

Análise comparativa de métodos de explicabilidade da Inteligência Artificial no cenário educacional: um estudo de caso sobre evasão

Francisco da C. Silva¹, Rodrigo M. Feitosa¹, Luiz A. Batista¹, André M. Santana²

¹Departamento de Ensino – Instituto Federal do Maranhão (IFMA)
Jardim Renascença, São Luís - MA, CEP: 65075-441 – Brasil

²Departamento de Computação/CCN – Universidade Federal do Piauí
Bairro Ininga, CEP: 64.049-550 - Teresina - PI – Brasil

{franisco.conceicao,rodrigo.feitosa,luiz.neto}@ifma.edu.br, andreamacedo@ufpi.edu.br

Abstract. *With the increasing application of Artificial Intelligence in education, it is essential to understand the reasons behind the results obtained to ensure safe decisions. Therefore, this paper presents preliminary results of experiments with XAI methods applied to analyze dropout in student data. Three methods were analyzed: SHAP, LIME and ANCHOR. SHAP and LIME presented detailed explanations, which may require additional technical knowledge, especially SHAP, for interpretation by managers and teachers, for example. The ANCHOR method, with its rule-based approach, proved to be simpler and more intuitive, facilitating the understanding of predictions and becoming a more accessible option for the educational context.*

Resumo. *Com a crescente aplicação da Inteligência Artificial na educação, é essencial entender os motivos por trás dos resultados obtidos para garantir decisões seguras. Desta forma, este trabalho apresenta resultados preliminares de experimentos com métodos de XAI aplicados na análise de evasão em dados de alunos. Três métodos foram analisados: SHAP, LIME e ANCHOR. SHAP e LIME apresentaram explicações detalhadas, que podem exigir conhecimento técnico adicional, principalmente o SHAP, para interpretação por gestores e professores, por exemplo. O método ANCHOR, com sua abordagem baseada em regras, se mostrou mais simples e intuitivo, facilitando a compreensão das previsões e tornando-se uma opção mais acessível para o contexto educacional.*

1. Introdução

Nos últimos anos, a Inteligência Artificial (AI) tem se destacado por sua capacidade de automatizar e otimizar decisões, especialmente através do Aprendizado de Máquina (ML), que analisa grandes conjuntos de dados para identificar padrões úteis [Rachha and Seyam 2023]. Na educação, a mineração de dados usa essas técnicas para apoiar o ensino-aprendizagem, empregando-as em atividades como recomendação de materiais e análise de interações em ambientes online [Chitti et al. 2020].

No entanto, a falta de explicabilidade dos modelos de ML, especialmente os "caixa-preta", é um desafio significativo que pode levar a resultados enviesados e errôneos [Fernandes et al. 2022]. Uma abordagem emergente para melhorar a transparência desses sistemas é a eXplainable AI (XAI), que busca fornecer explicações

claras para as decisões da AI. A importância da explicabilidade é amplamente reconhecida na literatura, com autores como [Alamri and Alharbi 2021], [Neto et al. 2021] e [Qin et al. 2020] enfatizando que a transparência é crucial para construir confiança e compreensão dos resultados, especialmente em contextos educacionais. Em se tratando dos termos interpretabilidade e explicabilidade, o estudo de [Adadi and Berrada 2018] mostra que há uma interligação entre ambos e tais termos são usados como sinônimos no presente trabalho, visto estar fora do escopo do estudo a discussão mais ampla sobre isso.

A evasão escolar é um problema global persistente, e no Brasil, fatores diversos foram identificados. Entre 173 milhões de pessoas com 14 anos ou mais, 25,2 milhões frequentaram cursos de formação profissional, representando 14,5% da população [IBGE 2023]. Deste quantitativo, a necessidade de trabalho (40,2%), falta de interesse (24,7%), e para as mulheres, gravidez e tarefas domésticas, foram determinantes para evasão de alunos [IBGE 2023]. Este cenário reforça que há uma preocupação compartilhada entre instituições de ensino, profissionais da educação, coordenadores e a sociedade em geral em relação à alta taxa de evasão de alunos no contexto escolar. A XAI pode desempenhar um papel importante no enfrentamento à evasão escolar, oferecendo uma compreensão mais aprofundada dos fatores que contribuem para o abandono e permitindo a implementação de estratégias mais eficazes para manter os alunos engajados.

Quando se trata de comparar métodos de XAI, é possível destacar algumas iniciativas. O estudo de [Nayebi et al. 2022] aborda como os algoritmos de *deep learning* lidam com decisões complexas, utilizando métodos de XAI para analisar séries temporais de lesões cerebrais traumáticas. Entre as técnicas comparadas, o SHAP foi considerado robusto e confiável, porém menos compreensível, enquanto o método ANCHOR, mais acessível, é limitado a dados tabulares. O estudo de [Canha 2022] discute o crescimento da AI em setores críticos como medicina e finanças, propondo um *framework* de *benchmark* para métodos de XAI, exemplificado no estudo de previsão de doenças cardíacas com destaque para o método CIU. O estudo de [Da Gama Neto 2022] enfatiza a importância da XAI na simplificação de modelos complexos de AI, introduzindo o *Textual SHAP* para tornar interpretações mais acessíveis em contextos textuais, beneficiando usuários não especializados na compreensão de desempenho escolar. Entre estes e outros trabalhos pesquisados, apenas o de [Da Gama Neto 2022] se aproximou da evasão escolar, o que pode indicar a necessidade de mais estudos sobre o tema.

Nesse sentido, o presente trabalho visa apresentar resultados preliminares de experimentos com métodos de XAI aplicados a dados de alunos em uma instituição de ensino técnico, focando na evasão escolar. Três métodos de XAI foram comparados quanto à sua compreensibilidade, fidelidade e estabilidade, com uma análise crítica de suas vantagens e desvantagens no contexto educacional. Este estudo realiza comparações de métodos de XAI com foco na evasão escolar e avalia o potencial dos métodos SHAP, LIME e ANCHOR para explicar os fatores que influenciam a evasão.

2. Materiais e métodos

O ambiente de execução adotado neste estudo foi a plataforma do Google Colabory, que é um serviço hospedado do Jupyter Notebook que não requer configuração para ser usado e fornece acesso gratuito a recursos de computação, incluindo GPUs e TPUs.

2.1. Fonte de dados

O conjunto de dados utilizado neste estudo provém de cursos técnicos de nível médio do

Instituto Federal de Educação, Ciência e Tecnologia do Maranhão (IFMA), composto por informações pessoais, acadêmicas e socioeconômicas de 15.084 alunos, distribuído em 30 colunas, com os seguintes quantitativos de instâncias: cursos integrados (10.070 alunos), cursos subsequentes (3.902 alunos), cursos concomitantes (718 alunos) e cursos integrados EJA (394 alunos). Os registros incluem detalhes como renda bruta, escolaridade dos pais, ano de ingresso, percentual de frequência, entre outros. O pipeline desenvolvido para esta metodologia, desempenho dos modelos de ML, resultados dos métodos de XAI, conjunto de dados e relatórios de experimentação adotados neste trabalho estão disponíveis no material de apoio, em: https://osf.io/eap5h/?view_only=662bff844f814447a898d2402825c894.

A análise se concentra na classe-alvo 'situacao', que foi transformada em uma variável binária. Alunos que concluíram foram representados pela Classe 0 (Concluído), enquanto situações que podem resultar em evasão, conforme entendimento da referida instituição, como cancelamento e transferência externa, por exemplo, foram agrupados na Classe 1 (Evadido). Este ajuste foi necessário devido à variedade das situações de saída dos alunos dos cursos. Apesar da simplificação para uma classe binária, os dados originais apresentaram um desbalanceamento significativo: 78% dos registros indicaram conclusão e apenas 22% indicaram evasão. Para lidar com isso, foi utilizado o algoritmo SMOTEENN [Batista et al. 2004] para balancear as classes pela combinação de sub e super amostragem. O desbalanceamento pode impactar negativamente o desempenho e a explicabilidade dos modelos de ML, onde os padrões da classe majoritária tendem a dominar, dificultando a identificação de padrões nas classes minoritárias [Cristobal et al. 2013], como o que poderia ocorrer na análise dos motivos que contribuíram para evasão (classe minoritária) de cursos técnicos de nível médio, tratados neste estudo, sem o balanceamento de classes.

Para melhorar a qualidade dos dados, atributos que identificavam diretamente os alunos foram excluídos, assim como aqueles com mais de 20% dos valores ausentes, aliada à seleção de atributos com a técnica *Backward Selection*. Isso resultou em uma seleção cuidadosa dos atributos a serem utilizados pelos modelos de ML, visando garantir a consistência e a relevância das informações para os objetivos do estudo. Desta forma, foram selecionados 15 atributos com a aplicação da técnica *Backward Selection*: 'campus', 'anoingresso', 'periodoingresso', 'rendabruta', 'modalidade', 'idade', 'tipo_area_residencial', 'trabalha', 'percentual_frequencia', 'reprovacoes', 'genero_F', 'raca_3', 'estado_civil_Não declarado', 'estado_civil_Solteiro(a)', 'estado_civil_Viúvo(a)'.

2.2. Modelos de predição

A realização dos experimentos com ML envolveu a utilização de algoritmos amplamente reconhecidos na literatura. O *Random Forest* (RF) constrói múltiplas árvores de decisão com atributos aleatórios para prever resultados generalizados e evitar sobreajuste durante o treinamento [Huynh-Cam and Chen 2021], [Jayaprakash et al. 2020]. O *Decision Tree* (DT) particiona os dados com base em mecanismos específicos de divisão; a pureza das classes resultantes é geralmente avaliada pelo Índice de Gini [Chen et al. 2021]; é amplamente aplicado em contextos educacionais para resolver problemas de classificação [Huynh-Cam and Chen 2021]. O *Support Vector Machines* (SVM) é usado para classificar categorias por meio de hiperplanos aprendidos que maximizam as margens entre as categorias. Seu desempenho é sensível apenas às amostras próximas à superfície de decisão, resultando em boa generalização [Marbouti et al. 2016], sendo comumente usado

para prever o desempenho de aprendizado dos alunos.

Para possibilitar melhor desempenho dos modelos de ML, foi utilizada a técnica *GridSearchCV* na busca de hiperparâmetros. A avaliação do desempenho dos modelos foi realizada utilizando métricas bem estabelecidas na literatura, incluindo *precision*, *recall*, *f1-score*, *AUC* e *accuracy*.

2.3. Métodos de XAI

O conceito de XAI abrange métodos e técnicas aplicados em sistemas de AI para melhorar a compreensão dos resultados e decisões gerados, especialmente devido à complexidade de modelos do tipo “caixa-preta”, o que dificulta a explicação dos passos para alcançar um resultado específico [Oliveira 2020]. Assim, foram implementados e analisados três métodos de XAI neste estudo, que são brevemente descritos a seguir.

O *Local Interpretable Model-Agnostic Explanations* (LIME) é um método de explicabilidade local usado em modelos “caixa-preta”, oferecendo explicações simples e acessíveis [Ribeiro et al. 2016]. O LIME trata cada modelo de aprendizagem supervisionado como uma “caixa-preta” separada, gerando um modelo claro com base em observações selecionadas que introduzem variações mínimas nos parâmetros.

O *Shapley Additive exPlanations* (SHAP) é um método capaz de realizar a explicabilidade local e global, ou seja, pode explicar as previsões de um modelo para uma única amostra de dados, bem como para todo o conjunto de dados [Lundberg and Lee 2017]. Baseia-se na teoria dos jogos para calcular a contribuição média marginal de cada variável na predição [Lundberg and Lee 2017]. O valor de *Shapley* (*shap_value*) distribui de forma justa os resultados, simbolizando a contribuição proporcional de cada parâmetro para o ganho total do jogo, podendo ser aplicado em conjuntos de dados educacionais para explicar a influência de atributos na previsão do desempenho dos alunos.

O método ANCHOR esclarece previsões individuais utilizando regras “se-então” que “ancoram” a predição [Ribeiro et al. 2018], assegurando que a predição não mude mesmo com variações aleatórias nas demais características. A precisão do ANCHOR é medida pela porcentagem de amostras vizinhas que mantêm o rótulo previsto, estabelecida em 95% para garantir consistência nas explicações próximas à entrada.

2.4. Avaliação dos métodos de XAI

Os resultados foram comparados nos níveis local e global de explicabilidade. O objetivo da explicabilidade local é explicar uma única previsão focando em uma instância específica e tentando compreender como o modelo chegou à sua predição. Por outro lado, a explicabilidade global refere-se à capacidade de compreender a distribuição da saída da predição com base nas características de entrada [Carvalho et al. 2019]. Em seguida, os métodos de XAI selecionados foram analisados para cada uma das seguintes métricas:

- Estabilidade: A coesão de uma técnica de XAI sobre o mesmo conjunto de entrada é caracterizada como estabilidade [Alvarez-Melis and Jaakola 2018].
- Fidelidade: É a proximidade das explicações com as previsões do modelo, mostrando o quão bem a explicação se aproxima do modelo tipo “caixa-preta” [Carvalho et al. 2019].
- Compreensibilidade: É a capacidade de fornecer explicações compreensíveis para humanos [Carvalho et al. 2019].

3. Interpretação da predição dos modelos de ML com XAI

Para demonstração neste trabalho serão apresentadas as explicações obtidas com o uso de métodos de XAI sobre o modelo RF, que obteve o melhor desempenho entre os modelos adotados, conforme verificado na Tabela 1.

Tabela 1. Desempenho dos modelos de ML, aplicados em dados balanceados pela combinação de sub e super amostragem e seleção de atributos com a técnica *Backward Selection*.

Modelo	<i>precision</i>	<i>recall</i>	<i>f1-score</i>	<i>accuracy</i>	AUC
RF	0,842	0,849	0,845	0,848	0,849
SVM	0,293	0,500	0,369	0,585	0,500
DT	0,818	0,826	0,819	0,821	0,826

3.1. Interpretação local

SHAP, LIME, e ANCHOR foram aplicados para interpretações locais em algumas amostras classificadas com situação ‘Evadido’ do dataset. A Tabela 2 apresenta quatro destas amostras, em que três delas foram classificadas corretamente como evadidas e uma classificada erroneamente como concluída. Os percentuais apresentados na Tabela 2 para os explicadores referem-se à probabilidade de predição que cada explicador gerou para as amostras em análise.

Tabela 2. Acurácia do RF e probabilidade de predição dos Explicadores em amostras específicas.

Acurácia do RF: ~84%				Situação	
AMOSTRA (S)	SHAP	LIME	ANCHOR	Predita	Real
4181	86%	86%	97%	Evadido	Evadido
2286*	69%	69%	89%	Concluído	Evadido
3943	88%	88%	96%	Evadido	Evadido
2353	92%	92%	97%	Evadido	Evadido

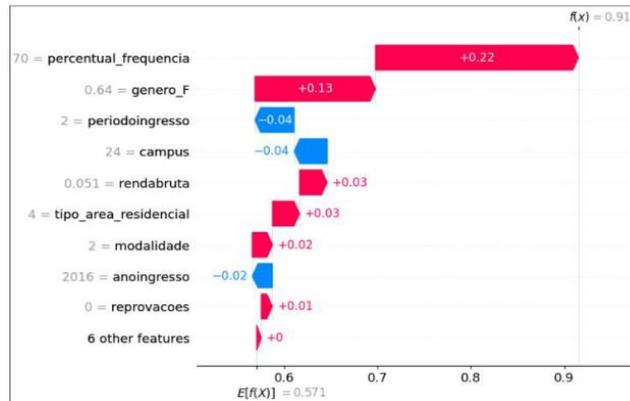
* Os percentuais da amostra 2286 referentes ao SHAP e LIME correspondem ao inverso do que é apresentado nos gráficos dos referidos explicadores, que alcançaram 31% para ‘Evadido’ na referida amostra.

Tomando como exemplo a amostra 2353, classificada corretamente com situação igual a “Evadido”, a Figura 1 traça um panorama do comportamento dos explicadores na explicação dos impactos dos atributos que levaram tal amostra a esta situação.

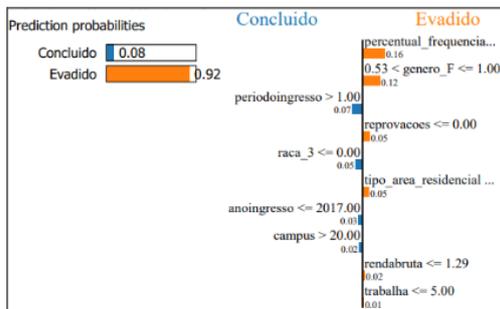
O item (a) da Figura 1 destaca que o SHAP parte de um valor base do modelo (neste caso, $E[f(x)] = 0.571$) e, a partir da soma das contribuições positivas e negativas dos atributos, alcança uma probabilidade de predição para a amostra em análise, que foi 0.915, como mostrado na função $f(x)$ do gráfico. O item (b) da Figura 1 mostra os pesos da aproximação linear da função de predição em torno da amostra 2353 usando o LIME. É ilustrado como cada variável contribui para o resultado com base nas pontuações de importância de cada atributo. Enquanto LIME e SHAP atribuem pontuações de relevância para cada recurso, ANCHOR oferece outros tipos de explicação, nomeadamente regras se-então, item (c) da Figura 1, e obteve acima de 95% de precisão para ‘Evadido’ na amostra em análise.

Considerando os principais atributos do SHAP e LIME (os seis primeiros de cima para baixo) e ANCHOR (da esquerda para a direita), os atributos ‘percentual_frequencia’

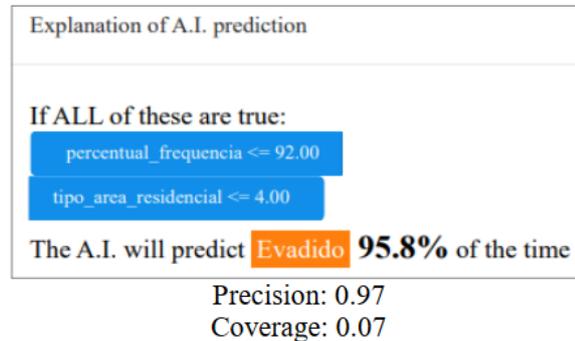
e ‘tipo_area_residencial’ (aluno oriundo da zona rural) estiveram presentes em todos os explicadores e, juntos, tiveram impacto positivo para evasão correspondendo a 25% no SHAP e 21% no LIME.



a) Gráfico do SHAP com as contribuições positivas e negativas dos atributos.



b) Gráfico do LIME com as contribuições positivas e negativas dos atributos.



c) Gráfico do ANCHOR com as regras geradas.

Figura 1. Resultados de diferentes métodos de interpretação local em torno da amostra 2353. Em gráficos SHAP e LIME, características indicadas com as cores vermelho (ou laranja) e azul contribuem com resultados positivos e negativos para evasão, respectivamente.

Em [Nayebi et al. 2022], constatou-se que os resultados obtidos pelo ANCHOR eram inconsistentes com os do SHAP e do LIME para dados estruturados, com apenas uma das quatro regras do ANCHOR coincidindo com as principais variáveis identificadas por esses explicadores. No entanto, o presente trabalho observou que as variáveis do ANCHOR também apareceram entre as principais características identificadas por LIME e SHAP, o que pode indicar boa consistência entre os explicadores.

3.2. Interpretação global

As explicabilidades globais foram realizadas apenas com o SHAP, a partir do gráfico de resumo, que exibe o efeito de cada atributo na previsão do modelo avaliado (os atributos são dispostos em ordem decrescente, de cima para baixo). Cada ponto no gráfico representa o *shap_value* calculado para o atributo em uma amostra.

A Figura 2 apresenta o resultado da aplicação do SHAP no conjunto de dados definido para este estudo. As previsões do modelo são afetadas positivamente pelos pontos à direita do zero no eixo horizontal, e neste caso chegando mais perto da classe ‘Evadido’. Ao passo que, os valores à esquerda de 0 no referido eixo têm impacto negativo, o que tende a levar a previsão para a classe ‘Concluído’. Assim, os atributos de

percentual de frequência, gênero, renda bruta e ano de ingresso foram fortemente impactantes entre os 5 principais atributos do modelo. O atributo renda bruta indica claramente que valores baixos (representados em tons de azul) têm um impacto positivo na evasão do aluno.

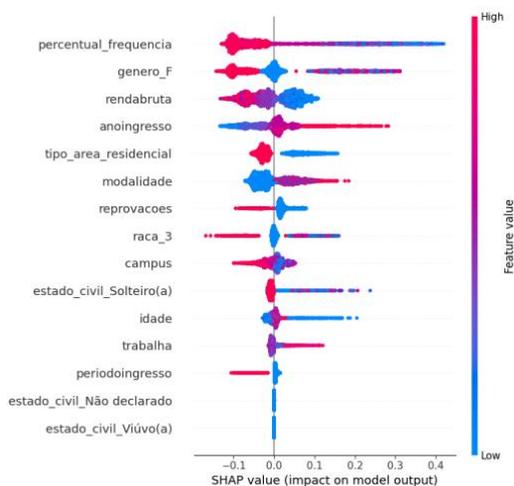


Figura 2. Principais atributos identificados pelas explicações globais do SHAP no modelo RF. Cada ponto representa uma amostra e o eixo horizontal reflete o valor SHAP da característica.

4. Comparação dos métodos de XAI

Considerando a Tabela 2, os métodos de XAI foram comparados a partir dos critérios de avaliação definidos na seção 2.4, que são atributos importantes para XAI, principalmente no contexto escolar, uma vez que o domínio educacional exige decisões confiáveis.

4.1. Estabilidade

A estabilidade avalia a consistência das explicações fornecidas entre diferentes amostras. Os métodos SHAP e LIME mostraram variações significativas nas amostras avaliadas, com probabilidade de predição variando de 31% a 92%, levando a sugerir alguma instabilidade. No entanto, dado o número limitado de amostras, essa variação pode ser específica para os casos testados e não necessariamente refletir uma instabilidade generalizada. O método ANCHOR apresentou maior estabilidade, na maioria das amostras, com probabilidade de predição (*precision*) de 97%, exceto na amostra 2286, onde a probabilidade foi de 89%. Apesar de uma variação menor nas probabilidades do método ANCHOR, mais amostras seriam necessárias para confirmar essa tendência.

4.2. Fidelidade

A fidelidade dos explicadores é essencial para garantir que as explicações reflitam corretamente as previsões do modelo. Para a amostra 2286, SHAP e LIME refletiram a classificação errônea do RF, apresentando uma probabilidade de predição de 31% para ‘Evadido’. Apesar disso, em outras amostras, ambos os métodos de XAI mostraram alta fidelidade, com probabilidade de predição variando de 86% a 92%. O método ANCHOR mostrou uma probabilidade de predição de 89% para ‘Concluído’ na mesma amostra, o que também reflete o erro da previsão do modelo RF. No entanto, este mesmo método também demonstrou alta fidelidade em outras amostras, com de 97% de precisão, o que sugere que é confiável na maioria dos casos.

4.3. Compreensibilidade

A compreensibilidade considera a facilidade com que as explicações podem ser interpretadas. O método ANCHOR demonstrou ser mais simples de entender, tendo em vista se tratar de uma abordagem baseada em regras para fazer a predição de como o modelo obteve o resultado. Tal constatação ocorre tanto pela interpretação dos autores deste trabalho quanto pelo próprio caráter autoexplicativo que as regras do ANCHOR proporcionam. LIME também é relativamente compreensível devido à simplicidade dos modelos locais, facilitando a interpretação das explicações. Em contraste, SHAP oferece explicações detalhadas e baseadas em valores de *Shapley*, que podem ser mais complexas e menos intuitivas, limitando a compreensão geral.

4.4. Implicações da adoção dos métodos de XAI no contexto educacional

O método ANCHOR se destaca por sua estabilidade e clareza, oferecendo explicações diretas e compreensíveis, com alta fidelidade na maioria das amostras. Essa simplicidade torna o ANCHOR particularmente adequado para gestores escolares, professores, etc., que podem se beneficiar de explicações intuitivas e acessíveis.

O SHAP, por outro lado, fornece explicações detalhadas e de alta fidelidade, mas a complexidade dos gráficos e a variação na precisão entre amostras podem tornar suas explicações mais difíceis de entender. Para tornar o SHAP mais útil no ambiente educacional, seria necessário simplificar as visualizações e oferecer tutoriais adicionais para melhorar a compreensão.

O LIME também apresenta um bom desempenho, mas enfrentou desafios com variações na fidelidade e precisão. Para ser mais eficaz, especialmente no ambiente educacional, o LIME precisaria de apresentações mais simplificadas e exemplos claros para melhorar a estabilidade e a clareza das explicações.

5. Conclusões e Considerações finais

Este trabalho realizou uma comparação de três abordagens de XAI no contexto educacional, apresentando resultados preliminares de um estudo em evolução. As descobertas iniciais indicam que os resultados de SHAP (com abrangência em escopo local e global) e LIME tem semelhanças significativas, embora possam apresentar algum nível de complexidade em seus gráficos. O método ANCHOR, por sua vez, é uma técnica mais compreensível, apesar do fato de sua abrangência, da mesma forma que o LIME, ser limitada ao escopo local. Constatações semelhantes foram encontradas em [Da Gama Neto 2022] e [Nayebi et al. 2022].

Cada estudo possui suas próprias limitações, e é essencial continuar pesquisando para aprimorar essas técnicas e abordagens. O estudo apresentado neste trabalho também identificou algumas limitações, que foram delineadas como áreas para pesquisas futuras: explorar outras possibilidades de análise de dados educacionais; ampliar a gama de métodos e métricas de avaliação de XAI a serem explorados; aumentar a variabilidade dos dados; expandir o estudo para o ensino superior; realizar análises comparativas mais amplas dos resultados e publicar novas contribuições para a comunidade científica.

Por fim, o estudo avança na compreensão da aplicação da XAI na educação, ressaltando a importância da explicabilidade dos modelos de ML nesse contexto. Os resultados obtidos com os métodos adotados podem aumentar a confiança na tomada de decisões e contribuir para criação de estratégias de enfrentamento à evasão de alunos.

Agradecimentos e reconhecimentos

O presente trabalho foi realizado com apoio do Instituto Federal de Educação, Ciência e Tecnologia do Maranhão (IFMA).

É importante destacar que este trabalho fez uso da ferramenta de AI *Chat GPT* como auxílio para: identificar informações nos trabalhos relacionados, proporcionando melhor apresentação da metodologia adotada; a partir da geração de conteúdo, realizar a análise comparativa dos métodos de XAI e; criar algumas referências.

Referências

- Adadi, A.; Berrada, M. (2018). Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE Access*, v. 6, p. 52138–52160. Disponível em: <https://doi.org/10.1109/ACCESS.2018.2870052>. Acesso em: 3 ago. 2024.
- Alamri, R. and Alharbi, B. (2021). Explainable student performance prediction models: A systematic review. *IEEE Access*, 9:33132–33143.
- Alvarez-Melis, D. and Jaakola, T. S. (2018). Towards Robust Interpretability with Self-Explaining Neural Networks. In: *Neural Information Processing Systems*.
- Batista, G., Prati, R. and Monard, M.-C. (2004). A study of the behavior of several methods for balancing machine learning training data. *SIGKDD Explorations*, v. 6, p. 20–29.
- Carvalho, D. V., Pereira, E. M. and Cardoso, J. S. (2019). Machine Learning Interpretability: A Survey on Methods and Metrics. *Electronics*, v. 8, n. 8, p. 832, 26 jul. Disponível em: <https://www.mdpi.com/2079-9292/8/8/832/htm>. Acesso em: 28 jul. 2024.
- Chen, J. et al. (2019). MOOC dropout prediction using a hybrid algorithm based on decision tree and extreme learning machine. *Mathematical Problems in Engineering*, v. 2019, p. 1-11. Disponível em: <https://doi.org/10.1155/2019/8404653>. Acesso em: 02 ago 2024.
- Chitti, M., Chitti, P. and Jayabalan, M. (2020). Need for interpretable student performance prediction. In *2020 13th International Conference on Developments in eSystems Engineering (DeSE)*, pages 269–272.
- Cristobal, R. et al. (2013). Web usage mining for predicting final marks of students that use Moodle courses. *Computer Application Engineering Education*, Wiley Periodicals, v. 21, n. 1, p. 135-146.
- Da Gama Neto, M. V. (2022). Análise comparativa das técnicas de Explainable AI e um novo método para geração de explicações textuais. [s.l.] Universidade Federal de Pernambuco, 10 mar.
- Canha, D. M. De Carvalho Martins. (2022). *Building a benchmark framework for explainable Artificial Intelligence (XAI) methods*. Instituto Superior Técnico, Lisboa.
- Fernandes, M. (2022). Inteligência artificial explicável aplicada a aprendizado de máquina: Um estudo para identificar estresse ocupacional em profissionais da saúde. Ano. Trabalho de Conclusão de Curso (Graduação em Engenharia de Computação) — Universidade Federal de Santa Catarina, Araranguá. Disponível em:

- https://repositorio.ufsc.br/bitstream/handle/123456789/243462/TCC_Milena_Fernandes-assinado.pdf?sequence=1&isAllowed=y. Acesso em: 27. Jul. 2024.
- Huynh-Cam, T.-T. Chen, L.-S. and Le, H. (2021). Using decision trees and random forest algorithms to predict and determine factors contributing to first-year university students' learning performance. *Algorithms*, v. 14, n. 11, p. 318, 2021.
- IBGE - Instituto Brasileiro de Geografia e Estatística. (2023). "Pesquisa Nacional por Amostra de Domicílios Contínua - Censos 2022". ISBN 978-85-240-4567-7. Rio de Janeiro, IBGE.
- Jayaprakash, S., Krishnan, S. and Jaiganesh, V. (2020). Predicting students' academic performance using an improved random forest classifier. In: *2020 International Conference on Emerging Smart Computing and Informatics (ESCI)*, 2020,. Proceedings... p. 238-243, 2020. DOI: 10.1109/ESCI48226.2020.9167547.
- Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.
- Marbouti, F.; Diefes-Dux, H. A.; Madhavan, K. (2016). Models for early prediction of at-risk students in a course using standards-based grading. *Computers & Education*, vol 103, pp 1-15.
- Nayebi, A. et al. (2022). An empirical comparison of explainable Artificial Intelligence methods for clinical data: A case study on traumatic brain injury. Disponível em: <<http://arxiv.org/abs/2208.06717>>. Acesso em: 20 jul. 2024.
- Neto, M. V. G., Vasconcelos, G. C. and Zanchettin, C. (2021). Mineração de dados aplicada à predição do desempenho de escolas e técnicas de interpretabilidade dos modelos. In *Anais do XXXII Simpósio Brasileiro de Informática na Educação*, pages 773–782. SBC.
- Oliveira, D. F. N. (2020). Dissertação de Mestrado. Escola Politécnica, Universidade de São Paulo. Disponível em: <https://doi.org/10.11606/D.3.2020.tde-08032021-102116>. Acesso em: 02 ago. 2024.
- Qin, F., Li, K., and Yan, J. (2020). Understanding user trust in artificial intelligence-based educational systems: Evidence from china. *British Journal of Educational Technology*, 51(5):1693–1710.
- Rachha, A. and Seyam, M. (2023). Explainable ai in education : Current trends, challenges, and opportunities. In *SoutheastCon 2023*, pages 232–239.
- Ribeiro, M. T., Singh, S. and Guestrin, C. (2016). "why should i trust you?": Explaining the predictions of any classifier. *KDD '16*, page 1135–1144, New York, NY, USA. Association for Computing Machinery.
- Ribeiro, M. T., Singh, S. and Guestrin, C. (2018). ANCHOR: high-precision model-agnostic explanations. Association for the Advancement of Artificial Intelligence. Disponível em: www.aaai.org. Acesso em: 02 ago. 2024.