



INTELIGÊNCIA ARTIFICIAL NA EDUCAÇÃO: aplicações do aprendizado de máquina para apoiar a aprendizagem adaptativa

Jean J. C. C. Leão¹
Abdiel B. Santos²
Thaiana M. Marques³
Eliane M. S. Santos⁴
Jeancarlo C. Leão⁵

Recebido em: 11/2020
Aprovado em: 01/2021

RESUMO

Este trabalho tem como objetivo demonstrar o uso de algoritmos de aprendizado de máquina no processo de ensino-aprendizagem. Fornecemos um conjunto de materiais (algoritmos e conjunto de dados) para possibilitar a reprodução dos nossos experimentos com as principais técnicas de aprendizado de máquina. Além disso, foram produzidos tutoriais sobre como esses materiais podem ser utilizados para resolver alguns problemas relacionados ao processo de ensino e aprendizado, assim como aprofundar no entendimento desses problemas. Este estudo possibilita melhorar a experiência de aprendizagem e apoiar a tomada de decisão sobre adaptações do conteúdo da aula ou das estratégias de ensino, fornecendo instrumentos eficientes e precisos para previsão de desempenho do aluno. Nossos experimentos foram executados em um conjunto de dados sintéticos de 395 alunos, contendo dentre outras propriedades, preferências por tipos de objetos de aprendizagem. Neste contexto, três tarefas de aprendizado de máquina foram consideradas: agrupamento, classificação e regressão linear. Nossos resultados revelaram meios para descobrir novos conhecimentos para satisfazer, prever e recomendar melhor as preferências específicas de cada aluno por tipos de objetos de aprendizagem.

Palavras-chave: Aprendizagem personalizada. Aprendizado de máquina. Agrupamento. Classificação. Regressão linear.

ARTIFICIAL INTELLIGENCE IN EDUCATION: machine learning applications to support adaptive learning

ABSTRACT

This work aims to demonstrate the use of machine learning algorithms in the teaching-learning process. We provide a set of materials (algorithms and dataset) to enable the reproduction of our experiments with the main machine learning techniques. In addition,

¹ Colégio Nazareth - Araçuaí / MG

² Instituto Federal do Norte de Minas Gerais- Araçuaí / MG

³ Instituto Federal do Norte de Minas Gerais- Araçuaí / MG

⁴ Instituto Federal do Norte de Minas Gerais- Araçuaí / MG

⁵ Instituto Federal do Norte de Minas Gerais- Araçuaí / MG



tutorials were produced on how these materials can be used to solve some problems related to the teaching and learning process, as well as to deepen the understanding of these problems. This study makes it possible to improve the learning experience and support decision making on adaptations of the content of the lesson or teaching strategies, providing efficient and accurate instruments for predicting student performance. Our experiments were carried out on a synthetic dataset of 395 students, containing, among other properties, preferences for types of learning objects. In this context, three machine learning tasks were considered: clustering, classification and linear regression. Our results revealed patterns of information that allow us to better satisfy, predict and recommend each student's specific preferences for types of learning objects.

Keywords: Personalized learning. Machine learning. Clustering. Classification. Linear regression.



1. INTRODUÇÃO

Diversas soluções baseadas em inteligência artificial (IA) possibilitam avanços em ambientes educacionais, pois podem informar melhor e mais rapidamente os alunos, professores e suas instituições e ajudá-los a prover benefícios na aprendizagem e no sucesso dos resultados educacionais. Por exemplo, com uso de algoritmos de IA, é possível prever, personalizar e recomendar tecnologias educacionais utilizadas em processos de ensino-aprendizagem e escolhas pedagógicas com base nas necessidades de cada aluno.

Assim, o uso de inteligência artificial e, mais especificamente, de algoritmos de aprendizado de máquina no processo de ensino-aprendizagem oportuniza ao educador descobrir mais rapidamente conhecimentos novos a partir de padrões ocultos nos dados brutos coletados no ambiente educacional. Portanto, pode ser aprimorada a percepção do docente sobre a sala de aula e sobre o comportamento dos alunos, além de possibilitar novos caminhos de aprendizagem, bem como refinar os métodos de aprendizagem a partir da tomada de decisões orientadas por dados (Romero, 2013; Lin, 2013).

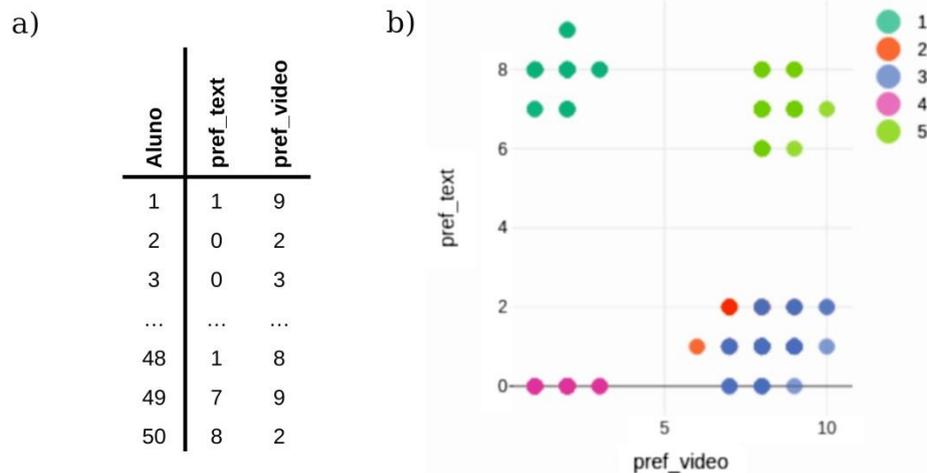
Ao estudante, tais algoritmos oportunizam melhorar a experiência de aprendizagem, enriquecendo os instrumentos de ensino e de avaliação com características mais atrativas. Nesse contexto, as interações comportamentais entre professores e alunos podem ser apoiadas por objetos de aprendizagem com interfaces mais personalizadas, inteligentes, interativas, dinâmicas e que favorecem a colaboração (Castro, 2016; Ciolacu, 2017). Assim, é possível tornar as tecnologias educacionais tão motivadoras quanto aquelas usadas para outros propósitos e que geralmente desviam a atenção do aluno, como os jogos eletrônicos e as redes sociais online. Dentre as tarefas realizadas por esses algoritmos estão as de agrupamento, classificação e regressão linear.

As tarefas de agrupamento (*clustering*) ou de detecção de comunidades (*community detection* ou *graph clustering*) consistem em particionar o conjunto de entidades presentes nos dados em grupos (ou comunidades), seguindo alguma função de dissimilaridade (Leão et al., 2018). Assim, um algoritmo de agrupamento procura atribuir a um mesmo grupo, entidades muito semelhantes, enquanto em grupos distintos, estão entidades tão diferentes quanto possível (Zaki e Meira, 2020).

Por meio da tarefa de agrupamento, é possível identificar, por exemplo, grupos de alunos que possuem semelhanças em suas preferências por objetos de aprendizagem. Para isso, primeiro é preciso representar a preferência por um tipo de objeto em um valor

numérico, por exemplo, a nota atribuída pelo aluno para o nível de satisfação que o tipo de objeto tem para seu aprendizado. A seguir, a Figura 1 ilustra a aplicação da tarefa de agrupamento.

Figura 1- Exemplo de aplicação do agrupamento de dados



Fonte: próprio autor

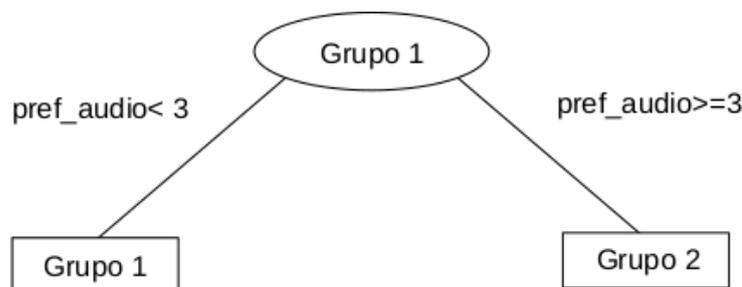
Observa-se primeiramente na Figura 1 (a), o conjunto de dados, apresentado na tabela que relaciona os estudantes às suas preferências por diferentes tipos de objetos de aprendizagem como aqueles baseados em texto ou vídeo. Em seguida, na Figura 1 (b), observa-se a representação gráfica dos dados, a partir da aplicação da tarefa de agrupamento, onde cada ponto representa um aluno e cada cor, o grupo (*cluster*) em que cada aluno foi associado, conforme preferências: grupo 1, alunos que preferem objetos baseados em texto; grupo 2 e 5, em vídeo; grupo 3, preferem ambos e; grupo 4, alunos que têm menor preferência por ambos tipos de objetos.

Assim, pode ser fornecido a cada grupo de alunos, os objetos que mais satisfaçam suas preferências, por exemplo: aos alunos do grupo 1, podem ser fornecidas mais atividades de leitura de texto; aos dos grupos 2 e 5, mais atividades com vídeos; aos do grupo 3, pelo menos um desses dois tipos de objetos de aprendizagem e; aos alunos do grupo 4, será preciso fornecer algum outro tipo de objeto de aprendizagem, diferente de vídeo e texto.

Quanto à tarefa de classificação, para RAUDYS (2012), a mesma busca solucionar o problema de identificar a que classe uma entidade pertence dentre um conjunto de possibilidades, a partir do histórico de dados obtidos, por exemplo, de um ambiente educacional. Assim, a tarefa de classificação, conforme ilustrado na Figura 2, pode ser

aplicada sobre os dados de preferências por objetos de aprendizagem, incluindo as informações descobertas pela técnica de agrupamento.

Figura 2 - Exemplo de classificação



Fonte: próprio autor

Para esta aplicação, primeiro devem ser identificados os grupos de alunos com preferências similares por tipos de objetos de aprendizagem (utilizando a técnica de agrupamento). Em seguida, cada grupo deve ser associado a uma classe ou estratégia de aprendizagem que permita extrair o melhor aproveitamento dos alunos considerando o estilo de aprendizagem⁶ do grupo a que pertencem.

A partir desses dados, o algoritmo gera um modelo de classificação dos estilos de aprendizagem dos alunos a partir das suas preferências por objetos de aprendizagem. Então, esse modelo pode ser utilizado sobre as preferências de um estudante, para sugerir automaticamente (classificar) o seu estilo de aprendizagem, quando esse estilo não for declarado previamente.

A tarefa de regressão, por sua vez, pode ser entendida segundo Menezes et al. (2018) como um modelo de equação linear que descreve a relação entre duas ou mais variáveis. Um exemplo de utilização desta tarefa é quando pretende-se prever a satisfação de uma proposta de personalização de aprendizagem.

Para isso, utiliza-se de um modelo de regressão, conforme Figura 3, para estimar o desempenho dos alunos em função da quantidade de objetos de aprendizagem disponibilizados para eles. Para gerar um modelo como esse é preciso considerar os valores históricos das quantidades de objetos utilizados em um determinado período e o desempenho obtido por cada aluno ao utilizar esses objetos. Note que o experimento deve ser conduzido

⁶ São exemplos de estilos de aprendizagem: o visual (que estimulam a visão); o auditivo (ouvindo uma explicação); o artístico (permitindo a produção criativa); o cinestésico (em atividades táteis e com movimentos corporais).

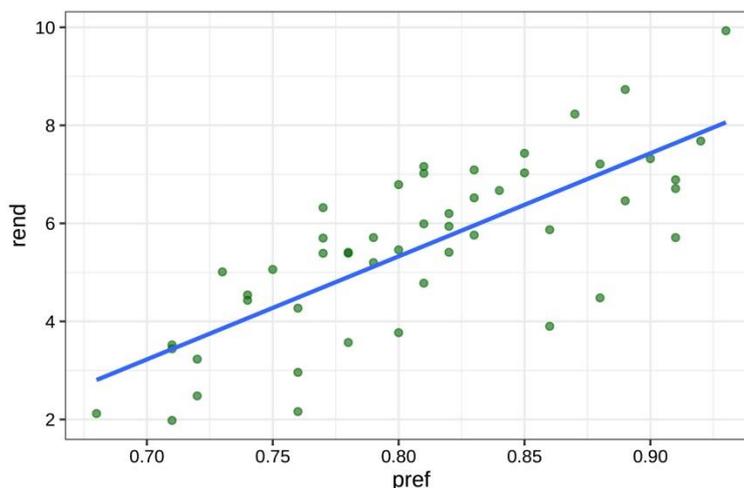
considerando os estilos de aprendizagem. Por isso, deve ser analisado individualmente para cada grupo de alunos com preferências similares por objetos de aprendizagem.

Figura 3 - Utilização da regressão linear para estimar o rendimento a partir da satisfação das preferências por tipos de objetos de aprendizagem.

a)

Aluno	pref	rend
1	0.7	3.2
2	0.9	7.4
3	0.8	5.8
...
48	0.8	6.2
49	0.8	7.0
50	0.8	7.1

b)



Fonte: próprio autor

Vale destacar que os exemplos descritos anteriormente ilustram algumas aplicações de algoritmos sobre dados hipotéticos em razão das particularidades deste tipo de informação a cada contexto educacional. Além disso, diante da diversidade de tarefas que podem ser aplicadas, surgem dois principais desafios.

Primeiro, é necessário conhecer cada uma dessas tarefas de modo que possa escolher a mais adequada para resolver um tipo de problema específico. Em seguida, é preciso modelar um problema educacional real para ser solucionado pela tarefa escolhida. Note que, podem existir outros aspectos que devem ser considerados na aplicação de uma tarefa de IA, como a escolha do melhor algoritmo dentre os múltiplos que se diferenciam em eficiência e eficácia (Leão et al., 2020). Apesar disso, deixamos para aprofundar sobre esses aspectos nos materiais complementares e na continuidade deste trabalho, por exigir dos profissionais da educação conhecimentos muito específicos da área da computação.

Assim, o objetivo deste trabalho é relevar aspectos gerais da aplicação de algumas das principais técnicas de aprendizado de máquina para a descoberta de novos conhecimentos no ambiente educacional. Especificamente, objetiva-se apresentar essas técnicas e favorecer o seu uso para apoiar o processo de ensino e aprendizagem.



As principais contribuições deste trabalho são: um conjunto de materiais (algoritmos e dados) para demonstração de uso das principais técnicas de aprendizado de máquina e reprodução dos nossos experimentos; instruções sobre como esses materiais podem ser utilizados para resolver alguns desafios importantes relacionados ao processo de ensino e aprendizagem, assim como buscar o entendimento desses desafios.

Ademais, o estudo sobre a aplicação de algoritmos de inteligência artificial na educação se justifica por possibilitar melhorar a experiência de aprendizagem do estudante e apoiar a tomada de decisão do professor sobre adaptações do conteúdo da aula ou das estratégias de ensino e por fornecer instrumentos eficientes e precisos para previsão de desempenho.

2. TRABALHOS RELACIONADOS

Diversas pesquisas que aplicam a inteligência artificial na educação foram conduzidas com diferentes abordagens e técnicas como o agrupamento (Ciolacu, 2017) e a classificação (Lin et al., 2013). Esses estudos buscam, através da IA, melhorar a formação de conteúdo (Adamu, 2018), sistematizar o aprendizado adaptável (Lin et al., 2013) ou descobrir conhecimento sobre o desempenho dos alunos (Ciolacu, 2017). Contudo, não observamos nesses estudos o foco em facilitar para que o usuário comum verifique e compreenda os seus resultados produzidos e para que possa replicar seus experimentos. Também carecem de instruções e materiais adequados para que esse tipo de usuário possa adaptar as aplicações propostas à natureza dinâmica do processo de ensino e aprendizado. Em geral, esses trabalhos são mais direcionados a verificar o desempenho das técnicas, deixando a facilitação de uso em segundo plano. Note que nem sempre a técnica mais eficiente será a mais fácil de ser compreendida e utilizada por profissionais de outras áreas.

3. METODOLOGIA

Para aplicar aprendizado de máquina em contextos de ensino e aprendizagem, serão utilizados algoritmos que executam diferentes tarefas e sobre um conjunto de dados sintéticos, ou seja, dados gerados a partir de modelos computacionais. Assim, as tarefas e algoritmos correspondentes selecionados para este estudo são: agrupamento com K-Means; classificação com árvore de decisão e regressão linear com a função dos mínimos quadrados.

O K-Means é um dos mais simples e populares algoritmos de agrupamento baseados no paradigma de representação. Por isso, esse algoritmo foi escolhido para uso neste trabalho,



considerando o propósito de iniciar educadores de diferentes áreas na utilização de aprendizado de máquina. A intuição do algoritmo K-Means se baseia em primeiro, escolher aleatoriamente uma posição inicial para os k centróides (cada um representando um grupo distinto) e a partir daí, iterar a associação de cada ponto ao grupo que possui o centróide mais próximo e mover o centróide para o ponto médio dos membros do seu grupo.

Essas iterações são feitas até a posição de todos os centróides convergirem (não movem mais ou atingiram uma quantidade de movimentos pré estabelecida). Mais formalmente, dado um conjunto de dados D com n pontos p_i (que representam entidades como alunos) em um espaço d -dimensional, e dado o número k de grupos desejados, o objetivo desse algoritmo é particionar o conjunto de dados em k grupos (agrupamento) denotados como $C = \{C_1, C_2, \dots, C_k\}$. Além disso, em cada grupo C_i existe o centróide μ_i , um ponto médio representativo de todos os pontos do grupo, ou seja, $\mu_i = 1/n_i \sum_{p_j \in C_i} p_j$ onde $n_i = |C_i|$, corresponde ao número de pontos no grupo C_i .

3.1. CONJUNTO DE DADOS

O conjunto de dados utilizado neste trabalho constitui registros que representam propriedades de 396 alunos, geradas por modelos computacionais (dados sintéticos). Essas propriedades são descritas na Tabela 1, e foram geradas por dois modelos matemáticos distintos.

O primeiro modelo (rotulado na Tabela 1 como RL) é uma combinação linear aleatória esparsa de propriedades, com ruído. Essas propriedades simulam o desempenho de alunos em função da qualidade atribuída por eles aos objetos de aprendizagem utilizados em um período de um curso.

O segundo modelo (rotulado na Tabela 1 como CL) corresponde a 10 atributos com valores que variam de 1 a 10 representando a preferência dos alunos sobre 10 tipos de objetos de aprendizagem.

Tabela 1 – Propriedades do conjunto de dados sintéticos

Rótulo	Descrição	Intervalo de valores	Modelo
Student	Identificador do estudante	1 a 395	NA
pref_audio	Preferência por objeto baseado em áudio	0 a 10	CL
pref_image	Preferência por imagens	0 a 10	CL



pref_infographic	Preferência por gráficos, diagramas	0 a 10	CL
pref_games	Preferência por atividades baseadas em jogos	0 a 10	CL
pref_webpgs	Preferência por páginas web	0 a 10	CL
pref_table	Preferência por tabelas	0 a 10	CL
pref_text	Preferência por leitura de textos	0 a 10	CL
pref_tutorial	Preferência por tutoriais de ensino	0 a 10	CL
pref_video	Preferência em assistir vídeos	0 a 10	CL
pref_forum	Preferência em participação em fóruns	0 a 10	CL
IRA	índice de rendimento acadêmico	0 a 1	RL
pref_avg	Estimativa de preferência	0 a 10	RL

Fonte: próprio autor

Note que decidiu-se utilizar dados sintéticos para essas propriedades em razão da dificuldade de coletar ou encontrar disponível todas elas. Além disso, para a proposta deste trabalho, a interpretação dos valores contidos nesses dados não é relevante, pois não influencia na obtenção e na qualidade dos resultados. A experimentação fez uso dos dados sintéticos, modelados considerando os seguintes cenários hipotéticos de coleta de dados:

3.1.1. Cenário 1: preferências por tipos de objetos de aprendizagem

1. Oferecer 10 diferentes tipos de objetos de aprendizagem (ex: imagem, infográfico, jogos, web pages, tabela, texto, tutorial, vídeo e fórum).
2. Pedir a cada aluno para atribuir uma nota de 0 a 10 para cada tipo de objeto de aprendizagem, conforme sua preferência.

3.1.2. Cenário 2: satisfação por uso de objetos de aprendizagem e rendimento do aprendizado

1. Planejar para uma aula, uma atividade que faça uso de diferentes tipos de objetos de aprendizagem.
2. Ofereça a cada aluno, um objeto de aprendizagem diferente, selecionado aleatoriamente, dentre os disponíveis.
3. Após a interação do aluno com o objeto, avalie seu rendimento no entendimento do conteúdo da aula. Repita até o passo 3 por 10 aulas.



4. Analise os objetos que foram oferecidos ao longo das 10 aulas e calcule o percentual de satisfação de cada um deles, para cada aluno. Para isso, considere a nota atribuída por cada aluno no experimento do Cenário 1 de cada objeto que ele utilizou no Cenário 2. Considere também, a nota correspondente que o aluno obteve em cada atividade (rendimento).

Note que, estes cenários podem ser adaptados para atender outros objetivos do processo de ensino e aprendizagem, ou para fazer uso de uma metodologia diferente de coleta de dados. Além disso, é importante lembrar que, em um cenário real, os experimentos precisam estar autorizados por um comitê de ética em pesquisa.

3.2. ALGORITMOS DE APRENDIZADO DE MÁQUINA

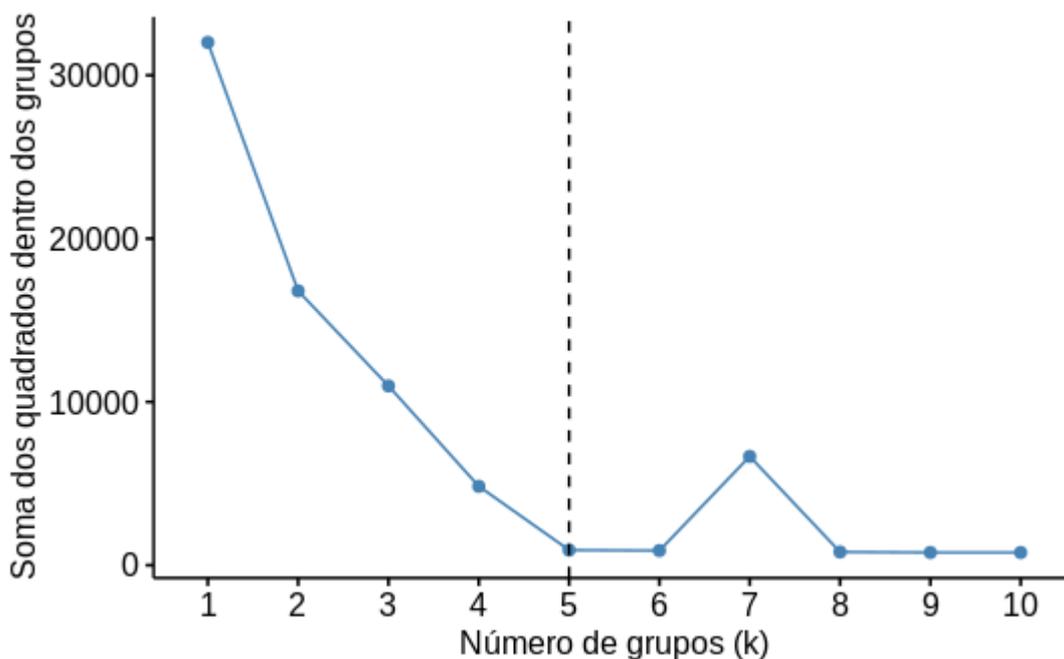
Os algoritmos utilizados neste trabalho foram selecionados dentre os mais populares para cada uma das três tarefas consideradas. Constituem algoritmos de aprendizado de máquina que podem ter sua implementação encontrada facilmente em diferentes linguagens de programação ou em ferramentas de mineração de dados como o Weka, Octave, Neural Designer, dentre outros. A seguir, descrevemos cada um desses algoritmos em suas respectivas tarefas.

3.2.1. Agrupamento

Na escolha do algoritmo para agrupamento das preferências por objetos de aprendizagem foi considerada a possibilidade de formar grupos de tamanhos aproximados, como normalmente ocorre em uma sala de aula. Assim, torna-se possível, além de identificar preferências similares para recomendação de tipos de objetos, ainda possibilita ter uma referência para sugerir membros de grupos em atividades colaborativas orientadas a diferentes atividades, como em metodologias como a de rotação por estações (Valente, 2018).

Para a escolha da quantidade de grupos de alunos, foi utilizado o método do cotovelo. O método do cotovelo permite buscar o melhor número de grupos para um agrupamento que é aquele que possui menor valor possível para a soma dos quadrados dentro dos grupos (do inglês, *within-clusters sum-of-squares* – *wcss*). A Figura 4 ilustra os diferentes valores de *wcss*, para quantidades de grupos variando de 1 a 10.

Figura 4 - Análise de agrupamento.

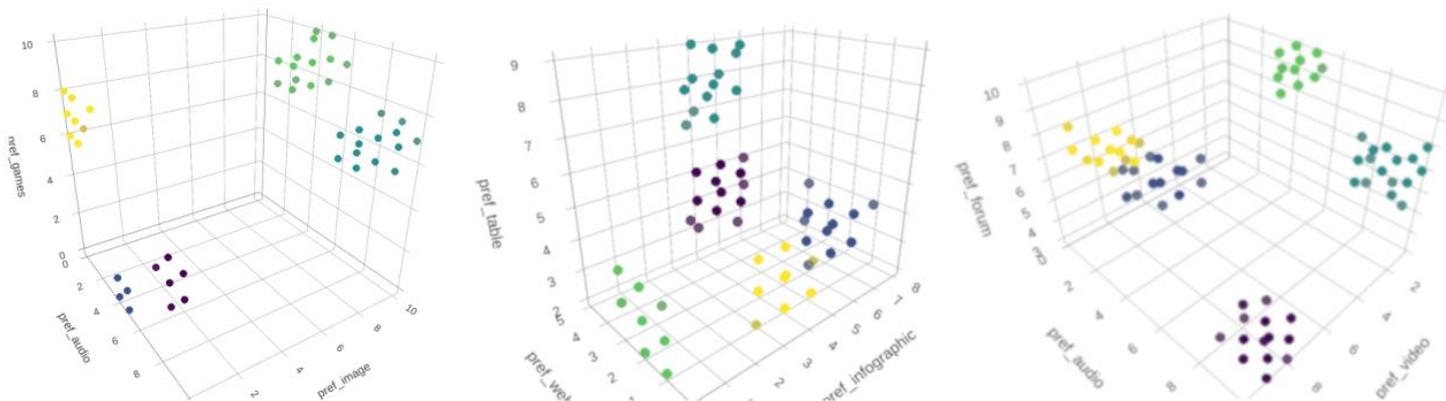


Fonte: próprio autor

Ao analisar os agrupamentos sobre o conjunto de dados sintéticos apresentados na Figura 4, observa-se que aqueles que possuem o menor valor possível para w_{css} possuem 5, 6, 8, 9 ou 10 grupos. Dentre esses, foi escolhido o menor número de agrupamentos, ou seja, com 5 grupos. Isso significa que cinco perfis distintos poderão ser identificados com base em respostas dos alunos sobre suas preferências por tipos de objetos de aprendizagem.

Apesar da análise de agrupamento considerar 10 diferentes tipos de objetos de aprendizagem, cada um representando uma dimensão, apenas poucas dimensões podem ser facilmente visualizadas simultaneamente devido às limitações da percepção humana. Por outro lado, os exemplos da Figura 5 que ilustram diversas perspectivas dos grupos em 3 dimensões, permitem perceber que esses cinco grupos são muito bem definidos, o que corrobora o baixo valor da w_{css} . Na prática, isso significa que é importante personalizar os tipos de objetos de aprendizagem para alunos que apresentam perfis como esses dos nossos dados sintéticos, pois eles possuem preferências diversas, mas que podem ser atendidas de forma otimizada através do uso da análise de agrupamentos.

Figura 5 - Visualização tridimensional do agrupamento de preferências, sob três das possíveis combinações de tipos de objetos de aprendizagem.



Fonte: próprio autor

Dessa forma, também é possível fazer essa análise manualmente pelo professor, usando algumas poucas horas para trabalhar sobre os dados informados pelos alunos para obter informações adicionais que contribuem com sua aula. Além disso, observe que as características de dados sintéticos como os que utilizamos facilitam a interpretação de sua visualização, o que é favorável ao propósito de demonstração deste trabalho. Por fim, um algoritmo de aprendizado de máquina pode ser utilizado sobre um conjunto de dados com muito mais registros de alunos, mais dimensões e com valores não tão bem comportados como esses que utilizamos. Ainda assim, é possível obter estatísticas e extrair em poucos segundos agrupamentos de boa qualidade e suas visualizações, o que seria inviável de ser feito em pouco tempo sem o uso de técnicas adequadas.

3.2.2. Classificação

Para a tarefa de classificação de dados, foi escolhida a técnica de árvore de decisão (*decision tree*), que é considerada como uma das mais intuitivas e de alta interpretabilidade. Essa técnica pode ser definida como um nodo de decisão que contém um teste, uma pergunta, sobre cada atributo de dados. Para cada resultado deste teste existe uma ligação para determinado conjunto de registros de dados. Cada conjunto tem uma estrutura de teste até o resultado final, como na Figura 2, onde o algoritmo classifica os alunos em três grupos com base no teste verificando a nota de cada um. O método utilizado para gerar a árvore foi o particionamento recursivo (BREIMAN et al., 1984).

3.2.3. Regressão Linear

A regressão linear é um modelo que permite estimar o valor de uma variável em um hiperplano p -dimensional referente às variáveis explicativas (Menezes et al., 2018). Mais formalmente temos:

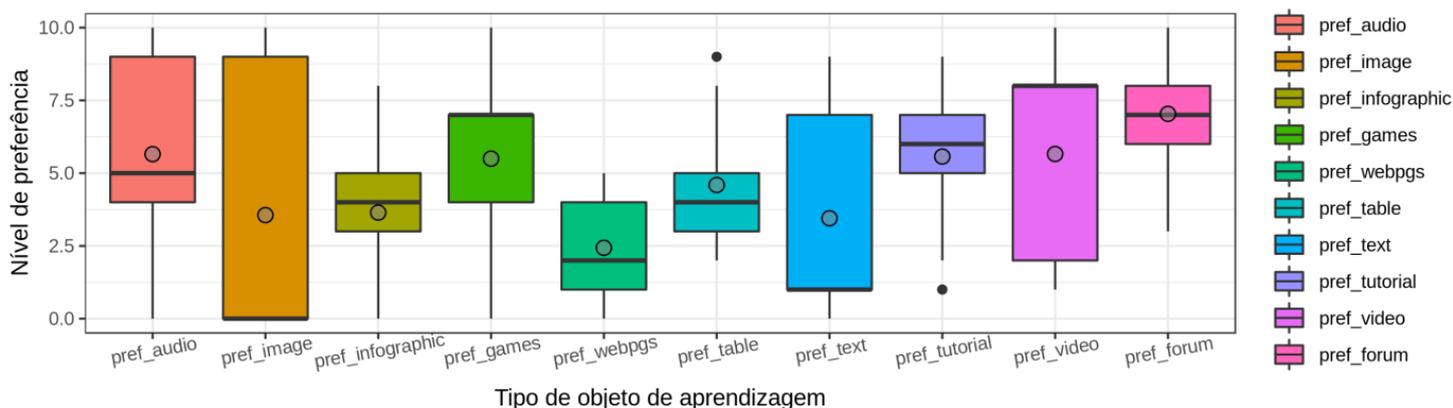
$$Q_w = \sum_{i=1}^n w_i (Y_i^* - \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip})^2$$

onde Y é o componente resposta, $\beta_1, \beta_2, \dots, \beta_p$ são os estimadores que minimizam a função de mínimos quadrados ponderados (MQP) e w_i é o desvio ponderado. Este método é interessante de ser aplicado quando é verificado um alto grau de relacionamento entre as variáveis explicativas. Para isso, deve ser feito um estudo de correlação.

4. RESULTADOS

A seguir, são apresentados os resultados da execução de cada uma das técnicas de aprendizado de máquina sobre o conjunto de dados sintéticos de alunos. Além disso, esse conjunto de dados foi disponibilizado publicamente⁷. O sumário das preferências contidas nesse conjunto de dados é apresentado na Figura 6.

Figura 6 - Estatística geral das preferências por tipo de objetos de aprendizagem



Fonte: próprio autor

Note que as estatísticas gerais apresentadas na Figura 6 complementam a precisão da percepção imediata do professor, mas dificilmente trariam muitas informações novas. Por exemplo, elas não informam qual o tipo de objeto de aprendizagem que os alunos mais

⁷ Dados, código e mais instruções para utilização dos experimentos exemplificados neste trabalho estão disponíveis em repositório público: <<https://jcloud.net.br/url/IAEduGit2020/>>.



gostam ou o que menos gostam e quais estudantes possuem o mesmo perfil de preferências. Vale ressaltar que seria complexo associar os alunos a grupos de tamanhos próximos com base em suas preferências sem o apoio de uma técnica algorítmica adequada para identificar os grupos e permitir uma análise individual, como descrito a seguir.

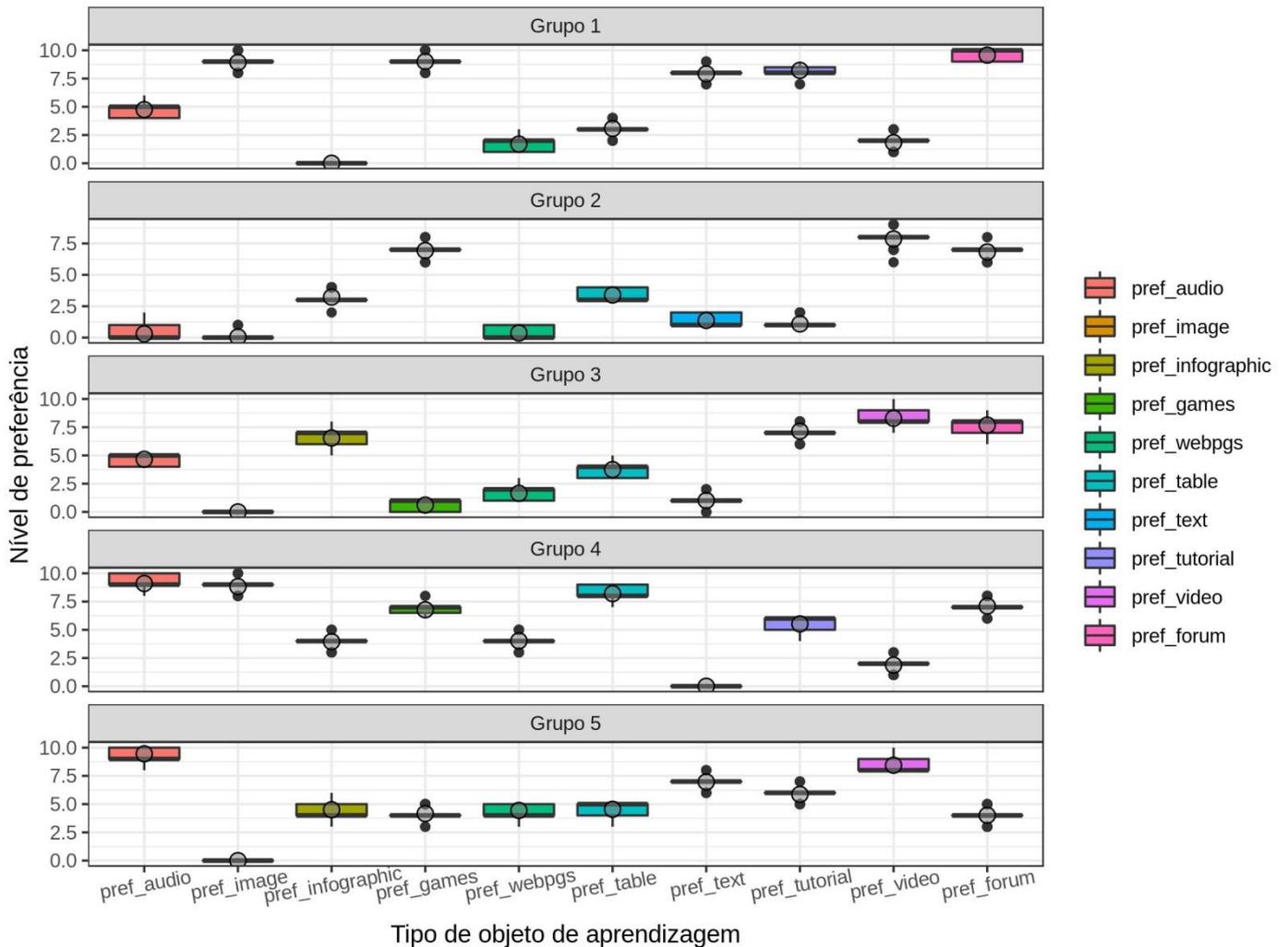
Para responder sobre quais os principais perfis de preferência existem nos dados, cada grupo de alunos foi analisado individualmente. Assim, além das estatísticas gerais deste conjunto de dados, são descritas a seguir as preferências compartilhadas por alunos que pertencem a um mesmo grupo. Além disso, a partir da análise de agrupamento foi possível verificar quais os objetos prevalecem nas preferências e quais os que seriam fundamentais para satisfazer os alunos. Foi observado que pelo menos um dos objetos vídeo, áudio ou fórum, são preferidos por todos os grupos, pois em todos os grupos um ou mais desses objetos receberam nota acima de 8 em relação à preferência. Para alunos pertencentes a grupos com essas descrições, pode ser direcionado conteúdo em mídias mais apropriadas assim como atividades colaborativas que envolvam o uso dessas mídias.

Uma situação natural na análise da preferência sobre objetos de aprendizagem é a de oferecer aos alunos aquele objeto que possui a maior estimativa de preferência das notas na consulta de opinião desses alunos. Isso pode ser facilmente obtido na estatística geral ao analisar os maiores valores de média (círculo no boxplot da Figura 6) e de mediana. Assim, conforme a Figura 6, os objetos dos tipos *vídeo* e *games* estão dentre os preferidos.

Contudo, essa análise geral e prematura deixou passar despercebido algo importante que pode ser verificado na análise de agrupamento: parte considerável dos alunos tem maior rejeição justamente a objetos desse tipo, apesar de estarem com maiores médias gerais. Isso pode ser verificado ao observar individualmente cada grupo, em especial, analisando na Figura 7 a menor preferência indicada nos grupos 1 e 4 por objetos do tipo *vídeo* e no grupo 3 por objetos do tipo *games*⁸. Casos como esses requerem a personalização das estratégias de aprendizado para maximizar a satisfação do aprendizado. Note que essa análise também permite identificar objetos pontuados por todos os grupos com notas razoáveis, como é o caso do objeto do tipo *fórum*.

⁸ Note que o aspecto importante a ser considerado neste exemplo é a vulnerabilidade da estimativa geral que pode ocorrer em dados simulados como também em dados reais.

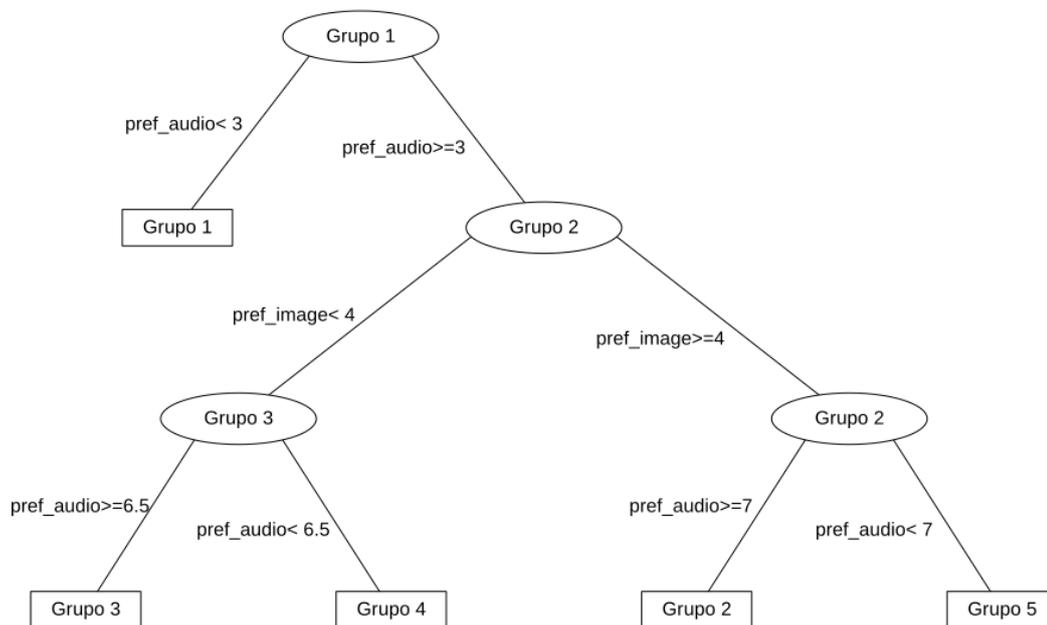
Figura 7 - Estatística das preferências por tipo de objetos de aprendizagem de cada grupo



Fonte: próprio autor

Na Figura 8 é possível observar a árvore de decisão obtida a partir dos dados de preferência de alunos combinados com os rótulos do grupo ao qual foram associados pela técnica de agrupamento. Esse modelo permite que novos alunos sejam classificados e em seguida associados a um grupo sem precisar executar a técnica de agrupamento sobre os dados de todos os demais alunos novamente. Além do custo computacional reduzido, o uso deste modelo permite entender como as propriedades do aluno foram consideradas para direcioná-lo para seu grupo.

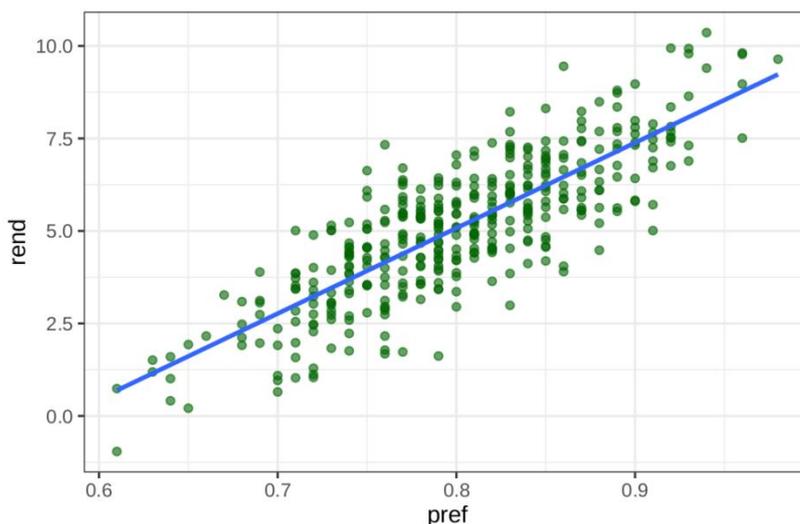
Figura 8 - Árvore de Decisão gerada pelo método de particionamento recursivo



Fonte: próprio autor

A Figura 9, ilustra a aplicação da regressão linear sobre os dados sintéticos de rendimento dos alunos em uma disciplina em função da satisfação de suas preferências pelos objetos de aprendizagem utilizados na disciplina.

Figura 9 - Modelo de regressão: rendimento e preferências



Fonte: próprio autor

Assim, na sala de aula, tal aplicação pode ajudar o professor a saber se o aprendizado do aluno está dentro do esperado (estimado pelo modelo) considerando a sua preferência por tipos de objetos de aprendizagem.



Note que, ao trabalhar com dados reais, além de permitir verificações como as apresentadas anteriormente, ainda é possível fazer a interpretação dos valores contidos nesses dados e obter resultados adicionais, como padrões de preferências peculiares. Além disso, torna-se possível que as informações descobertas com apoio de algoritmos sejam aplicáveis a um contexto específico de ensino-aprendizagem.

5. CONCLUSÃO

O presente trabalho objetivou demonstrar uso de algoritmos de aprendizado de máquina no processo de ensino-aprendizagem através de experimentos executados sobre um conjunto de dados sintéticos e de forma a favorecer a reprodução desses experimentos. Nossos resultados revelaram padrões de informações que seriam inviáveis de serem obtidas rapidamente sem o apoio de técnicas de aprendizado de máquina. Assim, foi possível concluir instruções acessíveis e materiais para reprodução dos experimentos com as técnicas de inteligência artificial consideradas, além de outras.

Uma limitação dos modelos gerados em nosso trabalho é não objetivar apresentar o comportamento de alunos de um ambiente real. Apesar disso, esses dados atenderam satisfatoriamente o objeto do trabalho e essa limitação pode ser facilmente superada reproduzindo os nossos experimentos sobre dados coletados de um ambiente educacional real, preferencialmente o mesmo onde serão utilizados. Outra limitação do trabalho surge da natureza dinâmica do ambiente educacional, que pode requerer, por exemplo, habilidades adicionais para minerar os seus dados que vão além do entendimento das técnicas apresentadas. Em trabalhos futuros, pretendemos executar novas tarefas de aprendizagem de máquina, assim como demonstrar outras soluções baseadas em inteligência artificial para solucionar outros tipos de problemas no processo de ensino e aprendizagem.

REFERÊNCIAS

- BAKER, R. S. J. **Data Mining for Education**. International Encyclopedia of Education, v. 7, n. 3, p. 112 – 118, 2010.
- BREIMAN, L. et al. **Classification and Regression Trees**. [s.l.] Wadsworth International Group, 1984.
- CASTRO, Juscilde Braga de. Construção do conceito de covariação por estudantes do Ensino Fundamental em ambientes de múltiplas representações com suporte das tecnologias digitais. 2016. Disponível em: <http://www.repositorio.ufc.br/handle/riufc/15908>. Acesso: 18 de janeiro de 2021.



CIOLACU, M. et al. **Education 4.0 - Fostering student's performance with machine learning methods**. 2017 IEEE 23rd International Symposium for Design and Technology in Electronic Packaging (SIITME). Anais...2017

CORTEZ, P.; SILVA, A. **Using Data Mining to Predict Secondary School Student Performance**. In A. Brito and J. Teixeira Eds., Proceedings of 5th FUTURE BUSINESS TECHNOLOGY CONFERENCE (FUBUTEC 2008) pp. 5-12, Porto, Portugal, April, 2008, EUROSIS, ISBN 978-9077381-39-7.

LEÃO, J. C. et al. **Who is really in my social circle?** Mining social relationships to improve detection of real communities. Journal of Internet Services and Applications, v. 9, n. 1, p. 20, 2018. Disponível em: <http://jcloud.net.br/url/JISA19/>. Acesso em: 01 de outubro de 2020.

LEÃO, J. C.; LAENDER, A. H. F.; VAZ DE MELO, P. O. S. **Overcoming Bias in Community Detection Evaluation**. Journal of Information and Data Management, 2020. Disponível em: <http://jcloud.net.br/url/LLD20JIDM/>. Acesso em: 05 de Janeiro de 2021.

LIN, C. F. et al. **Data mining for providing a personalized learning path in creativity: An application of decision trees**. Computers & Education, v. 68, p. 199–210, 2013.

MENEZES, L. S. et al. **Um Arcabouço para Integração de Microdados da Pesquisa Nacional por Amostra de Domicílios: Aplicação na Análise dos Fatores de Influência sobre a Renda do Trabalho na Região Sul do Brasil**. Anais dos Simpósios de Informática do IFNMG-Campus Januária, v. 10, p. 6, 2018. Disponível em: <https://jcloud.net.br/url/Lara2018PNADXSimposioJanuaria/>. Acesso em: 01 de outubro de 2020.

RAGHUPATHI, Wullianallur; RAGHUPATHI, Viju. **Big data analytics in healthcare: promise and potential**. Health information science and systems, v. 2, n. 1, p. 3, 2014.

RAUDYS, Sarunas. **Statistical and Neural Classifiers: An integrated approach to design**. Springer Science & Business Media, p.2, 2012.

ROLL, I.; WYLIE, R.; FORD, H. **Evolution and revolution in artificial intelligence in education**. International Journal of Artificial Intelligence in Education, v. 26, n. 2, p. 582–599, 2016.

ROMERO, C.; VENTURA, S. Data mining in education. WIREs Data Mining and Knowledge Discovery, v. 3, n. 1, p. 12–27, 2013.

SYAKUR, M. A. et al. Integration k-means clustering method and elbow method for identification of the best customer profile cluster. In: **IOP Conference Series: Materials Science and Engineering**. IOP Publishing, 2018. p. 012017.

VALENTE, J. A. **A sala de aula invertida e a possibilidade do ensino personalizado: uma experiência com a graduação em midialogia. Metodologias ativas para uma educação inovadora: uma abordagem teórico-prática**. Porto Alegre: Penso, p. 26–44, 2018.

ZAKI, M. J.; MEIRA, W. **Data Mining and Machine Learning: Fundamental Concepts and Algorithms**. [s.l.] Cambridge University Press, 2020.

ENDEREÇO DO AUTOR

Autor: Jean Jeremias Cardoso Campos Leão

E-mail: jj@itinganet.com.br



Autor: Abdiel Batista dos Santos

E-mail: abdielb@yahoo.com

Autor: Thaiana Martins Marques

E-mail: thaiana.marques@ifnmg.edu.br

Autor: Eliane Macedo Sobrinho Santos

E-mail: eliane.santos@ifnmg.edu.br

Autor: Jeancarlo Campos Leão

E-mail: jeancarlo.leao@ifnmg.edu.br