

# MO444 - Projeto Final - Resultado Final

Alan Motta Ganem  
RA 178777

*Instituto de Computação - Unicamp*

Novembro, 2021

## 1 Resumo

É sabido que existem processos históricos de exclusão na sociedade brasileira para recortes raciais e de gênero.

Vieses como este são sintomas de um processo histórico de exclusão, que por si só servem de insumo para o desenvolvimento de políticas de inclusão. Entretanto, eles dificultam uma análise causal mais profunda do efeito dos recortes racial e de gênero na renda média do indivíduo, uma vez que ocorre confundimento (viés do confundimento) entre as variáveis de interesse quando utilizamos os dados observacionais da PNADC do IBGE.

Entender o efeito causal do gênero e raça na renda média do indivíduo pode nos mostrar que, para além do processo histórico de exclusão, existe também um eco desse processo que permeia nossa sociedade atual, que faz com que indivíduos com histórias muito semelhantes, expostos às mesmas oportunidades de desenvolvimento durante a vida, mesma formação e habilidades técnicas, ainda assim possuam uma diferença salarial devido exclusivamente ao seu gênero ou raça. Conhecer o tamanho desse efeito pode auxiliar no desenvolvimento de estratégias para mitigá-lo.

A fim de determinar esse efeito causal, utilizou-se o Meta Learner T-Learner para inferência contrafactual de salários. Para o modelo de propensão foi utilizada uma regressão logística, por suas propriedades de calibração, e para o modelo de salários, foram utilizados Random Forests, uma para cada tratamento, como sugere o procedimento do T-Learner.

O resultado final mostra que mesmo excluindo os efeitos indiretos, frutos do confundimento, ainda assim persiste uma diferença salarial profunda entre diferentes recortes de gênero e raça.

## 2 Introdução

É sabido que existem vieses históricos na sociedade brasileira para recortes raciais e de gênero. Um exemplo muito evidente é que, enquanto 36% dos jovens brancos frequentam a universidade no Brasil, apenas 18% dos jovens pretos ou pardos também o fazem [10], ainda que seja 56% a proporção da população brasileira que se autodeclara negra segundo dados da Pesquisa Nacional por Amostra de Domicílios Contínua (PNADC), do IBGE [1]. Vieses como este são sintomas de um processo histórico de exclusão, que por si só servem de insumo para o desenvolvimento de políticas de inclusão. Entretanto, eles dificultam uma análise causal mais profunda do efeito dos recortes racial e de gênero na renda média do indivíduo, isso porque variáveis como "possui nível de formação superior" são preditoras tanto do nosso "tratamento" (recorte racial e de gênero) quanto da nossa variável resposta (salário), o que torna a variável "possui nível de formação superior" um confundidor entre o efeito causal do recorte racial e a renda média do indivíduo [4]. A proposta desse trabalho para a estrutura causal do problema é encontrada na figura 1.

O viés do confundimento é muito comum em estudos observacionais, diferentemente dos estudos aleatorizados controlados, em que são controladas certas variáveis a fim de evitar o viés do confundimento.

Entretanto, existem técnicas estatísticas que tentam diminuir os vieses nas estimativas de efeitos causais. Essas técnicas são oriundas de uma área de estudo chamada de inferência causal. O principal conceito que difere a inferência estatística e aprendizado de máquinas da inferência causal é conhecido como o problema fundamental da inferência causal [5]: para uma mesma unidade experimental que contem co-variantes  $\mathbf{X} = \mathbf{x}$ , é impossível observar a variável re-

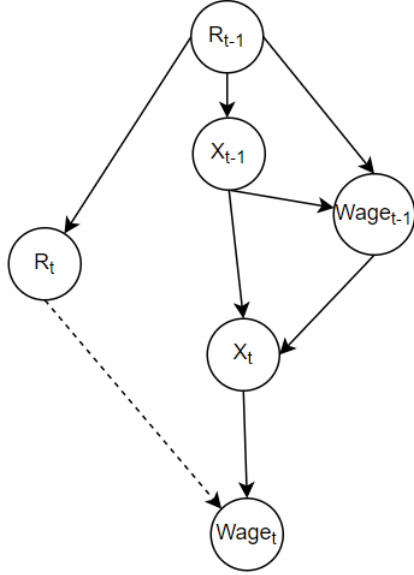


Figure 1: Representação gráfica da estrutura causal proposta pelo autor. Os sub-índices  $t$  e  $t-1$  representam a geração do indivíduo. As variáveis  $X$  são variáveis confundidoras/mediadoras, tais como maior nível de educação formal atingido, anos de educação, unidade federativa, se reside em região urbana ou rural... A variável  $R$  diz respeito ao tratamento (raça e/ou sexo) e a variável  $Wage$  é o salário. Setas pontilhadas dizem respeito ao mecanismo causal que estamos tentando inferir o efeito

sposta para os casos em que é aplicado um tratamento  $\mathbf{T}$  ( $\mathbf{Y}(\mathbf{T} = 1)$ ) e não é aplicado o mesmo tratamento ( $\mathbf{Y}(\mathbf{T} = 0)$ ), tornando impossível determinar então o efeito desse tratamento na variável resposta ( $\frac{\partial Y}{\partial T}(X = x)$  ou simplesmente  $Y(T = 1) - Y(T = 0)$  para tratamento binário).

Contudo, existem técnicas que tentam estimar esse efeito *contrafactual* [9] [3]. Contrafactual no sentido em que o resultado para a variável resposta observado é o factual, e agora gostaríamos de estimar qual seria o resultado dessa mesma variável caso o valor do tratamento fosse diferente.

Modelos de aprendizado de máquinas são capazes de identificar o efeito causal de um tratamento contanto que não haja viés do confundimento entre tratamento e resposta, o que sabemos ser dificilmente verdadeiro em estudos observacionais. Ainda assim, é possível utilizar modelos de aprendizado de máquinas genéricos, com algumas alterações, para que consigamos estimar efeitos causais.

### 3 Trabalhos relacionados

Existe uma vasta literatura abordando estudos de desigualdade salarial de gênero e raça, entretanto, esse espaço se torna mais restrito ao ponto em que buscamos por referências que tratem da temática sob a ótica da inferência causal e/ou do aprendizado de máquinas. Ainda assim, foram encontrados alguns estudos que se aproximam da temática mais específica desse trabalho. [causal'graph'approach'china] promove esforços para aproximar os estudos sociológicos de gênero e a inferência causal, utilizando-se de modelos de grafos, também conhecidos como modelos estruturais de causalidade. Esse estudo entretanto não faz considerações de recorte de raça. Já [7] por sua vez, utiliza uma abordagem preditiva com redes neurais densas identificar quais funcionários, dentro de uma empresa, podem estar susceptíveis a alguma viés de desigualdade salarial. Esse trabalho utiliza como classes sensíveis os recortes de gênero, raça e PCD (Pessoa Com Deficiência). Entretanto, nenhuma menção a inferência causal é encontrada no trabalho. Por fim [6] aborda as técnicas de ortogonalização e identificação causal do efeito de gênero e raça no salário sob a perspectiva da análise de mediação. O trabalho demonstra que métodos comuns de estimativa desses efeitos pode simplificar demais o processo de ortogonalização (ou exclusão do viés de confundimento e mediação) das variáveis sensíveis, propondo por fim maneiras de contornar esse problema.

### 4 Método

No trabalho foram utilizadas duas técnicas. A primeira dela foi uma implementação própria da ortogonalização através de resíduos [3], que como não trouxe bons resultados, não será abordado com afinho.

Desse modo, para o resultado final, foi utilizado o meta modelo **T-Learner** [9]. Esse meta modelo funciona da seguinte forma:

$$\hat{\tau}(X) = \hat{\mu}_t(X) - \hat{\mu}_c(X) \quad (1)$$

$$\hat{\mu}_i(X) = Y \sim X|T = i \quad (2)$$

Ou seja,  $\hat{\mu}_i(X)$  é um modelo arbitrariamente escolhido, treinado para prever  $Y$  a partir de  $X$  nos casos em que o tratamento é igual a  $i$ , e o efeito heterogêneo (dependente de  $X$ ) do tratamento  $\hat{\tau}(X)$  é

simplesmente a diferença entre a estimativa de Y feita pelo modelo do(s) tratamento(s) ( $\hat{\mu}_t(X)$ ) e a mesma estimativa feita pelo modelo do controle ( $\hat{\mu}_c(X)$ ).

Desse modo, temos que cada modelo  $\hat{\mu}_i(X = x)$  funciona como um contrafactual para uma unidade em que  $X = x$  e  $T = i$ .

As hipóteses que assumimos para a validade da estimativa gerada pelo T-Learner são:

1. **inconfundibilidade** (*unconfoundedness*): Os resultados potenciais (*potential outcomes*)  $Y_0, Y_1, \dots, Y_i$  são condicionalmente independentes da associação com T, dado X:

$$(Y_0, Y_1, \dots, Y_i) \perp T | X$$

2. **positividade/sobreposição** (*positivity*): o valor de propensão (propensity score) é maior que zero para todas as unidades e todos os tratamentos:

$$P(T_0, T_1, \dots, T_i) > 0$$

3. **Não interferência** (*no-interference*): o efeito do tratamento nas unidades é independente entre unidades, isso é, aplicar o tratamento em  $i$  não interfere o efeito do tratamento em  $j$ , para todo  $i \neq j$ :

$$Y_i(t_0, t_1, \dots, t_i, \dots, t_n) = Y_i(t_i)$$

No estudo final, cada tratamento será a combinação das variáveis binarizadas refrenets a sexo e recorte racial:

- Feminino E Branca ou Amarela
- Feminino E Preta, Parda ou Indígena
- Masculino E Branca ou Amarela
- Masculino E Preta, Parda ou Indígena

O caso base (ou controle) será a classe Masculino E Branca ou Amarela.

Além disso, será definido um *propensity-score* mínimo de 0.1 para todas as classes de tratamento, a fim de evitar extrapolação para regiões em que os tratamentos são raros e a estimativa se torna pouco confiável, mesmo que a hipótese de positividade ainda permaneça válida ( $P(T_0, T_1, \dots, T_n) > 0$ ).

## 5 Experimentos

### 5.1 Materiais

Todas as análises e treinos de modelo foram realizadas em um único computador, contendo 16GB de memória RAM e processador intel i7 oitava geração. Não houveram grandes problemas de processamento do alto volume de dados.

Para baixar e transformar os dados em um formato amigável, foi utilizada uma modificação do pacote PNADC disponível em <https://github.com/AlanGanem/pnadc>. A versão original do pacote pode ser encontrada em <https://github.com/ppnasser/pnadc>.

Para processamento de dados, foram utilizadas as bibliotecas Pandas (<https://github.com/pandas-dev/pandas>) e Dask (<https://github.com/dask/dask>).

Para visualização, foi utilizada a biblioteca saeorn (<https://github.com/mwaskom/seaborn>).

Para modelos de machine learning, foi utilizado o pacote scikit-learn (<https://github.com/scikit-learn/scikit-learn>).

A IDE utilizada para as análises foi o jupyter lab (<https://github.com/jupyterlab/jupyterlab>).

Todo o código utilizado nesse estudo está disponível em [https://github.com/AlanGanem/MO444/tree/master/trabalho\\_final](https://github.com/AlanGanem/MO444/tree/master/trabalho_final)

### 5.2 Dados utilizados

Nesse trabalho foram utilizados os microdados da Pesquisa Nacional por Amostra de Domicílios Contínua (PNADC) do Instituto Brasileiro de Geografia e Estatística (IBGE), disponíveis em [https://ftp.ibge.gov.br/Trabalho\\_e\\_Rendimento/Pesquisa\\_Nacional\\_por\\_Amostra\\_de\\_Domicilios\\_continua/Trimestral/Microdados/](https://ftp.ibge.gov.br/Trabalho_e_Rendimento/Pesquisa_Nacional_por_Amostra_de_Domicilios_continua/Trimestral/Microdados/). O subconjunto dos dados utilizados foi o de pessoas mais velhas que 14 anos que tem algum tipo de ocupação remunerada.

### 5.3 Seleção de features

Como teremos dois tipos de modelo, um de propensão e um de previsão dos targets, o processo de selecção de features foi feito levando em conta a intersecção das features selecionadas pelo processo abaixo, tendo como objetivo selecionar features com bom poder preditivo para uma regressão do salário e para classificação das classes sensíveis de recorte de gênero e raça.

Como o dataset em questão possui muitas features, o processo de seleção se deu em duas partes:

1. Filtro indutivo (selecionado pelo cientista de dados a priori)
2. Utilização de RandomForest + Permutation importances para filtro final, com as features selecionadas anteriormente [2].

As features finais utilizadas foram:

**UF** (estado), **V1022** (zona rural ou urbana), **V1023** (região do estado), **V3001** (sabe ler ou escrever?), **V4010-SubP** (subgrupo principal da ocupação de acordo com Estrutura-Ocupacao-COD), **V4010-Sub** (subgrupo principal da ocupação de acordo com Estrutura-Ocupacao-COD), **V4012** (Nesse trabalho,... era), **V4013-Sec** (seção do código CNAE da principal atividade do trabalho), **V4013-Div** (Divisão do código CNAE da principal atividade do trabalho), **V4014** (esfera do trabalho, Federal estadual ou municipal), **VD3004** (nível de instrução mais elevado alcançado), **V3009A** (curso mais alto que frequentou anteriormente), **V4029** (carteira assinada), **V2009** (idade), **VD3005** (anos de estudo), **VD4013** (Faixa das horas habitualmente trabalhadas)

## 5.4 Preprocessamento e Modelagem

O preprocessamento se deu da seguinte maneira: Variáveis numéricas serão normalizadas utilizando o RobustScaler do scikit-learn, mesmo o RandomForest não exigindo que as variáveis sejam normalizadas, ainda temos a regressão logística para o modelo de propensão. As variáveis categóricas serão codificadas usando OneHotEncoding.

A variável Y obedece uma distribuição lognormal, por isso, a tarefa de regressão será voltada para determinar o logaritmo natural do salário, ao invés do salário em si.

A implementação do meta modelo Tlearner foi feita inspirada no paradigma fit-transform, porém, incluindo métodos para inferência dos valores contrafactuais e para cálculo dos *propensity-score*, que é a propensão de uma unidade receber um tratamento  $i$ .

O algoritmo utilizado para determinar  $\hat{\mu}_i$  será um ensemble de árvores de decisão para regressão, implementado pelo scikit-learn (RandomForestRegressor). Para o modelo de *propensity-score* será utilizado a versão do classificador dessa mesma biblioteca (RandomForestClassifier). Os hiperparâmetros dos dois modelos serão os padrão da implementação, exceto por:

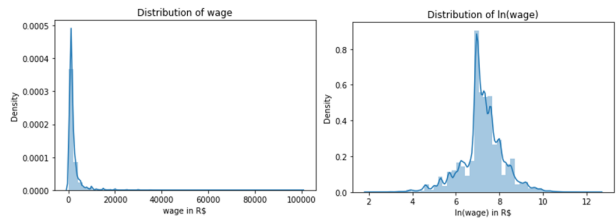


Figure 2: Distribuição de salário em escalas log e padrão: É possível observar que a distribuição do logaritmo do salário remonta uma distribuição normal, exceto pelo pico próximo de 7, que ocorre devido a existência de um salário mínimo, concentrando a distribuição nessa faixa.

1. `n_estimators` : 100
2. `min_samples_leaf` : 5
3. `n_jobs` : -1
4. `max_samples` : 0.2

Para o *propensity-score* será utilizada uma regressão logística, por sua capacidade de gerar distribuições de probabilidade bem calibradas com a distribuição real, uma propriedade muito importante para modelos de propensão.

## 6 Resultados

Como a inferência causal tenta inferir uma quantidade não mensurável (uma quantidade contrafactual), a validade dos resultados pode ser acessada através da verificação da validade das hipóteses relacionadas na seção 4.

### 6.1 Efeito global (*Average Treatment Effect*)

Garantindo a validade dessas hipóteses, observamos que de fato existe um viés de confundimento bastante pronunciado para o recorte racial (figura ??), evidenciando um processo histórico subjacente que determina a diferença salarial através da falta de oportunidades, como sugerido pelo modelo gráfico da figura 1.

### 6.2 Efeito heterogêneo (*Conditional Average Treatment Effect*)

É interessante também observar o efeito heterogêneo dos recortes raciais e de gênero para unidades com diferentes características.

Gender	Race	% wage	% wage - unbiased	% bias
Feminino	Branca ou Amarela	0.815	0.822	-0.006
	Preta, Parda ou Indígena	0.568	0.713	-0.146
Masculino	Branca ou Amarela	1.000	1.000	0.000
	Preta, Parda ou Indígena	0.720	0.884	-0.165

Figure 3: Efeito observado vs efeito causal estimado vs viés de confundimento.

Ao analisarmos o efeito causal das nossas variáveis sensíveis em função do salário real, observamos que a desigualdade se acentua para salários reais maiores, como observado na figura 4.

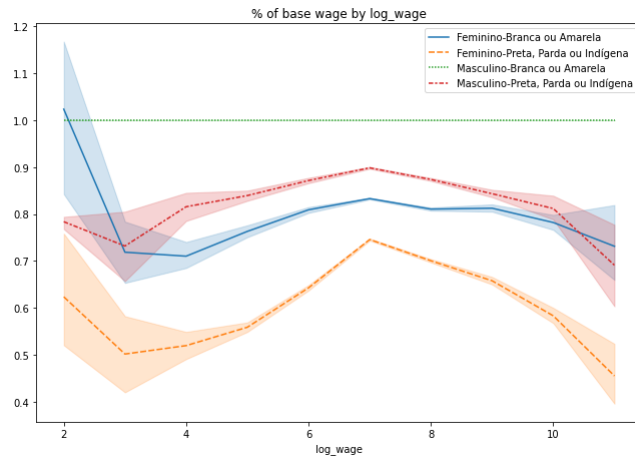


Figure 4: Percentual do salário referência em função de salário real: Para salários maiores, é possível observar que se acentua a diferença salarial entre homens brancos e os demais grupos.

Uma análise semelhante pode ser feita, levando em conta a idade (figura 5). É notável que para mulheres, a distância salarial se acentua a medida em que passam os anos, principalmente a partir dos 25 anos de idade. Essa tendência pode se dar em função da assimetria de respnsabilidade entre gêneros a partir do momento em que o casal tem um filho, algo que foi observado em [8].

Podemos observar também a desigualdade por estados da federação, nas figuras 7, 6 e 8

## 7 Conclusão

Do estudo, pode-se concluir que, uma vez garantidas as hipóteses necessárias para identificabilidade de efeitos causais, Meta Learners como o **T-Learner**

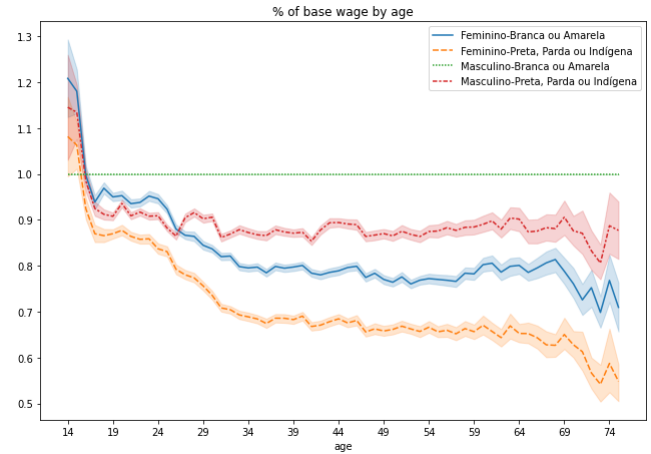


Figure 5: Percentual do salário referência em função de idade: A idade mais elevada acentua a distância salarial para as mulheres, principalmente a partir dos 25 anos.

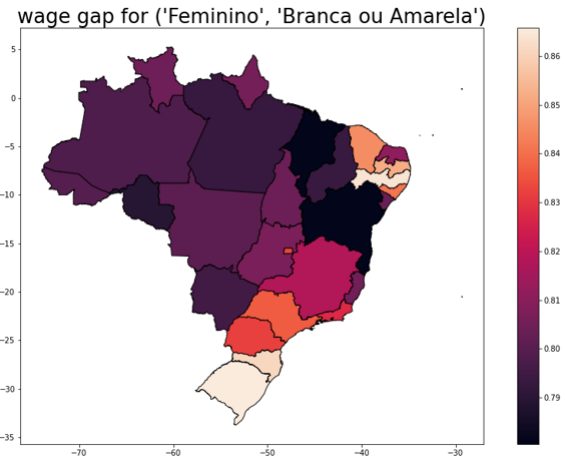


Figure 6: Desigualdade salarial causal por estado para mulheres brancas

pode ter uma serventia muito grande na análise causal de efeitos de desigualdade salarial. Uma dos maiores pontos de atenção desse estudo diz respeito à omissão de variáveis confundidoras, como por exemplo, que tipo de escola frequentou (público, privada, federal, militar), que tipo de faculdade frequentou (pública ou privada), escolaridade e condição socioeconômica dos pais... Todas essas variáveis não mensuradas na PNADc do IBGE podem ser fortes confundidoras dos efeitos medidos.

Por fim, o estudo mostra que de fato ainda existe uma grande lacuna de oportunidades, principalmente olhando o recorte racial. já olhando o recorte de gênero, a lacuna de oportunidade não é tão evidente, mas a desigualdade salarial é muito mais evi-

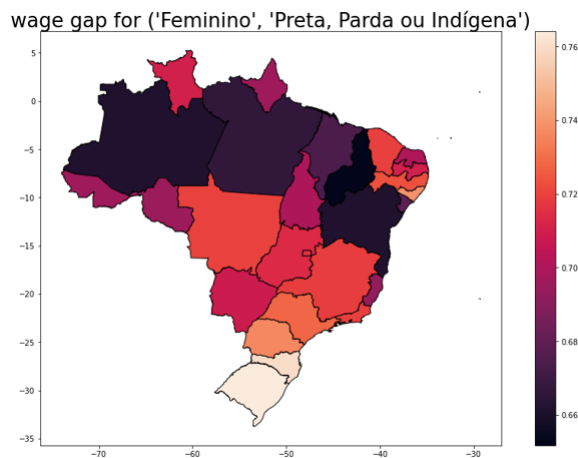


Figure 7: Desigualdade salarial causal por estado para mulheres não brancas

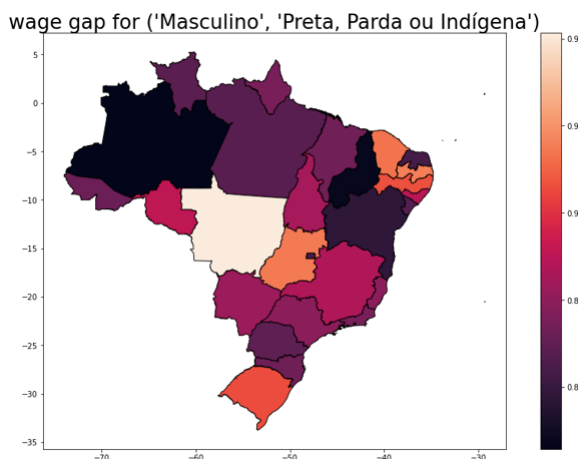


Figure 8: Desigualdade salarial causal por estado para homens não brancos

dente para unidades com oportunidades/convariantes semelhantes.

A conclusão é que o problema de desigualdade salarial no Brasil, tem várias origens e maneiras de se manifestar, seja por vias indiretas (desigualdade de oportunidades) ou vias diretas (discriminação salarial). Por esse motivo, políticas públicas e iniciativas da sociedade civil para mitigar esses efeitos tem uma grande possibilidade de prosperar caso leve em conta as especificidades de cada um desses tipos de desigualdade.

## References

- [1] Nathália Afonso. *Dia da Consciência Negra: números expõem desigualdade racial no Brasil*.

URL: <https://piaui.folha.uol.com.br/lupa/2019/11/20/consciencia-negra-numeros-brasil/>. (acessado em: 17.09.2021).

- [2] Nesrine Chehata, Li Guo, and Clément Mallet. *Airbone Lidar Feature Selection For Urban Classification Using Random Forests*. 2009.
- [3] Victor Chernozhukov et al. *Double/Debiased Machine Learning for Treatment and Structural Parameters*. 2016. eprint: [arXiv:1608.00060](https://arxiv.org/abs/1608.00060).
- [4] Matheus Facure. *Causal Inference for the Brave and True*. URL: <https://matheusfacure.github.io/python-causality-handbook/04-Graphical-Causal-Models.html>. (acessado em: 17.09.2021).
- [5] Paul W. Holland. *Statistics and Causal Inference*. 1986.
- [6] Martin Huber. *Causal Pitfalls in the Decomposition of Wage Gaps*. 2014.
- [7] Hamid R. Karimian et al. *A Machine Learning Framework to Identify Employees at Risk of Wage Inequality: U.S. Department of Transportation Case Study*. 2019.
- [8] Henrik Kleven, Camille Landais, and Jakob Egholt Sogaard. *Children and gender inequality: evidence from Denmark*. 2018.
- [9] Sören R. Künzela et al. *Meta-learners for Estimating Heterogeneous Treatment Effects using Machine Learning*. 2019. eprint: [arXiv:1706.03461](https://arxiv.org/abs/1706.03461).
- [10] Ana Carolina Moreno. *Taxa de jovens negros no ensino superior avança, mas ainda é metade da taxa dos brancos*. URL: <https://g1.globo.com/educacao/noticia/2019/11/06/taxa-de-jovens-negros-no-ensino-superior-avanca-mas-ainda-e-metade-da-taxa-dos-brancos.ghtml>. (acessado em: 17.09.2021).