

---

**Faculdade de Tecnologia de Americana "Ministro Ralph Biasi"**

**ANÁLISE PREDITIVA E EQUITATIVA COM AI FAIRNESS 360**

**PREDICTIVE AND EQUITABLE ANALYSYS WITH AI FAIRNESS  
360**

Camila Cardoso Freitas, Faculdade de Tecnologia de Americana (FATEC Americana), Email: [camila.freitas@fatec.sp.gov.br](mailto:camila.freitas@fatec.sp.gov.br)

Clerivaldo Jose Roccia, Faculdade de Tecnologia de Americana (FATEC Americana), Email: [clerivaldo.roccia@fatec.sp.gov.br](mailto:clerivaldo.roccia@fatec.sp.gov.br)

**Resumo**

A inteligência artificial é uma tecnologia que trouxe muitos benefícios para a humanidade, em questão de otimização de tempo e praticidade quanto aos serviços prestados, e por consequência, as ações comportamentais do ser humano passaram a ser armazenadas e analisadas, para posteriormente indicarem uma predição comportamental referente a serviços personalizados de cada indivíduo, englobando algoritmos de vieses discriminatórios já enraizados na sociedade. Portanto, o artigo abordará como a análise preditiva impacta diretamente as estratégias de negócios e as tomadas de decisões de mercado na economia da informação, resultando em implicações éticas, sociais e econômicas e como mitigar esses vieses em modelos de aprendizagem de máquina utilizando a ferramenta AI Fairness 360 em um estudo de caso, onde foi detectado vieses em atributos sensíveis de raça e gênero e conforme a aplicação de métodos e métricas de avaliação para o treinamento do modelo é perceptível uma redução nos efeitos negativos de dados preconceituosos resultando em análises preditivas mais justas e equitativas.

**Palavras-chave:** Análise preditiva, algoritmos discriminatórios, aprendizagem de máquina, AI Fairness 360.

**Abstract**

*Artificial intelligence is a technology that has brought many benefits to humanity, in terms of time optimization and practicality in the services provided. Consequently, human behavioral actions have started to be stored and analyzed to predict personalized service behaviors for each individual, encompassing algorithms with discriminatory biases deeply rooted in society. Therefore, this article will address how predictive analysis directly impacts business strategies and market decision-making in the information economy, resulting in ethical, social, and economic implications, and how to mitigate these biases in machine learning models using the AI Fairness 360 tool in a case study. In this study, biases in sensitive attributes such as race and gender were identified, and through the application of evaluation methods and metrics for model training, a reduction in the negative effects of prejudiced data is observed, resulting in fairer and more equitable predictive analyses.*

---

## Faculdade de Tecnologia de Americana "Ministro Ralph Biasi"

**Keywords:** *predictive analytics, discriminatory algorithms, machine learning, Ai fairness 360.*

### 1. Introdução

A inteligência artificial (IA) desempenha um papel crucial nas estratégias de negócios, permitindo que as empresas consigam antecipar tendências de mercado, por meio de análises preditivas e conseqüentemente acaba aderindo a vieses discriminatórios, o que pode resultar em distorções éticas.

De acordo com Goldsmith (2010) a inteligência artificial é uma tecnologia multidisciplinar que procura desenvolver e aplicar técnicas computacionais que simulem o comportamento humano em funções distintas.

Como descreve Goldsmith (2010), a inteligência artificial tem como principal propósito, coletar dados de cada ação humana para posteriormente fazer uma simulação preditiva de comportamentos em situações específicas. Com isso essas análises estão se tornando a cada dia mais precisas, e conseqüentemente é possível notar um aumento exponencial de empresas que utilizam a inteligência artificial como base para recomendar itens customizados com o objetivo de obter uma estratégia de marketing mais eficaz com retorno de lucro garantido. Segundo Marlon (2023) a Amazon por exemplo, utiliza ferramentas de IA para analisar o histórico de compras e pesquisas dos clientes para indicar produtos personalizados, enquanto de acordo com Hollingsworth (2022) a Google interpreta com exatidão os interesses de pesquisas dos usuários preenchendo de forma automática o sistema de busca.

Complementando os autores acima a SAS (2023) afirma que as análises preditivas são baseadas em coletas de informações e métodos de aprendizagem de máquina, além de modelos estatísticos, para determinar a possibilidade de eventos futuros tendo os dados históricos como fonte. O propósito é obter uma projeção precisa do que poderá acontecer adiante, por intermédio de informações

---

## **Faculdade de Tecnologia de Americana "Ministro Ralph Biasi"**

já armazenadas e analisadas.

Ou seja, a partir dessas análises preditivas é possível moldar a economia por conta das previsões de mudanças e tendências de mercado, podendo assim tomar decisões estratégicas embasadas nas preferências do consumidor final.

Complementando a ideia de SAS (2023), de acordo com Udacity (2017 *apud* SOARES, 2017, p.2) os recursos de análise preditiva têm como principal meta ascender o sucesso da empresa já que é possível prever as necessidades de forma personalizada e resolver problemas de demanda promovendo soluções para um aumento do fluxo de mercado e assim melhorar a performance do negócio.

Portanto, essas práticas de análises preditivas podem acarretar implicações éticas, sociais e econômicas por poderem gerar perfis de consumidores, influenciando diretamente a dinâmica de mercado e o comportamento dos usuários. Na área econômica por exemplo, impacta diretamente as estratégias de negócios e as tomadas de decisões da empresa com o objetivo de atender a demanda indicada nos dados. Quanto a implicações sociais, a análise preditiva pode aumentar a desigualdade por conta de discriminação algorítmica. Por exemplo, contratação para vagas de emprego onde os algoritmos podem dar preferência a gêneros, idades ou origens étnicas por intermédio de dados históricos que demonstram preconceito. Assim como concessões de créditos, os algoritmos podem negar acesso a certos grupos sociais por morar em locais menos desfavorecidos ou por pertencer a uma minoria étnica. Além, de publicidades direcionadas o que pode aumentar estereótipos e marginalizar outros o que pode provocar uma ascensão de desigualdade social sem precedentes reforçando privilégios e preconceitos existentes na sociedade.

Desta forma, o artigo abordará os impactos sociais e éticos e os meios de mitigação desses vieses utilizando a ferramenta de equidade AI Fairness 360 da IBM com o objetivo de tornar o desenvolvimento e a implementação da Inteligência

---

## **Faculdade de Tecnologia de Americana "Ministro Ralph Biasi"**

Artificial mais justa e igualitária sem perpetuar algoritmos discriminatórios que prejudiquem certos grupos sociais de maneira indesejada.

### **2. AI Fairness 360**

Segundo Queiroz (2020) a ferramenta AI Fairness 360 (AIF360) criada pela IBM é um projeto de código aberto em Python que analisa o aprendizado da máquina com o objetivo de evitar que passem despercebidos resultados de caráter segregado, por conta de preferências históricas nocivas que estão imbuídas na sociedade. Sendo assim, sua funcionalidade é baseada em diferentes estágios do processo de desenvolvimento de um modelo de IA disponibilizando recursos para analisar, detectar, treinar e assim reduzir vieses discriminatórios de diversas áreas, como finanças, saúde, educação e recursos humanos.

De acordo com os autores Mahoney, Varshney e Hind (2020). A implementação da ferramenta AIF360 é fundamentada a partir do pré-processamento dos dados passando por etapas como normalização, tratamento de valores ausentes e seleção de atributos relevantes.

Sendo assim a etapa de normalização é utilizada para garantir que atributos diversos estejam na mesma escala. O próximo estágio seria o tratamento de valores ausentes para evitar distorções de resultado e também é adotado a seleção de atributos relevantes destacando os dados mais significativos para a execução da tarefa em questão. O propósito dessas ações é estabelecer um conjunto de dados sólidos, para a comparação precisa dessas características, mesmo quando medidas em diferentes escalas.

Em seguida, ocorre a etapa de classificação, na qual o modelo passa por treinamento com dados pré-processados, onde a máquina assimila padrões existentes nos dados rotulados, passando a desenvolver habilidades de prever

## Faculdade de Tecnologia de Americana "Ministro Ralph Biasi"

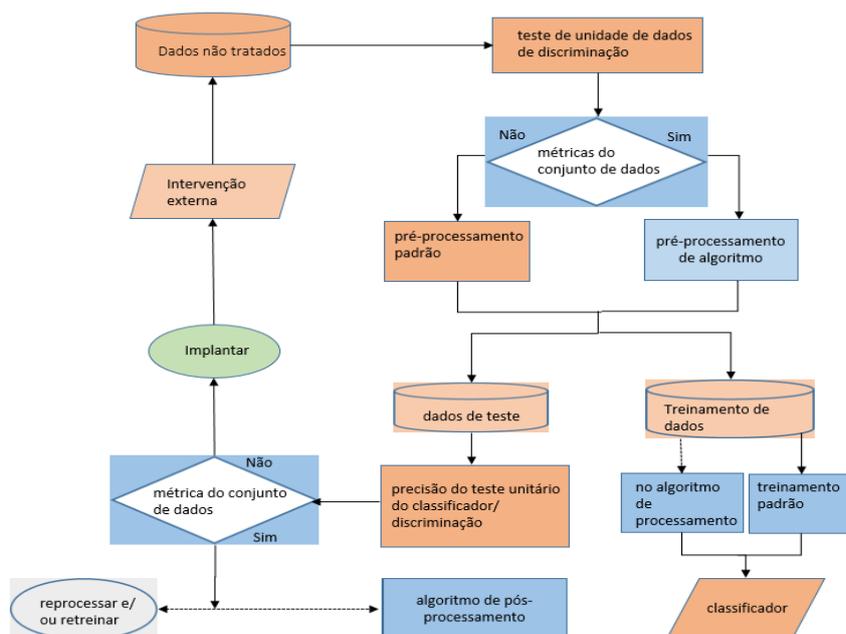
novas instâncias.

No centro da linha de processamento da *pipeline*, são apresentadas as métricas de conjunto de dados que são ferramentas utilizadas para avaliar a equidade e a imparcialidade em máquinas de inteligência artificial. Sendo assim, é possível detectar discriminação algorítmica em diversos aspectos como, gênero, raça, religião entre outros, e assim ter uma análise mais precisa quanto a desigualdades e privilégios relacionadas a um certo grupo de pessoas.

Após a classificação dos dados ocorre o pós-processamento, onde são adotadas medidas de calibração do modelo para que as previsões sejam refletidas de forma precisa e confiável, assim como eventuais ajustes para reduzir vieses encontrados nos resultados.

Com base na explicação acima, a Figura 1 destaca como os algoritmos discriminatórios podem se infiltrar no processo de aprendizagem de máquina em etapas distintas da *pipeline*.

**Figura 1: Viés no *pipeline* de aprendizado de máquina**



Fonte: Baseado em AIFairness (2020)

---

## Faculdade de Tecnologia de Americana "Ministro Ralph Biasi"

Caso ainda haja detecção de algoritmos discriminatórios durante o pós-processamento, a máquina passa por um novo processo de treinamento para diminuir ou eliminar esses vieses com medidas específicas como ajustes de parâmetros ou estratégias de nova amostragem dos dados. Por isso, o AIF360 oferece uma abordagem completa para promover a equidade em todas as fases do desenvolvimento de modelos de IA proporcionando ferramentas práticas para mitigar vieses e garantir sistemas mais justos e éticos.

### 3. Processo de Implementação

Com base na explicação anterior os algoritmos são divididos em três classes: que é a do pré-processamento que é utilizado os métodos *reweighing*, termo esse que pode ser entendido como uma (nova pesagem), ou *Impact Disparate Remove* (Removedor de impacto de Disparidade). Quanto ao *Adversarial debiasing* (Desvio Adversário) se refere a uma técnica de processamento dos dados. E por fim o *Calibrated Equalized Odds PostProcessing* (Pós-processamento Calibrado para Igualdade de Chances) que como o nome já diz é uma abordagem do pós-processamento. Importante ressaltar que existem outros meios de mitigação de viés que fazem parte da ferramenta e apenas alguns estão sendo abordados por conta da extensão de variedades de métricas disponibilizadas para a implementação.

Segundo Kamiran e Calders (2012) o *reweighing* é uma técnica de pré-processamento de dados que modifica os pesos atribuídos a cada instância dependendo de cada grupo ou rótulo sem mudar as categorias originais, com o intuito de reduzir as diferenças estatísticas e garantir que a probabilidade de resultados positivos entre os grupos seja preservada.

## Faculdade de Tecnologia de Americana "Ministro Ralph Biasi"

Sendo assim, a *reweighing* pode ser aplicada por definições de grupos privilegiados que são os candidatos majoritários e os não privilegiados que é o grupo minoritário.

A partir disso são utilizados métodos para calcular os pesos de forma mais igualitária em relação a cada instância, treinando o modelo e aplicando transformações nos dados para manter a equidade antes da classificação. Os métodos utilizados são:

O *fit* que calcula os pesos de cada exemplo com o intuito de balancear a distribuição de cada instância. Outro procedimento complementar é o *fit\_predict* que após o cálculo treina o modelo para fazer previsões justas. Quanto ao *fit\_transform* após a medição faz o ajuste dos pesos nos dados. Além do *predict* que usa os pesos que já sofreram ajustes para garantir que as previsões não contenham vieses e sejam mais igualitárias. E por fim, o método *transform* que é utilizado para ajustar os dados originais para equilibrar a distribuição entre os grupos.

Outro algoritmo específico aplicado durante o pré-processamento dos dados é o *Impact Disparate Remover* (Removedor de impacto de disparidade) que de acordo com Feldman *et al.* (2015), é aplicado para ajustar os valores dos dados sem alterar a classificação de cada grupo com o intuito de tornar as decisões mais justas entre diferentes grupos. Ou seja, o algoritmo é um removedor de viés que se baseia na métrica Impacto de disparidade que é utilizada para avaliar as desigualdades nos dados e uma pequena redução indica que o viés está sendo mitigado.

Conforme Zhang *et al.* (2018) na etapa de processamento é aplicado o *Adversarial debiasing* (Desvio Adversário) que é uma técnica que ensina um classificador a fazer previsões precisas sem revelar informações sensíveis como raça ou gênero. Pode ser citado como exemplo de utilização do método Desvio

## Faculdade de Tecnologia de Americana "Ministro Ralph Biasi"

Adversário o ensinamento de um sistema de reconhecimento facial onde o intuito é reconhecer rostos sem distinção entre raças. Desta forma o modelo se torna mais justo por não aderir características discriminatórias já inseridas na sociedade.

Quanto a etapa final *Calibrated Equalized Odds PostProcessing* (Pós-processamento Calibrado para Igualdade de Chances) segundo Pleiss *et al.* (2017) é uma técnica de pós-processamento utilizada após o treinamento da máquina com o intuito de ajustar as previsões para garantir a igualdade entre os grupos corrigindo vieses presentes nas previsões.

### 4. Métricas de equidade

A AIF360 possui muitas métricas de equidade por se tratar de um código aberto que está em constante atualização. Por essa razão, serão analisadas somente as principais, com o intuito de avaliar a eficácia na redução de vieses dessas métricas em modelos de aprendizagem de máquina.

Segundo Feldman *et al.* (2015) o *Disparate Impact Ratio* (impacto de disparidade) ajusta os valores dos atributos para tornar os grupos mais justos, enquanto mantém a ordem de classificação dentro de cada grupo. Um valor próximo de 1 significa oportunidades igualitárias entre os grupos, sendo especificada pela fórmula.

$$\frac{Pr(\hat{Y} = \text{pos\_label} | D = \text{unprivileged})}{Pr(\hat{Y} = \text{pos\_label} | D = \text{privileged})}$$

Portanto é necessário colocar o atributo protegido como binário onde o valor 1 é atribuído para grupos privilegiados e 0 aos grupos não privilegiados. Assim, como a coluna de rótulos onde são atribuídos o valor 1 para resultado positivo e 0 para resultados não satisfatórios.

## Faculdade de Tecnologia de Americana "Ministro Ralph Biasi"

A métrica *Statistical Parity Difference* (Diferença de Paridade Estatística) de acordo com Aasheim *et al.* (2020) significa que tanto os indivíduos do grupo privilegiado quanto do grupo desprivilegiados devem ter a mesma chance de receber uma classificação favorável. Um valor ideal é zero e um valor negativo significa que o grupo desprivilegiado está em desvantagem, como indicado pela fórmula.

$$Pr(Y = 1|D = \text{unprivileged}) - Pr(Y = 1|D = \text{privileged})$$

Em relação a *Average Odds Difference* (Diferença de Oportunidade Igual) conforme a Blow *et al.* (2023) a métrica é utilizada para a verificação, onde todos os envolvidos possuem chances igualitárias de beneficiamento, por intermédio da taxa de verdadeiro positivo que mostra de forma precisa se o resultado positivo é reconhecido em todos os grupos. Desta forma, o valor 0 significa uma equidade perfeita e um valor positivo indica preconceito a grupos não privilegiados e o negativo aos privilegiados.

Quanto a *Average Odds Difference* (Diferença Média de Chances) segundo Pandraju *et.al.* (2021) a métrica compara as taxas de falsos positivos que seria a classificação de uma amostra de forma errada e verdadeiros positivos que significa que a mostra é de fato verdadeira entre os grupos privilegiados e os não privilegiados. O valor ideal é 0.

### 5. Análise da Implementação da ferramenta Ai Fairness 360

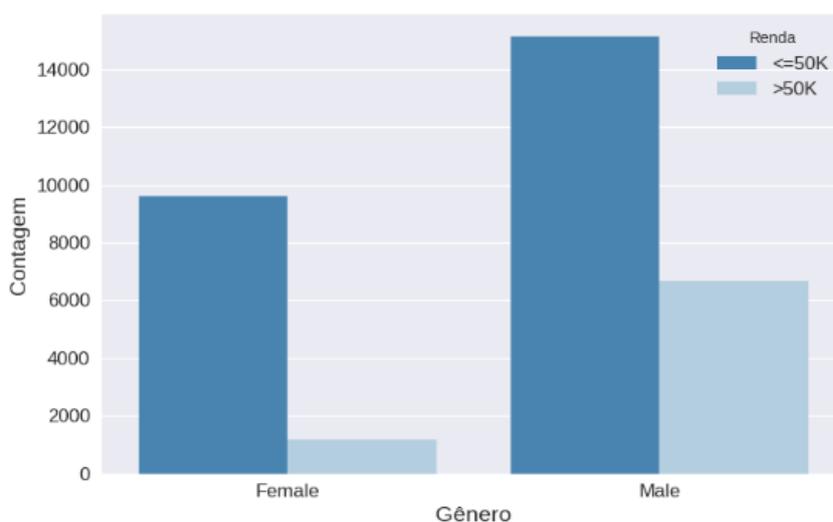
Neste caso será analisado o conjunto de dados *Adult Census Income* por Kohavi *et al.* (1994) com o intuito de prever se a renda anual de uma pessoa ultrapassa os 50 mil dólares. Com base nesse conjunto de dados a IA já consegue identificar atributos protegidos ou sensíveis por conseguir correlacionar padrões

## Faculdade de Tecnologia de Americana "Ministro Ralph Biasi"

que já estão embutidos na sociedade. Os dados informados possuem 15 atributos no total onde dois deles são sensíveis como raça e gênero.

Conforme apresenta a Figura 2 é perceptível a disparidade no resultado, onde com base na contagem dos dados, em torno de 6.000 pessoas do gênero masculino obteve melhor renda acima dos R\$50 mil dólares em comparação com o gênero feminino que ficou abaixo da contagem de 2.000 pessoas. Deixando em evidência a desvantagem do gênero feminino em relação a renda com oportunidades mais justas e equitativas.

Figura 2 - Atributos com viés gênero feminino e masculino

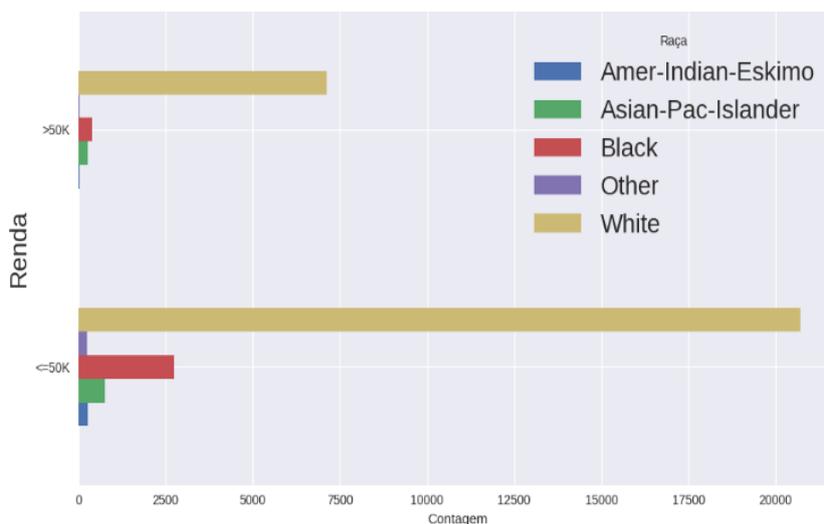


Fonte: Próprio Autor. Baseado no conjunto de dados *Adult Census Income*.

Com base na figura 3 é evidente a desigualdade ressaltando uma remuneração melhor de pessoas brancas onde em torno de 7.000 cidadãos americanos recebem em média uma renda de 50 mil dólares anualmente comparada a outras etnias que ficaram em uma notável disparidade reforçando assim a desigualdade entre as raças.

**Faculdade de Tecnologia de Americana "Ministro Ralph Biasi"**

**Figura 3 - Atributos com viés Raça**



**Fonte: Próprio Autor. Baseado no conjunto de dados *Adult Census Income*.**

A partir da identificação desses vieses é utilizada a métrica do impacto dispar que será aplicado antes do treinamento. Desta forma, o atributo sensível é transformado em número binário onde o valor 1 é para o grupo privilegiado e 0 para o grupo não privilegiado. Assim, como as colunas de rótulos que será atribuído o valor 1 para quem ganha mais \$50 mil por ano e 0 para quem ganha menos. Como a renda é a principal referência do conjunto de dados a coluna será utilizada como índice na análise dos atributos. Conforme a figura 4 são destacados os atributos gênero e raça os valores ficaram abaixo de 1 sendo detectado assim os vieses algoritmos.

**Figura 4 – Métrica de equidade impacto de disparidade**



**Fonte: Baseado na aplicação da métrica *Disparate impact*.**

## Faculdade de Tecnologia de Americana "Ministro Ralph Biasi"

### 6. Mitigação do viés

Com base na análise acima é possível realizar qualquer um dos métodos referente as categorias de pré-processamento, processamento ou pós-processamento para a mitigação dos vieses discriminatórios.

O procedimento aplicado foi o Desvio Adversário que é um método utilizado durante o processamento dos dados, para tornar o modelo preditivo mais justo. Deste modo, foi criada uma nova instância com o parâmetro `'pro_attr='sex'` pelo fato de gênero ser o atributo protegido onde o viés algoritmo deve ser mitigado. Quanto ao `'random_state'` tem o intuito de garantir resultados mais consistentes mesmo que haja aleatoriedade no treinamento do modelo o que ajuda a identificar e corrigir problemas no código.

As bibliotecas da ferramenta *Tensorflow* também foram importadas para o treinamento da máquina, onde o modelo de mitigação utiliza os dados de entrada dos atributos protegidos como `x_train`, e o `y_train`, que são usados como os rótulos para fazer previsões precisas ao mesmo tempo em que é ensinada a remover o viés relacionado ao gênero nos dados de análise. Baseado na explicação acima é apresentado na Figura 5 o código de aplicação para a redução do algoritmo discriminatório.

**Figura 5 – Aplicação do método Desvio Adversário**

```
✓ [504] from aif360.sklearn.inprocessing import AdversarialDebiasing
352 import tensorflow as tf
tf.compat.v1.disable_eager_execution()
adv_deb = AdversarialDebiasing(prot_attr='sex', random_state=RANDOM_STATE)
adv_deb.fit(X_train, y_train)
```

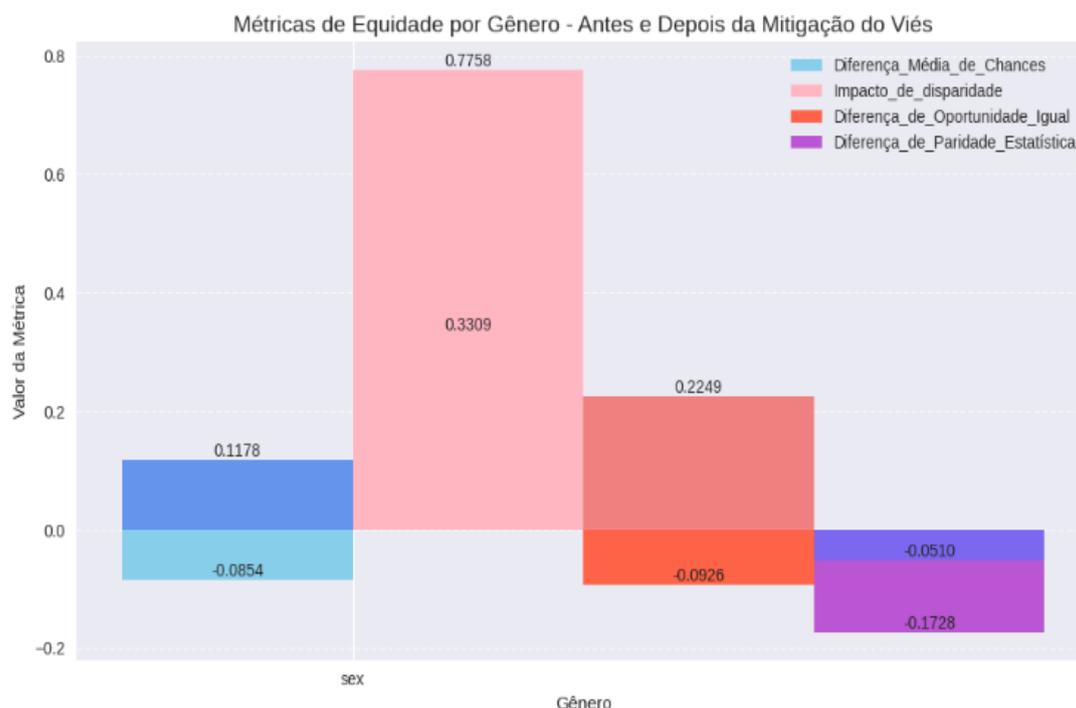
AdversarialDebiasing  
AdversarialDebiasing(prot\_attr='sex', random\_state=7654321)

Fonte: Próprio Autor. Baseado na aplicação do método *Adversarial Debiasing*.

## Faculdade de Tecnologia de Americana "Ministro Ralph Biasi"

A seguir é apresentado a figura 6 onde com base nos resultados é possível notar que o Desvio Adversário teve um efeito positivo na equidade, reduzindo o viés inicial em relação ao atributo sensível "gênero", já que no gráfico antes do uso do modelo de mitigação os valores das métricas ficaram abaixo de 0 e na métrica de impacto díspar mediu abaixo de 1 que seria o valor ideal, o que identifica viés nos algoritmos dos atributos sensíveis. Quanto ao gráfico analisado, após ao uso do Desvio Adversário como modelo de mitigação, indicou-se uma melhora nos resultados chegando mais próximo do desempenho recomendado. No entanto, ainda podem ser feitas melhorias com o intuito de alcançar uma equidade mais completa dos dados.

**Figura 6 – Antes e Depois da aplicação do método Desvio Adversário**



Fonte: Próprio autor.

## **Faculdade de Tecnologia de Americana "Ministro Ralph Biasi"**

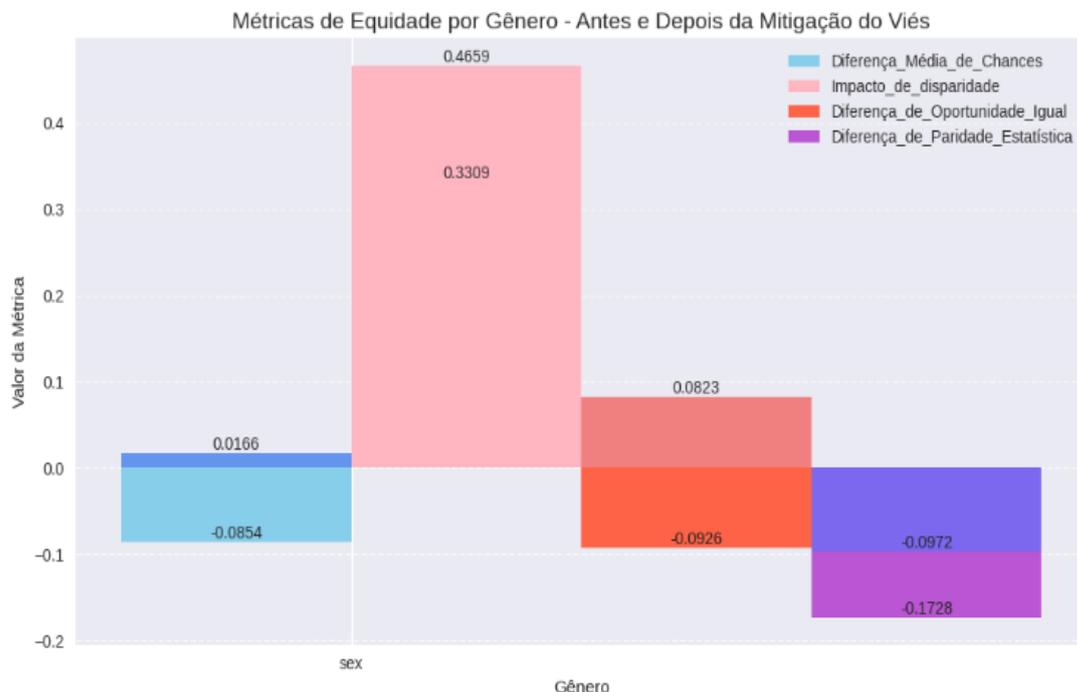
Outro método utilizado foi o pós-processamento calibrado para igualdade de chances, para reduzir o viés do atributo gênero. Antes da aplicação foi usado a regressão logística que de acordo com a IBM (2024) é um modelo estatístico utilizado para fazer previsões estimando características de um evento com base em variáveis independentes e como o resultado é uma probabilidade a variável dependente é classificada entre 0 e 1.

Desta forma a regressão logística é usada para prever se uma pessoa tem uma renda superior a R\$50.000,00 por ano ou outro valor de renda com base em características pessoais. Onde ter 50 mil é igual a 1 por ser a variável dependente e não ter 50 mil por ano é igual a 0. As variáveis independentes são os atributos do conjunto de dados, como idade, raça entre outros.

A partir desta métrica é implementada o modelo de mitigação Pós-processamento Calibrado para Igualdade de Chances onde é aplicado o conceito de falso positivo que baseado no conjunto de dados, ocorre quando uma pessoa do gênero feminino é classificada como tendo uma renda acima dos 50 mil que será representado pelo valor 1, quando na verdade a renda está abaixo deste resultado indicando o valor 0. No falso negativo é representado pelo 1 quando a renda é alta e 0 quando a renda estiver baixa. Conforme apresentado na figura 7 é perceptível uma melhora nos resultados em comparação com o antes e depois da redução do viés onde há uma redução mais ainda há resquícios da desigualdade no atributo protegido como indica a métrica Diferença de Paridade Estatística que alcançou o valor de -0,0972, porém o resultado ainda é promissor.

**Figura 7 – Antes e Depois da aplicação do método Pós-processamento Calibrado para Igualdade de Chances**

## Faculdade de Tecnologia de Americana "Ministro Ralph Biasi"



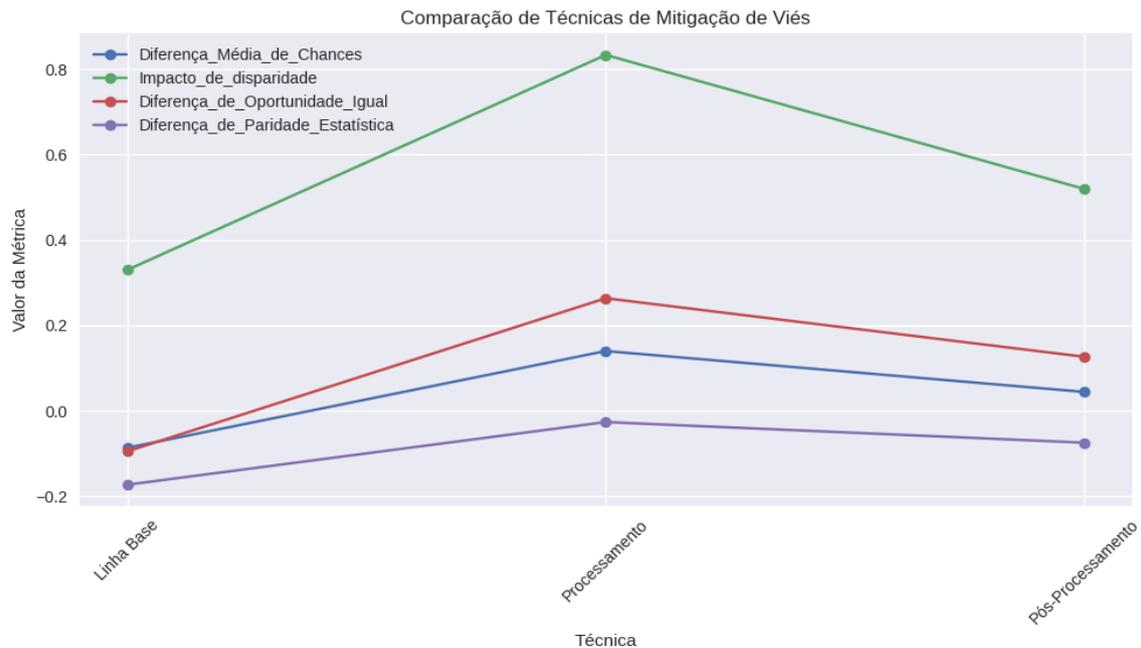
Fonte: Próprio autor

### 7. Comparação do resultado dos modelos de Mitigação do viés

A seguir é apresentado a figura 8 utilizando os dois modelos de redução de viés algoritmo como comparativo que é o Desvio Adversário e o pós-processamento calibrado para igualdade de chances onde com base nas técnicas utilizadas é notável a redução do viés em ambos os modelos que foram aplicados em etapas de processamento e pós-processamento dos dados como Desvio Adversário que teve 83% de precisão quanto a classificação correta em relação a classes de renda previstas e o Pós-processamento Calibrado para Igualdade de Chances com 81% de assertividade. Sendo assim, é perceptível que a fase do processamento teve um resultado um pouco melhor. Porém ambas conseguiram alcançar um modelo mais justo e imparcial e menos discriminatório.

## Faculdade de Tecnologia de Americana "Ministro Ralph Biasi"

Figura 8 - comparativo das métricas utilizadas



Fonte: Próprio autor.

## 8. Considerações finais

No artigo foi aplicado um estudo de caso de renda do kohavi (1994) e com base nos algoritmos discriminatórios encontrados, foram aplicados métodos de mitigação e métricas que permitiram uma compreensão mais clara quanto aos impactos da reprodução de vieses em máquinas preditivas, destacando a importância de reduzir preconceitos nos dados garantindo uma imparcialidade mais equilibrada quanto a tomada de decisões estratégicas, proporcionando melhores resultados para os grupos envolvidos.

## Faculdade de Tecnologia de Americana "Ministro Ralph Biasi"

Desta forma, a implementação da ferramenta AIF 360 é importante por contribuir em pesquisas de modelos preditivos e aprendizagem de máquina, pois, ajuda na identificação e correção de preconceitos. Além de contribuir na compreensão dos desafios éticos e também a melhorar as práticas no desenvolvimento de tecnologias mais responsáveis.

### REFERÊNCIAS

AASHEIM, Tor. H. *et al.* **BIAs mitigation with AIF360**: a comparative study. 2020. Bergen /Noruega: Universidade de Bergen. Disponível em: <https://bora.uib.no/bora-xmlui/handle/11250/2764230>. Acesso em: 03 nov. 2023.

BLOW, C. H. *et al.* Comprehensive validation on reweighting samples for BIAs mitigation via AIF360. 2023. arXiv: 2312.12560 [cs.LG]. Disponível em: <https://arxiv.org/pdf/2312.12560> Acesso em 31 mar. 2024

FELDMAN, M. *et al.* AIF360 **Algorithms**: disparateImpactRemover. Documentação do AIF360 0.6.0. 2023. Disponível em: [https://aif360.readthedocs.io/en/latest/modules/generated/aif360.algorithms.pre\\_processing.DisparateImpactRemover.html#id1](https://aif360.readthedocs.io/en/latest/modules/generated/aif360.algorithms.pre_processing.DisparateImpactRemover.html#id1). Acesso em: 31 mar. 2024.

GOLDSCHMIDT, R. R. **Uma introdução à inteligência computacional**: fundamentos, ferramentas e aplicações. Rio de Janeiro: IST-Rio, 2010. Disponível em: [https://www.facterj-rio.edu.br/downloads/ebooks/ebook\\_Intro\\_Inte\\_comp.pdf](https://www.facterj-rio.edu.br/downloads/ebooks/ebook_Intro_Inte_comp.pdf). Acesso em: 31 mar.2024

HOLLINGSWORTH, S. **Preenchimento automático do Google**: um guia completo de SEO. Search Engine Journal, 17 Jan. 2022. Disponível em: <https://www.searchenginejournal.com/google-autocomplete-a-complete-seo-guide/251407/>. Acesso em: 21 abr. 2024.

IBM. SPSS Statistics SaaS: regressão logística 2023. Disponível em: <https://www.ibm.com/docs/pt-br/spss-statistics/saas?topic=regression-logistic>. Acesso em: 07 mai. 2024.

## Faculdade de Tecnologia de Americana "Ministro Ralph Biasi"

KAMIRAN, F.; CALDERS, T. **Data preprocessing techniques for classification without discrimination**: knowledge and information systems. 2012. Disponível em: <[https://aif360.algorithms.preprocessing.Reweighting — aif360 0.6.0 documentation](https://aif360.algorithms.preprocessing.Reweighting—aif360.0.6.0.documentation)>. Acesso em: 07 abr. 2024.

KOHAVI, R.; Becker, B. **Adult census income dataset**. Kaggle, 1994. United States: Census. Disponível em: <https://www.kaggle.com/datasets/uciml/adult-census-income/data>. Acesso em: 31 mar. 2024.

MAHONEY, T.; VARSHNEY, K. R.; HIND, M. **How to measure and reduce unwanted BIAs in machine learning**. 1. ed. United States of America: O'Reilly Media, Inc., 2020. Disponível em: <https://community.ibm.com/community/user/ai-datascience/blogs/trisha-mahoney/2020/04/09/how-to-measure-and-reduce-unwanted-bias-in-machine/>. Acesso em: 07 mai. 2024.

MARLON, B. Inteligência artificial revoluciona o universo do marketing. **Promarketing 360**, 31 out. 2023. Disponível em: <https://promarketing360.com.br/inteligencia-artificial-revoluciona-o-universo-do-marketing/>. Acesso em: 21 abr. 2024.

PANDRAJU, S.; GANESHA, S. AI fairness: uma breve introdução ao AI fairness 360. **Transformer NLP**. <https://transformernlp.medium.com/ai-fairness-a-brief-introduction-to-ai-fairness-360-b2e39c96ca49>. Acesso em 31 mar. 2024.

PLEISS, G., *et. al.* AIF360 **Algorithms: CalibratedEqualizedOdds**. Documentação do AIF360 0.6.0. 2023. Disponível em: <https://aif360.readthedocs.io/en/latest/modules/generated/aif360.sklearn.postprocessing.CalibratedEqualizedOdds.html>. Acesso em: 31 mar. 2024.

QUEIROZ, G. Diversidade em equipes torna inteligência artificial mais justa - **Mentalidades Matemáticas**. 2020. Disponível em: <https://mentalidadesmatematicas.org.br/diversidade-em-equipes-torna-inteligencia-artificial-mais-justa/>. Acesso em: 07 nov. 2023.

SAS. **Análises preditivas**: o que são e qual sua importância? 2023. Disponível em: [https://www.sas.com/pt\\_br/insights/analytics/analises-preditivas.html](https://www.sas.com/pt_br/insights/analytics/analises-preditivas.html). Acesso em: 07 nov. 2023.

SOARES, J. F. Análise preditiva: importância e vantagens para os negócios. 2018. 11) - **Anima Educação**, Curitiba, PR, 2018. Disponível em: <https://repositorio.animaeducacao.com.br/handle/ANIMA/12186> Acesso em: 07 de novembro de 2023

---

## Faculdade de Tecnologia de Americana "Ministro Ralph Biasi"

TENSORFLOW. TensorFlow - Guia de início rápido para iniciantes. [s.d.]. Disponível em: <https://www.tensorflow.org/tutorials/quickstart/beginner?hl=pt-BR>. Acesso em: 08 abr. 2024.

ZHANG, YUNFENG. *et al.* **AI fairness 360: an extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic BIA**. ArXiv: 1810.01943 [cs.AI]. 2018. Disponível em: <https://arxiv.org/abs/1810.01943/>. Acesso em 31 mar. 2024.