, Outros
Informações sobre condições psicológicas	Situação do Dispositivo Virtual	Situação do dispositivo virtual/computador	Computador/Notebook pessoal e/ou Emprestado, Celular/Notebook pessoal e/ou Emprestado, Uso de dispositivo dentro/Emprestado do campus,
	Doença Psicológica	Se possui alguma doença psicológica que possa afetar o desempenho	Depressão, Ansiedade, Bipolaridade, Síndrome do Pânico, Não possuo doença deste tipo
Informações referentes a adaptabilidade e conhecimento do aluno sobre EAD	Nível de Estímulo Negativo do Estresse	Situações que posam desencadear estímulos negativos em estudantes	Necessidade de estudar várias matérias, Acúmulo de atividades, Dificuldade de absorção do conteúdo, Preocupação com o futuro, Não possuo estímulos
	Foco para EAD	Nível de foco para realização das atividades	Bom foco, Médio foco, Nenhum foco
	Dificuldade em Sanar Dúvidas via EAD	Nível de dificuldade em sanar dúvidas via EAD	Nenhuma dificuldade, Pouca dificuldade, 0 Muita dificuldade
	Ânimo para Atividades EAD	Possui ânimo para execução de atividades EAD	Muito positivo, Positivo, Negativo, Muito negativo
	Adaptabilidade às Ferramentas EAD	Nível de adaptabilidade às ferramentas EAD	Fácil, Média, Dificil adaptação
	Nível de Perfil do Aluno EAD	Características sobre o perfil do aluno EAD mais em comum com as que possui	Proativo, Organizado, Responsável, Disciplinado, Flexível, Possui Autonomia, Enxerga professores como orientadores
	Tecnologias EAD que conhecem	Algumas tecnologias que os entrevistados conhecem ou não	Hangout Meet Google Education, Microsoft Teams, Moodle, Skype, Slack, Nenhuma das alternativas
Informações referentes ao engajamento do aluno	Tecnologias Fundamentais Necessárias em Plataformas EAD	Opinião sobre quais das tecnologias listadas precisam ser fundamentais em plataformas EAD	Videoaulas, Áudio e Videoconferência, Chats, Fóruns, Bibliotecas Virtuais, Outros
	Uso do Software Educativo	Opinião sobre sugestão de aplicação de softwares educacionais para auxílio das atividades EAD	Sim, Talvez, Não
	Importância da Interação Presencial com Professores	Nível de importância da interação presencial entre aluno e professor	Muito importante, Importante, Pouco importante
Informações sobre situação atual do aluno em relação as aulas e a pandemia	Importância da Interação Presencial com Colegas	Nível de importância da interação presencial entre aluno e colegas de classe	Muito importante, Importante, Pouco importante
	Desvantagem pela suspensão das Aulas	Se foi prejudicado pela suspensão das aulas	Sim, Não
Informações referente a opinião dos alunos sobre aplicação do EAD	Grupo de risco Covid-19	Se está no grupo de risco da Covid-19	Sim, Não
	Optaria pelo Método EAD fora de pandemia	Se optaria pelo método EAD fora de pandemia	Sim, Não
	Optaria pelo Método EAD durante Pandemia	Se optaria pelo método EAD durante pandemia	Sim, Não

A base é formada, quase que em sua totalidade, por atributos nominais. Assim, fez-se necessário transformar os valores dos atributos para posterior aplicação dos algoritmos de aprendizado de máquina. Como houve respostas abertas, os valores foram codificados criando-se uma representação numérica para cada valor do atributo, na qual todas as possibilidades de respostas das questões foram identificadas e enumeradas. Após este processo, foi aplicado ao conjunto a normalização min-max (HAN, KAMBER, 2006) para normalizar os valores e submetê-los a execução dos algoritmos.

Os experimentos foram realizados no *software Weka* versão 3.8.4. O *Weka (Waikato Environment for Knowledge Analysis)* é um pacote composto por um conjunto de implementações de diversos algoritmos de Mineração de Dados (BOUCKAERT *et al.*, 2018). Os algoritmos trabalhados foram executados utilizando os parâmetros de configuração padrão do *Weka*. Após a etapa de Preparação dos Dados, estes foram submetidos a aplicação dos algoritmos de aprendizagem de máquina referente a fase de Modelagem.

Os algoritmos *Self-organizing maps* (SOM), e *k-means* (*k-means clustering*) foram usados para o experimento. Tais algoritmos foram escolhidos porque apresentam bons resultados em trabalhos na literatura. O algoritmo SOM foi usado para determinar automaticamente uma quantidade de grupos, sendo gerado 4 agrupamentos. Para o *k-means*, as escolhas dos valores de *k* consideraram a quantidade de agrupamentos gerados anteriormente pelo SOM, sendo estes *k*=3 e *k*=5, a fim de trabalhar em proporções próximas um do outro.

4. Resultados

Para representar os resultados dos experimentos utilizou-se a sigla *Ci-ALG* em que *i* corresponde a variação da quantidade de *cluster* gerados pelos algoritmos e ALG é o algoritmo de agrupamento utilizado. O algoritmo SOM gerou 4 grupos, logo têm-se as seguintes representações: C0-SOM, C1-SOM, C2-SOM e C3-SOM. Para o algoritmo *k-means*, têm-se: C0-k3, C1-k3 e C2-k3 usando *k*=3 e C0-k5, C1-k5, C2-k5, C3-k5 e C4-k5 para *k*=5. A Tabela 1 apresenta a quantidade de instâncias atribuídas, para cada grupo gerado, pelos algoritmos de agrupamento.

Tabela 1 – Resultado dos agrupamentos sendo os *clusters* e suas quantidade de instâncias

Algoritmos	SOM				<i>k</i> =3			<i>k</i> =5				
Grupos	C0-SOM	C1-SOM	C2-SOM	C3-SOM	C0-k3	C1-k3	C2-k3	C0-k5	C1-k5	C2-k5	C3-k5	C4-k5
Quantidade de Instâncias	138	150	93	102	125	154	204	91	67	119	165	41

A partir dos grupos gerados, foram analisados os respectivos valores de cada um dos atributos em cada grupo com o objetivo de identificar os valores mais predominantes nas características, i.e., os atributos da base, possibilitando realizar uma análise dos grupos e dos indivíduos.

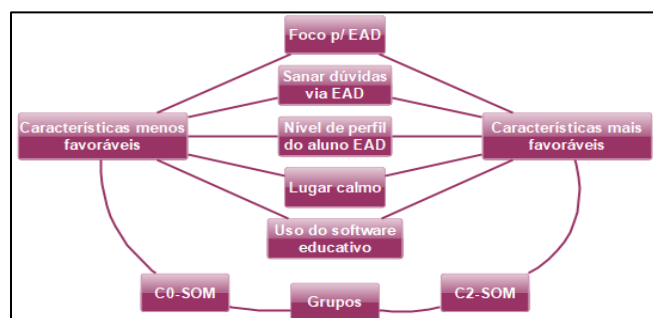
4.1 Discussão dos Resultados

Nesta seção são apresentados os resultados alcançados pelo SOM e *k-means*.

4.1.1. Resultados SOM

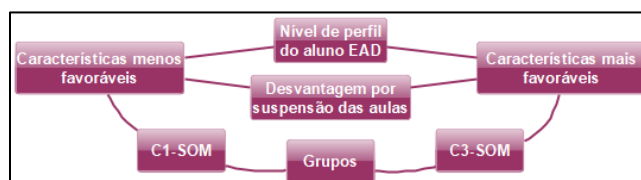
O algoritmo SOM gerou quatro agrupamentos. Os grupos C0-SOM e C2-SOM se mostraram contrários a aplicação EAD tanto durante a época de pandemia quanto fora dela. C1-SOM e C3-SOM se mostraram mais favoráveis, com C1-SOM a favor do EAD apenas durante a época de pandemia e o C3-SOM a favor em ambas as situações. A Figura 4 apresenta alguns dos atributos que dividiram os grupos C0-SOM e C2-SOM.

Figura 4 - Diferenças entre os grupos C0-SOM e C2-SOM



Observa-se que as características como ‘foco_para_ead’, ‘sanar_duvidas_ead’, ‘nivel_perfil_aluno_ead’, ‘lugar_calmo’ e ‘uso_software_educativo’, foram alguns dos divisores desses grupos, onde as pessoas inseridas no grupo C0-SOM mostraram índices maiores de dificuldade e as do C2-SOM com menores. A Figura 5 apresenta os atributos que dividiram os grupos C1-SOM e C3-SOM.

Figura 5 - Diferenças entre os grupos C1-SOM e C3-SOM

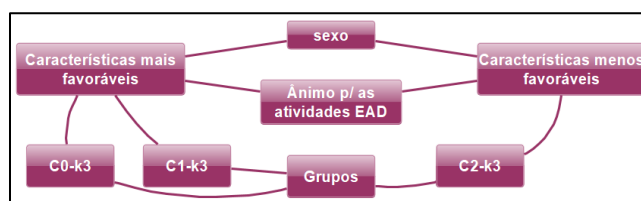


Nos agrupamentos C1-SOM e C3-SOM, as características ‘nivel_perfil_aluno_ead’ e ‘desvantagem_suspensao_aulas’ separaram os grupos. Indivíduos com mais características de adaptação compuseram o C3-SOM. O C1-SOM selecionou indivíduos que possuíram alguma desvantagem por conta da suspensão das aulas.

4.1.2. Resultados *k-means*

Para o *k-means* com $k=3$, os grupos C0-k3 e C1-k3 selecionaram indivíduos que optariam pelo EAD durante pandemia, mas não fora dela, e o C2-k3 selecionou indivíduos que não optariam em nenhuma das ocasiões. A Figura 6 apresenta os atributos que dividiram os três agrupamentos gerados.

Figura 6 - Diferenças entre os grupos C0-k3, C1-k3 e C2-k3



Analisando algumas características dos três agrupamentos o atributo ‘sexo’ e ‘animo_para_atividades_ead’ foram caracterizados como principais separadores entre os *clusters*. Em relação a C2-k3, características como ‘foco_para_ead’, ‘sanar_duvidas_ead’, ‘doenca_psicologica’, ‘nivel_perfil_aluno_ead’ e, se possui ‘lugar_calmo’ foram fatores negativos mais presentes nesse grupo.

Para $k=5$, os agrupamentos C0-k5 e C4-k5 foram compostos por indivíduos que optariam apenas pelo EAD durante pandemia. Nos grupos C1-k5 e C2-k5, os indivíduos não optariam pelo EAD em ambas as situações, e no grupo C3-k5, separados os indivíduos que optariam pelo EAD durante pandemia e fora dela. A Figura 7 apresenta os atributos que dividiram os grupos C0-k5 e C4-k5.

Figura 7 - Diferenças entre os grupos C0-k5 e C4-k5



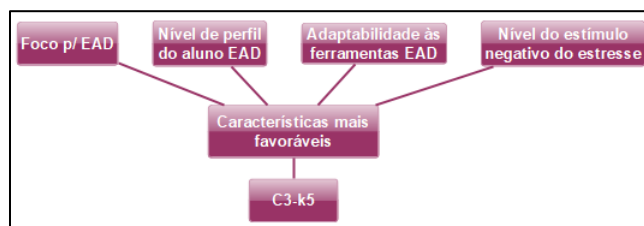
Entre estes grupos observou-se que os atributos ‘nivel_academico’, ‘situacao_emprego’ e ‘nivel_estimulo_negativo_estresse’ influenciaram a divisão dos indivíduos, sendo o grupo C0-k5 formado por pessoas a nível de graduação e menos adaptáveis à metodologia e o grupo C4-k5 indivíduos mais velhos, na maioria da pós-graduação e com características mais positivas em relação ao EAD. A Figura 8 apresenta os atributos que separaram os grupos C1-k5 e C2-k5.

Figura 8 - Diferenças entre os grupos C1-k5 e C2-k5



Nestes grupos poucas discrepâncias foram encontradas entre os indivíduos, pois possuem quase as mesmas características, os atributos ‘grupo_de_risco_covid19’ e ‘doença_psicologica’ foram os principais divisores entre esses agrupamentos. A Figura 9 denota as características para o grupo C3-k5.

Figura 9 – Características mais favoráveis do grupo C3-k5



Por fim, o grupo C3-k5 é composto por indivíduos a favor dos métodos EAD, as características ‘foco_para_ead’, ‘nivel_perfil_aluno_ead’, ‘adaptabilidade_ferramentas_ead’ e ‘nivel_estimulo_negativo_estresse’ foram fatores mais positivos em relação aos outros quatro agrupamentos.

4.2. Discussão Comparativa

A experimentação com diferentes algoritmos possibilitou gerar diversas saídas para comparação das informações descobertas. Os resultados classificaram indivíduos de acordo com características positivas e negativas em relação a fatores de adaptação, condições psicológicas e condições estruturais. Analisou-se que indivíduos não adeptos ao EAD, em sua maioria, possuíam maiores estímulos negativos do estresse, média adaptação às ferramentas, relataram dúvidas quanto a aplicação de softwares educativos para auxílio das atividades, mostraram-se mais dependentes por considerarem a interação presencial com os professores fundamental, além de apresentarem baixos níveis de perfil do aluno EAD. Características como doença psicológica e se possuíam lugar calmo para realizar as atividades também se mostraram importantes.

Entre os indivíduos adeptos ao EAD, estes exibiram maiores características de perfil do aluno EAD, fácil adaptação para as ferramentas, optam pelo uso do software educativo, relataram poucos estímulos negativos do estresse e mostram ser mais independentes, são mais autogerenciáveis. Em alguns casos a idade se mostrou relevante, classificando indivíduos mais jovens como menos adeptos, assim como indivíduos de maior grau acadêmico como mais adeptos à metodologia. Também, foi possível inferir que muitos alunos podem ainda não estar preparados para aplicação da metodologia EAD,

e apenas optariam por esta por conta de alguma desvantagem ocasionada pela suspensão das aulas presenciais. Por fim, foi possível realizar uma projeção de divisão dos grupos estipulada por meio dos resultados, projetando ao fim três grupos, sendo estes formados por pessoas que optariam pelo método EAD durante a pandemia e fora de pandemia, outro composto por pessoas que optariam pelo EAD apenas durante a pandemia e o último composto por pessoas que não optariam pelo EAD em ambas as situações.

5. Conclusões

As características descobertas na tentativa de modelar o perfil dos estudantes em relação ao EAD durante a época de pandemia, foram trabalhadas em dados acadêmicos e pessoais de alunos de diferentes universidades. Os resultados mostraram alunos em condições mais adversas de adaptação, econômica, psicológica ou de acesso com menor ânimo para desenvolver as atividades. Alunos com boas condições de acesso, infraestrutura, vínculo com a instituição de ensino e melhor adaptabilidade mostraram-se mais engajados com a metodologia EAD. Apesar da resposta positiva de diversos alunos, muitos escolheram em não optar pelo método EAD, mesmo apresentando bons níveis de adaptação, o que leva a uma busca por maiores informações sobre estes estudantes.

Como trabalhos futuros, é possível acrescentar questões em relação as atividades práticas e teóricas dos cursos, buscando entender o nível de ocorrências destas aulas e aplicar uma análise para seleção de características/atributos da base para eliminar variáveis que mostram pouca contribuição com a tarefa desejada.

6. Referências

BAKER, R. S. J. et al. Mineração de Dados Educacionais: Conceitos, Técnicas, Ferramentas e Aplicações. In: **JORNADA DE ATUALIZAÇÃO EM INFORMÁTICA NA EDUCAÇÃO**, 2012. v. 1, n. 1, p. 1 – 29.

BAZALDUA, D. A. L.; BAKER, R. S.; SAN PEDRO, M. O. Z. Comparing Expert and Metric-Based Assessments of Association Rule Interestingness. In: **INTERNATIONAL CONFERENCE ON EDUCATIONAL DATA MINING**, 7, 2014, London. Proceedings. London, jul. 2014, p. 44 – 51.

BOUCKAERT, R. R. et al. **Weka Manual for Version 3-8-3**. Sep 4, 2018. Disponível em: <<https://www.cs.waikato.ac.nz/ml/weka/>> Acesso em: 16 de jun de 2020.

CIEB. Nota Técnica nº 16 de 30 de outubro de 2019. **Inteligência Artificial na educação**. Disponível em: <cieb.net.br/wp-content/uploads/2019/11/CIEB_Nota_Tecnica16_nov_2019_digital.pdf>. Acesso em: 04 set. 2020.

DO NASCIMENTO, R. L. S.; DA CRUZ JUNIOR, G. G.; DE ARAÚJO FAGUNDES, R. A. Mineração de Dados Educacionais: Um estudo sobre indicadores da educação em bases de dados do INEP. **RENOTE - Revista Novas Tecnologias na Educação**, v. 16, n. 1, 2018.

FANG, Y. et al. Clustering the Learning Patterns of Adults with Low Literacy Skills Interacting with an Intelligent Tutoring System. In: **INTERNATIONAL CONFERENCE ON EDUCATIONAL DATA MINING**, 10, 2018, Buffalo. Proceedings. Buffalo, jul. 2018, p. 348 – 354.

FAVA, R. **Educação para o século 21: a era do indivíduo digital**. São Paulo: Saraiva, 2016.

- HAN, J; KAMBER, M. **Data Mining: Concepts and Techniques**. Elsevier, 2006.
- HIJAZI, S.T.; NAQVI, S. Factors Affecting Students' Performance: A Case of Private Colleges. **Bangladesh e-Journal of Sociology**, v. 3, n. 1, p. 1-10, jan. 2006.
- KAMPFF, A. J. C.; REATEGUI, E. B.; LIMA, J. V. Mineração de dados educacionais para a construção de alertas em ambientes virtuais de aprendizagem como apoio à prática docente. **RENOTE - Revista Novas Tecnologia na Educação**, v. 6, n. 2, 2008.
- KOHONEN, T. The self-organizing map. **Proceedings of the IEEE**, v. 78, n. 9, p. 1464-1480, 1990.
- LOPEZ, M. I. et al. Classification via Clustering for predicting final marks based on student participation in forums. In: **INTERNATIONAL CONFERENCE ON EDUCATIONAL DATA MINING**, 5, 2012. Pittsburgh. Proceedings. Chania, jun. 2012, p. 148 – 151.
- LORENZE, S.; HJULER, N.; ALSTRUP, S. Tracking Behavioral Patterns among Students in an Online Educational System. In: **INTERNATIONAL CONFERENCE ON EDUCATIONAL DATA MINING**, 11, 2018, Buffalo. Proceedings. Buffalo, jul. 2018, p. 280 – 285.
- LUCKIN, R.; HOLMES, W.; GRFFITHS, M.; FORCIER, L.B. (2016). **Intelligence Unleashed: An Argument for AI in Education**. London: Pearson. Disponível em: <static.googleusercontent.com/media/edu.google.com/pt-BR//pdfs/IntelligenceUnleashed-Publication.pdf>. Acesso em: 21 set de 2020.
- PIMENTEL, E. P.; DE FRANÇA, V. F.; OMAR, N. A identificação de grupos de aprendizes no ensino presencial utilizando técnicas de clusterização. In: **SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO**, 2003, Rio de Janeiro. Anais. Rio de Janeiro: Simpósio Brasileiro de Informática na Educação, 2003, p. 495-504.
- RAU, M. A.; MASON, B.; NOWAK, R. How to Model Implicit Knowledge? Similarity Learning Methods to Assess Perceptions of Visual Representations. In: **INTERNATIONAL CONFERENCE ON EDUCATIONAL DATA MINING**, 9, 2016, North Carolina, Proceedings. North Carolina, jun. 2016, p. 199 – 206.
- SHIH, B.; KOEDINGER, K. R.; SCHEINES, R. A Response Time Model for Bottom-Out Hints as Worked Examples. In: **INTERNATIONAL CONFERENCE ON EDUCATIONAL DATA MINING**, 1, 2008, Montreal. Proceedings. Montreal, jun. 2008, p. 117 - 126.
- VICARI, R. M. **Tendências em inteligência artificial na educação no período de 2017 a 2030: sumário executivo**. 2018. Disponível em: <<http://www2.fiescnet.com.br/web/uploads/recursos/d1dbf03635c1ad8ad3607190f17c9a19.pdf>>. Acesso em: 14 de jun de 2020.
- WIRTH, R.; HIPPEL, J. CRISP-DM: Towards a standard process model for data mining. In: **INTERNATIONAL CONFERENCE ON THE PRACTICAL APPLICATIONS OF KNOWLEDGE DISCOVERY AND DATA MINING**, 2000. Proceedings. London, 2000. p. 29-39.
- YANG, T. Y. et al. Active Learning for Student Affect Detection. In: **INTERNATIONAL CONFERENCE ON EDUCATIONAL DATA MINING**, 11, 2019, Montréal. Proceedings. Montréal, jul. 2019, p. 208 – 217.