Um Mapeamento Sistemático Sobre as Técnicas Emergentes de Pré-Processamento para Mitigação de Viés em Dados de Treinamento

Marco Antonio da Silva Neves Filho[1], Julio Cezar Costa Furtado[1]

[1]Departamento de Ciências Exatas e Tecnológicas
Universidade Federal do Amapá (Unifap) – Macapá, AP – Brasil

mrcantoniofilho@gmail.com, furtado@unifap.br

Resumo. *Este trabalho apresenta um mapeamento sistemático da literatura sobre técnicas emergentes de mitigação de viés em dados de treinamento para aprendizado de máquina, com foco no pré-processamento entre 2020 e 2024. Foram analisados 86 estudos a partir das bases IEEE Xplore e ACM Digital Library, resultando na identificação de 244 técnicas, das quais 22 foram aplicadas em mais de um estudo. Os resultados evidenciam o crescimento de abordagens híbridas e o uso de múltiplas métricas de avaliação. O estudo reforça a importância do pré-processamento para promover maior equidade nos sistemas de aprendizado de máquina.*

Abstract. *This paper presents a systematic literature mapping on emerging bias mitigation techniques in training data for machine learning, focusing on preprocessing from 2020 to 2024. A total of 86 studies from IEEE Xplore and ACM Digital Library were analyzed, identifying 244 techniques, with 22 used in more than one study. Results show increased use of hybrid approaches and multiple evaluation metrics. The study highlights the role of preprocessing in enhancing fairness in machine learning systems.*

1. Introdução

Machine Learning (ML) tem se consolidado não apenas como uma ferramenta técnica, mas como um agente de transformação econômica (ATHEY, 2019), social e cultural (FEHER e ZELENKAUSKAITE, 2020). Seu uso permite otimizar decisões, personalizar experiências e antecipar comportamentos, ampliando a capacidade humana de lidar com dados e complexidade. Compreender ML, portanto, vai além da capacitação técnica: trata-se de se preparar para um futuro em que essa tecnologia estará cada vez mais central na organização da sociedade. Contudo, os benefícios trazidos por ML também vêm acompanhados de riscos importantes, principalmente quando os modelos são mal implementados. Por essa razão, é necessário adotar uma postura crítica ao estudar e aplicar técnicas de aprendizado de máquina. Avaliar essas tecnologias envolve não apenas conhecimento técnico, mas também responsabilidade social, econômica e ética (SIAM e BHATTACHARJEE, 2024), o que reforça a importância de uma formação sólida e atualizada na área.

O pré-processamento de dados é uma etapa essencial na construção de modelos de ML, pois os algoritmos aprendem diretamente com os dados

fornecidos. Caso esses dados estejam inconsistentes, incompletos ou enviesados, o desempenho e a equidade do modelo serão afetados. Técnicas como normalização, tratamento de valores ausentes, remoção de ruídos e codificação adequada são fundamentais para garantir resultados confiáveis (SURESH e GUTTAG, 2019), sendo o primeiro passo crítico antes mesmo do treinamento dos modelos.

Diante do ritmo acelerado de evolução do campo, é crucial acompanhar as tendências emergentes em ML, especialmente no que diz respeito a práticas mais eficientes, acessíveis e éticas. Ignorar essas inovações representa um risco de estagnação, uma vez que técnicas tradicionais, ainda que validadas no passado, podem tornar-se obsoletas ou ineficazes. Além disso, *benchmarks* e estudos recentes apontam quais métodos estão sendo mais eficazes em contextos reais, como saúde, finanças e tecnologia embarcada, permitindo antecipar soluções com maior impacto prático. Nesse contexto, estudar ML é essencial não só por seu valor técnico, mas pelo seu papel estratégico e social. Acompanhando tendências de pré-processamento torna-se possível promover maior equidade e melhorar o desempenho dos sistemas.

Assim, o objetivo deste estudo é identificar, categorizar e analisar tais técnicas emergentes, publicadas entre 2020 e 2024, compreendendo suas aplicações e seus impactos sobre a justiça algorítmica e a performance dos modelos.

2. O Mapeamento Sistemático

Este estudo adota a metodologia de Mapeamento Sistemático da Literatura (MSL), conforme as diretrizes estabelecidas por Kitchenham e Charters (2007). O processo metodológico teve início com a elaboração de um protocolo de pesquisa, que contemplou a definição da questão de pesquisa principal, os critérios de inclusão e exclusão, bem como a estratégia de busca a ser empregada.

A condução do mapeamento seguiu as seguintes etapas: definição das questões de pesquisa, identificação das bases de dados relevantes, construção e refinamento da *string* de busca com base em termos-chave derivados da questão primária, e aplicação dos critérios de seleção previamente estabelecidos. Após a realização das buscas, os estudos recuperados foram submetidos a uma triagem inicial, seguida por uma análise mais aprofundada, com base nos critérios definidos no protocolo. Os estudos selecionados foram então organizados e sintetizados para possibilitar a análise final. Nesse estágio, além da resposta à questão primária, também foram abordadas as questões secundárias relacionadas ao tema investigado.

A etapa de análise final visa atender ao objetivo central do mapeamento, que consiste em responder de forma abrangente à questão primária. Os resultados obtidos, devidamente sistematizados, têm o propósito de subsidiar futuras pesquisas e aplicações práticas na área de Aprendizado de Máquina (ML), especialmente no que se refere ao tratamento de viés nos dados de treinamento. A questão de pesquisa primária (QP) deste estudo é formulada nos

seguintes termos: *Quais são as técnicas emergentes para mitigação de viés em dados de treinamento no contexto de Aprendizado de Máquina?*

A formulação da QP foi orientada pelo modelo PICO (*Population, Intervention, Comparison, Outcomes*), conforme proposto por Sackett et al. (2000). No entanto, o elemento *Comparison* foi omitido, por não se mostrar pertinente ao escopo deste mapeamento sistemático. Dessa forma, a questão foi estruturada com base nos três componentes relevantes, População, Intervenção e Resultados , conforme demonstrado no Quadro 1.

Assim, o principal objetivo deste trabalho é identificar e analisar técnicas emergentes empregadas entre 2020 e 2024 para mitigar viés em dados de treinamento em aprendizado de máquina (ML). A Questão Primária foi formulada com base no modelo PIO: População – dados de treinamento de ML; Intervenção – técnicas de mitigação de viés; Resultado – abordagens recentes identificadas na literatura.

Além da formulação da Questão Primária (QP), este estudo também definiu um conjunto de questões secundárias (QS) com o propósito de orientar a análise dos estudos primários selecionados e enriquecer a compreensão das características relevantes para a síntese da QP. As questões secundárias foram delineadas com foco na contextualização, avaliação e abrangência das técnicas investigadas nos estudos, conforme segue:

- QS1: Os estudos analisados são de iniciativa acadêmica ou industrial?
- QS2: Que métricas são utilizadas para avaliar a eficácia das técnicas apresentadas?
- QS3: Os estudos realizados utilizam de medidas além de técnicas pré-processamento para a mitigação de viés?

A seleção das fontes de dados para este mapeamento sistemático foi conduzida com base em critérios que asseguram a viabilidade, acessibilidade e relevância das publicações recuperadas. As bases de dados consideradas elegíveis devem:

- Possibilitar consultas por meio digital, com acesso online e suporte a mecanismos de busca automatizados;
- Permitir o acesso integral aos textos completos das publicações, seja por meio do Portal de Periódicos da CAPES, pelo domínio institucional da UNIFAP, ou ainda Google e Google Scholar;
- Incluir publicações redigidas em pelo menos um dos seguintes idiomas: português ou inglês;

Com base nesses critérios, foram selecionadas as bases de dados IEEE Xplore e ACM Digital Library. Além de atenderem integralmente às condições estabelecidas, ambas são reconhecidas por sua abrangência, qualidade e relevância científica, constituindo fontes consolidadas nas áreas de computação, engenharia e tecnologia.

A formulação da string de busca partiu da extração de termos-chave da Questão Primária (QP), bem como dos elementos definidos pelos critérios PIO. Considerando o escopo deste trabalho, que busca identificar tendências recentes na área, o período de análise foi delimitado entre janeiro de 2020 e

dezembro de 2024, abrangendo os últimos cinco anos de produção científica no tema. Abaixo a string de busca utilizada:

> *(bias OR fairness OR discrimination) AND (dataset* OR "data set*" OR "training data*") AND (adjust* OR mitigat* OR remov* OR correct* OR optimizat* OR reduc* OR minimiz*) AND ("machine learning" OR "deep learning" OR "supervised learning") AND ("case study" OR experiment OR empirical OR "real-world" OR implement* OR practica* OR propos* OR novel) NOT (interview OR survey OR "position paper")*

Os critérios de seleção são divididos em Critério de Inclusão (CI) e Critério de Exclusão (CE), descritos no Quadro 2. Estes são estabelecidos pelos pesquisadores de forma a garantir que a seleção possa classificar os estudos adequadamente (KITCHENHAM e CHARTES, 2007).

Quadro 2. Descrição dos critérios de seleção.

Código	*Descrição*
CI1	Estudos que apresentem, primária ou secundariamente, técnicas de mitigação de viés em dados de treinamento para Machine Learning (ML).
CE1	Artigos que não estejam disponíveis livremente para consulta ou download nas fontes de pesquisa ou por meio de busca manua via Google e/ou Google Scholar.
CE2	Artigos não relacionados aos objetivos da pesquisa.
CE3	Artigos repetidos (em mais de uma fonte de busca) tiveram apenas sua primeira ocorrência considerada.
CE4	Estudos enquadrados como resumos, keynote speeches, cursos, tutoriais e afins.
CE5	Artigos que não mencionam as palavras-chave da pesquisa no título, resumo ou nas palavras-chave do artigo.
CE6	Excluir se o estudo não estiver apresentado em uma das linguagens aceitas (Inglês e Português).

A seleção dos estudos foi inicialmente feita com base nos títulos, resumos e palavras-chave. Quando essas informações não eram suficientes, o texto completo era analisado. Após essa triagem, aplicaram-se os critérios de inclusão e exclusão, resultando na seleção final. Para cada estudo incluído, foram extraídos metadados como tipo e ano de publicação, instituições e países de origem, editora, autoria, além do título, resumo e palavras-chave, organizando-se essas informações para análise posterior.

3. Resultados do MSL

A busca foi realizada nas bases de dados IEEE Xplore e ACM Digital Library, resultando inicialmente em 844 e 408 publicações, respectivamente, totalizando 1.252 trabalhos. Após a triagem por títulos, resumos e palavras-chave, 169 estudos foram pré-selecionados. Com a aplicação dos critérios de inclusão e exclusão definidos no protocolo, 86 estudos foram mantidos para análise e extração de dados.

A partir dessa amostra final, procedeu-se à identificação e sistematização das técnicas relacionadas à mitigação de viés, com foco exclusivo naquelas aplicadas durante a etapa de pré-processamento dos dados. As técnicas apresentadas referem-se diretamente à Questão Primária (QP) deste mapeamento sistemático, cujo objetivo é identificar as técnicas emergentes de mitigação de viés aplicadas à etapa de pré-processamento de dados em

modelos de Aprendizado de Máquina. Para isso, foram extraídas de cada estudo apenas as abordagens utilizadas ou introduzidas especificamente nessa fase, desconsiderando-se intervenções realizadas em etapas posteriores, como em-processamento ou pós-processamento.

Ao todo, foram identificadas 244 técnicas distintas. No entanto, considerando o foco deste estudo na identificação de tendências emergentes, optou-se por listar somente aquelas que ocorreram em pelo menos dois estudos distintos. Essa filtragem permite destacar as técnicas com maior recorrência na literatura recente, oferecendo uma visão mais clara das abordagens que vêm ganhando espaço na comunidade científica.

As técnicas mais frequentemente identificadas nos estudos analisados incluem: Instance Reweighting (11 ocorrências), Fair Representation Learning (9), SMOTE – Synthetic Minority Oversampling Technique (8), o Quadro 3 resume este resultado. A presença recorrente dessas abordagens nos estudos analisados indica sua crescente adoção como estratégias promissoras para a mitigação de viés na etapa de pré-processamento.

Quadro 3. Técnicas de Mitigação de Viés no Pré-processamento

Técnica	*Ocorrências*	*Técnica*	*Ocorrências*
Instance Reweighting	11	Minority Class Augmentation	3
Fair Representation Learning	9	Neighbourhood-based Undersampling	3
SMOTE	8	PCA (Principal Component Analysis)	3
Random Oversampling	6	SMOTE-ENN	3
Class Imbalance Rebalancing	5	Statistical Parity Correction	3
Generative Data Augmentation	4	Sensitive Attribute Suppression	3
Label Adjustment	4	Adversarial Debiasing	2
ADASYN (Adaptive Synthetic Sampling)	3	Data Normalization	2
Balanced Training Data	3	Missing Data Handling	2
Fairness-driven Binning	3	Soft Label Smoothing	2
Hybrid Resampling	3	Undersampling	2

As questões secundárias deste mapeamento sistemático forneceram suporte essencial para a análise qualitativa dos estudos selecionados, complementando a resposta à Questão Primária. Em relação à QS1, verificou-se que nenhuma das publicações é de autoria exclusivamente industrial. A maioria dos estudos (73) é proveniente de iniciativas exclusivamente acadêmicas, enquanto os 12 restantes resultam de colaborações entre instituições acadêmicas e industriais. Quanto à QS2, foram identificadas 153 métricas distintas utilizadas para avaliar a eficácia das técnicas propostas, totalizando 341 ocorrências. No entanto, cerca da metade dessas ocorrências (49,56%) concentra-se em apenas 14 métricas. As mais frequentes foram: Accuracy (29 ocorrências), Sensitivity (21), F-Measures (17), AUROCC (14), SPD (13), Precision (12), Statistical Parity (11), Confusion Matrices (10), G-Mean (9), Specificity (9), EOD (7), Equalized Odds (7), Disparate Impact (6) e Spectral Analysis (6). Por fim, em relação à QS3, observou-se que 27 dos 86 estudos

analisados adotam abordagens híbridas para mitigação de viés, combinando técnicas de pré-processamento com estratégias de em-processamento e/ou pós-processamento.

6. Considerações finais

Os dados sintetizados por este MSL podem auxiliar pesquisadores e profissionais da área a entender o atual estado da pesquisa científica sobre mitigação de viés em pré-processamento, o que fornece base para decisões metodológicas mais embasadas. Ao reunir e categorizar técnicas recentes durante uma fase específica do treinamento, o estudo oferece um panorama valioso para orientar a aplicação de abordagens existentes, assim como o desenvolvimento de novas soluções mais eficazes e justas em sistemas de aprendizado de máquina.

Sugere-se, para trabalhos futuros, a ampliação da pesquisa para outras bases científicas importantes ou o foco em fases seguintes do processo de treinamento, ainda no contexto de busca e sintetização das informações. Outro caminho interessante seria a realização de estudos empíricos que avaliassem e comparassem o desempenho e a efetividade das técnicas mapeadas em diferentes domínios de aplicação. Também seriam relevantes o estudo e desenvolvimento de abordagens automatizadas de seleção de técnicas de mitigação, com base nas individualidades da área de aplicação, das características dos dados e das métricas de equidade desejadas.

Agradecimentos

Em alinhamento com os princípios da ciência aberta, todo o material utilizado e produzido nesta pesquisa está publicamente disponível em https://zenodo.org/records/16293903.

Referencias Bibliográficas

Athey, S. (2019). “The impact of machine learning on economics”. In: Agrawal, A., Gans, J., and Goldfarb, A. (Eds.), The Economics of Artificial Intelligence: An Agenda, pp. 507–547. Chicago: University of Chicago Press.

Feher, K. and Zelenkauskaite, A. (2020). “AI in society and culture: decision making and values”. In: Conference on Human Factors in Computing Systems Extended Abstracts – CHI 2020, Honolulu. Proceedings. New York: ACM.

Kitchenham, B. and Charters, S. M. (2007). Guidelines for performing systematic literature reviews in software engineering. Version 2.3. EBSE Technical Report EBSE-2007-01. Keele; Durham: Software Engineering Group, Keele University; Department of Computer Science, University of Durham. A

Siam, M. K. H., Bhattacharjee, M., Mahmud, S., Sarkar, M. S., and Rana, M. M. (2024). “The impact of machine learning on society: an analysis of current trends and future implications”.

Suresh, H. and Guttag, J. V. (2019). “A framework for understanding sources of harm throughout the machine learning life cycle”. Communications of the ACM, 64(8), pp. 62–71.