



Gutenberg Assunção Vieira

# **Uso de aprendizado de máquina na avaliação de políticas públicas: uma revisão de escopo**

Brasília-DF

Setembro, 2023



Gutemberg Assunção Vieira

## **Uso de aprendizado de máquina na avaliação de políticas públicas: uma revisão de escopo**

Dissertação apresentada ao Programa de Mestrado Profissional em Avaliação e Monitoramento de Políticas Públicas da Escola Nacional de Administração Pública - ENAP como requisito para obtenção do título de Mestre em Avaliação e Monitoramento de Políticas Públicas.

Escola Nacional de Administração Pública – ENAP

Orientadora: Profa. Dra. Flávia Lúcia Chein Feres

Brasília-DF

Setembro, 2023

Ficha catalográfica elaborada pela equipe da Biblioteca Graciliano Ramos da Enap

---

V658u      Vieira, Gutemberg Assunção  
              Uso de aprendizado de máquina na avaliação de políticas  
              públicas: uma revisão de escopo / Gutemberg Assunção  
              Vieira. -- Brasília: Enap, 2023.  
              164 f. : il.

              Dissertação (Mestrado - Programa de Mestrado  
              Profissional em Avaliação e Monitoramento de Políticas  
              Públicas) -- Escola Nacional de Administração Pública, 2023.

              Orientação: Profa. Dra. Flávia Lúcia Chein Feres  
              1. Políticas Públicas. 2. Avaliação de Políticas Públicas. 3.  
              Aprendizado de Máquina. 4. Programação. I. Título. II. Feres,  
              Flávia Lúcia Chein orient.

CDD 320.6

---

Bibliotecária: Kelly Lemos da Silva – CRB1/1880



**ATA DA BANCA DE DEFESA DE TRABALHO DE CONCLUSÃO DO CURSO  
DE MESTRADO PROFISSIONAL EM AVALIAÇÃO E MONITORAMENTO  
DE POLÍTICAS PÚBLICAS**

**Aluno (a):** Gutemberg Assunção Vieira

**Ano de Ingresso:** 2021

**Título da Dissertação:** Uso de aprendizado de máquina na avaliação de políticas públicas: uma revisão

**Orientador:** Profa. Dra Flávia Lúcia Chein Feres

**Avaliadores:** Prof. Dr. Paulo de Martino Jannuzzi

Prof. Dr. João Gabriel de Moraes Souza

Avaliação:

Aprovado

Não aprovado. Reapresentação agendada para \_\_\_\_/\_\_\_\_/\_\_\_\_.

Brasília, 18 de setembro de 2023

DocuSigned by:  
*PAULO DE MARTINO JANNUZZI*  
345174BF8A58495...

Avaliador

DocuSigned by:  
*Joao Gabriel de Moraes Souza*  
33498CE63B0F4F8...

Avaliador

DocuSigned by:  
*Flavia Lucia Chein Feres*  
B81BDDCZAC5B461...

Orientadora



# Agradecimentos

Agradeço especialmente à minha esposa pelo apoio incondicional e compreensão dos momentos de dedicação necessários para participação no programa de mestrado e desenvolvimento deste projeto.

Agradeço aos meus pais por terem provido os meios e incentivos para que eu pudesse trilhar o caminho da educação.

Agradeço à minha orientadora pela confiança e pelos momentos de discussão, compreensão e direcionamento para consecução do trabalho.

Agradeço aos colegas de trabalho, principalmente àqueles que atuaram na etapa de seleção de estudos, dando mais robustez ao resultado final. Ainda no âmbito institucional, agradeço aos meus superiores, presentes e passados, que contribuíram com o meu desenvolvimento.

Agradeço aos avaliadores da banca, professores do curso e colegas de classe pelos ensinamentos compartilhados.

Finalmente, agradeço a todos colaboradores da ENAP pela oportunidade e por contribuírem com o aprimoramento da qualidade do serviço público por meio da oferta de cursos de alta qualidade.





*“Technology is capable of doing great things.  
But it doesn’t want to do great things.  
It doesn’t want anything.  
That part takes all of us”.*  
*(Tim Cook)*



# Resumo

Cada vez mais o aprendizado de máquina tem sido aplicado para desempenhar atividades que requeriam a execução por seres humanos, inclusive no âmbito governamental. Nesse contexto, questiona-se como ocorre a utilização desse tipo de ferramenta na avaliação de políticas públicas e sobre quais intervenções já foram aplicadas. Dito isso, o objetivo desta dissertação é o mapeamento de evidências de uso de aprendizado de máquina na avaliação de políticas públicas com a finalidade de ser utilizado como fonte de referência por pesquisadores, profissionais e outros interessados no assunto. Foi utilizada a metodologia de revisão de escopo com implementação de revisão cegada para redução de viés na seleção dos estudos. Ao final foram escolhidos 64 estudos para mapeamento, agrupamento, sumarização e reporte nas perspectivas de métodos de aprendizado de máquina empregados, objetos avaliados e características gerais das publicações. Confirmou-se que se trata de uma área de aplicação recente em termos de publicação, com a maioria dos estudos concentrados nos últimos cinco anos, e não abrange ações da maioria das áreas governamentais. Não foi identificado estudo no Brasil dentro do escopo definido e estratégia de busca implementada, destacando-se a oportunidade de pesquisa aplicada no tema. Os métodos de aprendizado de máquina foram codificados quanto ao objetivo de aplicação, resultando em nove formas de aplicação. Todos os dados mapeados, codificações criadas, bem como outras menções às tecnologias empregadas foram incluídos no texto e nos apêndices para referências futuras.

**Palavras-chave:** avaliação de política pública, aprendizado de máquina.



# Abstract

Machine learning has increasingly been applied to perform activities that required human beings to perform them, including government services. In this context, it is questioned how this type of tool is used in the evaluation of public policies and in which interventions have already been applied. Therefore, the objective of this dissertation is to map evidence of the use of machine learning in the evaluation of public policies in order to be used as a reference source by researchers, professionals and others interested in the subject. The scope review methodology was used and blinded review was applied to reduce bias in the selection of studies. In the end, 64 studies were chosen for mapping, grouping, summarizing and reporting from the perspectives of machine learning methods employed, evaluated objects and general characteristics of the publications. It was confirmed that this is a recent area of application in terms of publication, mostly of the them published in the last five years, and does not cover actions in most government areas. No study was identified in Brazil within the defined scope and search strategy, highlighting the opportunity for applied research on the topic. The machine learning methods were coded according to the application objective, resulting in nine ways of application. All mapped data, coding created, as well as other mentions to the technologies employed were included in the text and in the appendices for future reference.

**Keywords:** public policy evaluation, machine learning.



# Lista de ilustrações

|   |    |
|---|----|
| Figura 1 – Fluxograma de decisão de seleção de estudos . . . . .  | 44 |
| Figura 2 – Esquema da etapa de seleção com revisão cegada . . . . .   | 46 |
| Figura 3 – Quantidade de estudos por ano de publicação . . . . .  | 52 |
| Figura 4 – Quantidade de estudos por ano de publicação (selecionados e bola de neve) . . . . .  | 52 |
| Figura 5 – Quantidade de estudos por país de afiliação do autor . . . . .   | 53 |
| Figura 6 – Quantidade de estudos por objetivo . . . . .   | 54 |
| Figura 7 – Nuvem de palavras dos títulos . . . . .  | 56 |
| Figura 8 – Nuvem de palavras da descrição das intervenções avaliada . . . . .   | 57 |
| Figura 9 – Quantidade de estudos por função de governo da intervenção avaliada . . . . .  | 58 |
| Figura 10 – Quantidade de estudos pelo tamanho do conjunto de dados . . . . .   | 60 |
| Figura 11 – Quantidade de estudos pelo tipo de análise (causal ou não) . . . . .  | 62 |
| Figura 12 – Quantidade de estudos por objetivo da aplicação de aprendizado de máquina . . . . .   | 62 |
| Figura 13 – Quantidade de estudos (bola de neve) por objetivo da aplicação de aprendizado de máquina da referência originária . . . . . | 67 |
| Figura 14 – Quantidade de estudos por tipo de aprendizado do método utilizado . . . . .   | 67 |
| Figura 15 – Quantidade de estudos por métodos utilizados . . . . .  | 69 |
| Figura 16 – Outros métodos utilizados . . . . .   | 70 |
| Figura 17 – Quantidade de estudos por linguagem de programação . . . . .  | 73 |





# Lista de quadros

|   |     |
|---|-----|
| Quadro 1 – Expressão inicial . . . . .                                      | 34  |
| Quadro 2 – Primeiro teste no Scopus . . . . .                               | 37  |
| Quadro 3 – Teste Scopus incluindo o termo programa . . . . .                | 38  |
| Quadro 4 – Teste Scopus incluindo o termo “avaliação de programa” . . . . . | 38  |
| Quadro 5 – Teste Scopus incluindo o termo “avaliação de política” . . . . . | 39  |
| Quadro 6 – Teste Scopus incluindo o termo “programa AND política” . . . . . | 40  |
| Quadro 7 – Critérios de inclusão e exclusão . . . . .                       | 47  |
| Quadro 8 – Categorização de pacotes . . . . .                               | 74  |
| Quadro 9 – Classificação por características . . . . .                      | 101 |
| Quadro 10 – Classificação por objeto . . . . .                              | 111 |
| Quadro 11 – Classificação por método . . . . .                              | 127 |
| Quadro 12 – Estudos eliminados na elegibilidade . . . . .                   | 155 |
| Quadro 13 – Novos estudos decorrentes de exploração das citações . . . . .  | 161 |



# Lista de tabelas

|   |    |
|---|----|
| Tabela 1 – Quantidade de estudos encontrados por pesquisa realizada . . . . . | 41 |
|---|----|



# Lista de abreviaturas e siglas

|       |  |
|-------|--|
| ACM   | <i>Association for Computing Machinery</i>               |
| AGL   | Alexandre Gomide Lemos                                   |
| AJRF  | Arthur José Rodrigues Farias                             |
| ANN   | <i>Artificial Neural Network</i>                         |
| ANOVA | Análise de variância                                     |
| CATE  | Efeitos de tratamento médio condicional                  |
| DBB   | Demétrius Batista Borges                                 |
| DL    | <i>Deep learning</i>                                     |
| ESP   | Eduardo Soares de Paiva                                  |
| FLCF  | Flávia Lúcia Chein Feres                                 |
| FSP   | Fernando Sola Pereira                                    |
| GATE  | Efeitos de tratamento médio condicional por grupo        |
| GAV   | Gutemberg Assunção Vieira                                |
| HCB   | Hermann Cortizo de Biagi                                 |
| HIV   | Vírus da imunodeficiência humana                         |
| HLRO  | Hamilton Luiz Rodrigues de Oliveira                      |
| IA    | Inteligência artificial                                  |
| IATE  | Efeitos de tratamento médio condicional individualizado  |
| IEEE  | <i>Institute of Electrical and Electronics Engineers</i> |
| LASSO | <i>Least Absolute Selection and Shrinkage Operator</i>   |
| LATE  | Efeitos médios de tratamento local                       |
| LQTE  | Efeitos de tratamento local por quantil                  |
| MC    | <i>Matrix Completion</i>                                 |

|       |   |
|-------|---|
| ME    | Ministério da Economia                          |
| ML    | <i>Machine learning</i>                         |
| MS    | <i>Microsoft</i>                                |
| NATE  | Efeitos médios de tratamento não-paramétrico    |
| nATE  | Efeitos médios de tratamento naturais           |
| ODA   | <i>Optimal Discriminant Analysis</i>            |
| PCC   | População, Conceito e Contexto                  |
| rATE  | Efeitos médios de tratamento aleatórios         |
| RD    | Rodrigo Dewes                                   |
| RIS   | <i>Research Information Systems file format</i> |
| RNM   | Ricardo Nagamine Motta                          |
| SAS   | <i>Statistical Analysis System</i>              |
| SC-EN | <i>Synthetic Controls with Elastic Net</i>      |
| SMMB  | Stella Mendes Meireles Bonifácio                |
| SOF   | Secretaria de Orçamento Federal                 |

# Sumário

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>INTRODUÇÃO</b>   | <b>23</b> |
| <b>2</b> | <b>REFERENCIAL TEÓRICO</b>  | <b>25</b> |
| 2.1      | Aprendizado de máquina  | 25        |
| 2.2      | Avaliação de políticas públicas                                   | 26        |
| 2.3      | Avaliação de políticas públicas com uso de aprendizado de máquina | 27        |
| <b>3</b> | <b>METODOLOGIA</b>  | <b>31</b> |
| 3.1      | Identificação da questão de pesquisa                              | 32        |
| 3.2      | Identificação dos estudos relevantes                              | 33        |
| 3.3      | Seleção de estudos  | 34        |
| 3.4      | Mapeamento dos dados  | 35        |
| 3.5      | Agrupamento, sumarização e reporte dos resultados                 | 35        |
| <b>4</b> | <b>IDENTIFICAÇÃO, SELEÇÃO E MAPEAMENTO</b>                        | <b>37</b> |
| 4.1      | Identificação dos estudos relevantes                              | 37        |
| 4.1.1    | Refinamento da expressão de busca                                 | 37        |
| 4.1.2    | Buscas nas bases de dados   | 41        |
| 4.2      | Seleção de estudos  | 44        |
| 4.3      | Mapeamento dos dados  | 48        |
| <b>5</b> | <b>RESULTADOS</b>   | <b>51</b> |
| 5.1      | Características dos estudos                                       | 51        |
| 5.2      | Objetos de avaliação  | 56        |
| 5.2.1    | Intervenção   | 56        |
| 5.2.2    | Conjunto de dados   | 59        |
| 5.3      | Métodos de aprendizado de máquina                                 | 61        |
| 5.3.1    | Objetivo da aplicação de aprendizado de máquina                   | 61        |
| 5.3.2    | Métodos de aprendizado de máquina                                 | 66        |
| 5.3.3    | Linguagens e recursos de programação                              | 72        |
| <b>6</b> | <b>CONSIDERAÇÕES FINAIS</b>                                       | <b>75</b> |
|          | <b>REFERÊNCIAS</b>  | <b>79</b> |

|  |            |
|--|------------|
| <b>APÊNDICES</b>   | <b>99</b>  |
| <b>APÊNDICE A – CLASSIFICAÇÃO POR CARACTERÍSTICAS . . .</b>                              | <b>101</b> |
| <b>APÊNDICE B – CLASSIFICAÇÃO POR OBJETO . . . . .</b>                                   | <b>111</b> |
| <b>APÊNDICE C – CLASSIFICAÇÃO POR MÉTODO . . . . .</b>                                   | <b>127</b> |
| <b>APÊNDICE D – ESTUDOS ELIMINADOS NA ELEGIBILIDADE .</b>                                | <b>155</b> |
| <b>APÊNDICE E – NOVOS ESTUDOS DECORRENTES DE EXPLO-<br/>RAÇÃO DAS CITAÇÕES . . . . .</b> | <b>161</b> |



# 1 Introdução

O aprendizado de máquina tem sido utilizado em diversas áreas por meio da criação de sistemas que aprendem com o conhecimento prévio sobre atividades antes desempenhadas apenas por seres humanos. Diferentemente do modelo tradicional de programação, por meio da qual o programador codifica explicitamente todos os passos que o computador deve seguir, o sistema aprende as regras a partir dos dados.

Sistemas assim são empregados, por exemplo, para direção de automóveis, recomendações de compras e entretenimento, reconhecimento de imagens, geração de textos, entre outros. No âmbito governamental, observa-se exemplos de uso pelo poder público como triagem de denúncias<sup>1</sup>, análise de prestações de contas<sup>2</sup> e diversas atividades do judiciário<sup>3</sup>.

Nesse contexto, questiona-se como ocorre a utilização desse tipo de ferramenta na avaliação de políticas públicas e sobre quais intervenções já foram aplicadas. Há algumas evidências tanto na preparação dos dados quanto como método principal no processo, porém, não encontramos ainda uma referência abrangente que mapeie esse corpo de conhecimento, principalmente os casos concretos de aplicação.

Isso posto, o objetivo deste estudo é mapear evidências de uso de aprendizado de máquina na avaliação de políticas públicas. A finalidade principal é servir como fonte de referência para avaliadores de políticas públicas, seja no âmbito acadêmico, por profissionais da iniciativa privada ou servidores da administração pública, como os pertencentes a órgãos de avaliação de políticas públicas e de controle.

Assim, elaboramos a seguinte questão e subquestões para guiar a pesquisa:

1. Como o aprendizado de máquina tem sido utilizado na avaliação de políticas públicas?
  - a) Quais as características gerais dos estudos?
  - b) Quais os objetos de avaliação?
  - c) Quais métodos de aprendizado são aplicados?

O mapeamento de evidências de utilização de aprendizado de máquina na avaliação de políticas públicas pode contribuir para disseminar as técnicas utilizadas, servir como referência para pesquisadores e profissionais, bem como identificar lacunas que possam

<sup>1</sup> <[https://repositorio.cgu.gov.br/handle/1/66840?locale=pt\\_BR](https://repositorio.cgu.gov.br/handle/1/66840?locale=pt_BR)>

<sup>2</sup> <<https://www.gov.br/cgu/pt-br/assuntos/noticias/2022/12/malha-fina-do-fnde-e-vencedor-do-26o-concurso-inovacao-no-setor-publico>>

<sup>3</sup> <<https://www.conjur.com.br/dl/metade-cortes-brasileiras-projeto.pdf>>

ser exploradas. Nesse cenário, escolhemos a metodologia de revisão de escopo, que é uma espécie de revisão de literatura mais adequada ao mapeamento de evidências.

Ainda sobre a justificativa para realizar esse trabalho, além das oportunidades de contribuição acima, vale mencionar que o autor atua no Órgão Central do Sistema de Controle Interno do Poder Executivo Federal, que dentre suas competências constitucionais abrange a avaliação da execução de programas de governo. Em que pese a avaliação executada no âmbito da auditoria interna governamental frequentemente ser normativa, as demais pesquisas avaliativas também estão no campo de interesse da função. Ademais, o órgão compõe o Conselho de Monitoramento e Avaliação de Políticas Públicas, ao qual compete avaliar políticas públicas financiadas por gastos diretos ou subsídios da União. Dessa forma, a proximidade com o tema e potencial aplicação prática e aprimoramento das atividades da instituição empregadora motivam a realização do projeto.

Os demais capítulos do documento apresentam o referencial teórico sobre os conceitos envolvidos ([Capítulo 2](#)), a metodologia ([Capítulo 3](#)), o processo de identificação, seleção e mapeamento dos estudos ([Capítulo 4](#)), os resultados ([Capítulo 5](#)) e, finalmente, as considerações finais ([Capítulo 6](#)).

## 2 Referencial Teórico

Neste capítulo contextualizamos brevemente os principais conceitos relacionados ao objetivo deste trabalho, mais especificamente o aprendizado de máquina, a avaliação de política pública e exemplos de aplicação do primeiro para operacionalizar o segundo. O objetivo, portanto, é apenas trazer breves conceitos abordados no desenvolvimento do texto.

### 2.1 Aprendizado de máquina

Inicialmente, os sistemas de Inteligência Artificial (IA) eram construídos com a codificação explícita de regras para a resolução de tarefas específicas, por exemplo, jogar xadrez. O avanço da capacidade de processamento, disponibilidade de dados, entre outros, possibilitou a criação de modelos analíticos baseados no aprendizado de máquina (ML, do inglês *machine learning*). Os algoritmos de ML permitem automatizar a criação de modelos analíticos a partir de dados de treinamento, sem a necessidade de codificação explícita de regras, com o objetivo de executar tarefas cognitivas específicas, como detecção de objetos e processamento de linguagem (JANIESCH; ZSCHECH; HEINRICH, 2021).

A grosso modo, pode-se dizer que o sistema aprende com a experiência consignada em dados disponíveis. Uma opção é fornecer ao algoritmo dados históricos de tarefas previamente executadas contendo o gabarito da “resposta certa”, ou seja, exemplos de como fazer aquilo que se espera automatizar. Nesse caso, a máquina aprende pelos exemplos conhecidos, ao que se dá o nome de aprendizado supervisionado. Quando se tem dados sobre algo, mas não se conhece as respostas corretas, pode-se aplicar um algoritmo que tenta identificar grupos e relacionamentos a partir das características dos dados. Esse é um exemplo de aprendizado não-supervisionado. Há ainda algoritmos que se valem de dados rotulados e não-rotulados, chamados de semi-supervisionados (BI et al., 2019). Somada a esses, existe o aprendizado por reforço, que abrange algoritmos que avaliam o comportamento ótimo em determinado contexto para aprimorar sua eficiência, guiados por políticas de recompensa e penalização (SARKER, 2021).

A rede neural artificial (ANN, do inglês *artificial neural network*) é um algoritmo de aprendizado de máquina baseado no funcionamento do cérebro humano, constituída de neurônios artificiais que se comunicam por sinais e são organizados em redes com diferentes camadas. O número de camadas e outras propriedades, como a função de ativação que controla o processamento dos sinais entre neurônios, são definidas manualmente. Nesse contexto, é definido o aprendizado profundo (DL, do inglês *deep learning*), como um subcampo do ML. O DL abrange as redes neurais profundas (*deep neural networks*), que

se distinguem das ANN pelo fato de que podem aprender a melhor representação para a tarefa desejada, além de possibilitar neurônios mais avançados ou múltiplas ativações em um mesmo neurônio (JANIESCH; ZSCHECH; HEINRICH, 2021).

Existem diversos algoritmos dentro de cada um desses grupos. A Seção 2.3 traz exemplos de aplicação.

## 2.2 Avaliação de políticas públicas

Na definição de Pal (2005), política pública compreende um curso de ação ou inação das autoridades públicas para resolução de um problema ou conjunto inter-relacionado de problemas. A definição de problema não deve ser vista de forma restritiva. Nesse sentido, trazemos a explicação de Dente e Subirats (2014), os quais definem política pública como um conjunto de decisões e atividades para solução de um problema coletivo, que pode ser insatisfação quanto a uma necessidade, uma demanda ou uma oportunidade públicas.

A política pública pode abranger diversos programas, que surgem com a criação de condições, alocação de recursos e cronograma para sua implementação e, estes, por sua vez, podem ser formados por projetos e atividades que visam alcançar os objetivos do programa (CUNHA, 2018).

A avaliação de políticas públicas é uma forma de prover informações cientificamente válidas e socialmente legítimas sobre uma intervenção de forma que os diferentes atores envolvidos possam se posicionar sobre a mesma ou seus componentes com a finalidade de, entre outras, conhecê-la, adotá-la, aprimorá-la, decidir sobre sua continuidade ou determinar seus efeitos. Dessa forma, as análises podem compreender desde a pertinência da intervenção, a coerência com os objetivos, a produtividade e qualidade dos processos, os efeitos, a eficiência e interação com o contexto da implantação (CHAMPAGNE et al., 2011).

Governos ao redor do mundo têm estabelecido estratégias nacionais e/ou implementado técnicas de inteligência artificial para melhorar seus processos e serviços prestados à sociedade. Nesse contexto, Valle-Cruz et al. (2020) analisa como a inteligência artificial pode afetar o ciclo de política pública nas etapas de construção de agenda, formulação da política e processo decisório, implementação e avaliação.

Com relação à avaliação, os autores destacam o potencial de a avaliação se tornar contínua, presente em todas as fases do ciclo de políticas públicas. Como um exemplo do que pode ser útil na avaliação, cita o reconhecimento facial em aeroporto em conjunto com dados do governo chinês. Na sua leitura, isso seria uma entrada útil na avaliação de política pública de segurança e migração.

Ao final, os autores propõem um arcabouço conceitual, mas não mencionam

exemplos práticos de avaliação de política pública utilizando IA. Dito isso, e após pesquisas iniciais em motores de busca da internet, não encontramos uma revisão abrangente de aplicações de aprendizado de máquina na avaliação de políticas públicas, apesar de nos depararmos com alguns estudos descritos na seção seguinte.

A ausência de um estudo sistematizando o tema nos motivou a realizar esse projeto de revisão dada a oportunidade de contribuição sobre esse tema recente com servidores públicos, pesquisadores e demais profissionais que têm interesse na avaliação de políticas públicas.

## 2.3 Avaliação de políticas públicas com uso de aprendizado de máquina

Nessa seção trazemos exemplos de aplicação de aprendizado na avaliação de políticas públicas. [Linden e Yarnold \(2016a\)](#) propõem a utilização do algoritmo de aprendizado de máquina *Optimal Discriminant Analysis* (ODA) com o objetivo de avaliar o balanceamento das covariáveis entre os grupos de tratamento e controle, após a aplicação de uma técnica de pareamento, para definição dos grupos que seriam utilizados na avaliação dos efeitos de tratamento de um programa. A demonstração utilizou score de propensidade com pareamento de um para um. Contudo, os autores afirmam que qualquer técnica poderia ser utilizada. O desempenho do ODA foi comparado com um método convencional baseado nos mínimos quadrados ordinários a partir de dados de avaliação de um programa piloto de cuidados médicos primários em casa para pacientes com doenças crônicas ou para os quais foi predito que teriam altos custos de saúde no ano seguinte.

Os resultados do ODA foram consistentes com o método tradicional. Segundo os autores, as vantagens de se usar o ODA sobre métodos tradicionais de avaliação do balanceamento de covariáveis envolvem: lidar com qualquer tipo de variável, admitir múltiplos grupos de tratamento, ser insensível a dados assimétricos e contendo valores atípicos (*outliers*), aceitar técnicas de pareamento que produzem pesos para balanceamento, permitir utilizar validação cruzada para verificar a generalização do modelo para outros indivíduos.

[Brahma e Mukherjee \(2018\)](#) avaliaram um programa de alimentação escolar da Índia com o objetivo de verificar se o montante de recursos transferidos e a quantidade de grãos, repassados pelo Governo Central aos estados, servem de determinante para o número de crianças alimentadas. No ano fiscal 2016-17, o programa atendeu mais de 900 milhões de estudantes de 1 milhão de escolas. Os dados da avaliação compreenderam os quatro anos de 2011-12 a 2014-15. Uma das dificuldades apontadas para se usar métodos usuais, como mínimos quadrados, é que, com a inclusão de variáveis específicas dos estados e interações entre elas, a quantidade de variáveis poderia facilmente exceder o número de

observações.

Os autores utilizaram LASSO (*Least Absolute Selection and Shrinkage Operator*) e LASSO hierárquico para selecionar os preditores importantes da cobertura do programa. Esse último é apropriado para o caso de haver muitas covariáveis. O LASSO padrão é utilizado para verificação das conclusões principais a partir das covariáveis, sem suas interações. Ambas selecionam as covariáveis mais importantes e descartam as demais.

Os autores ressaltam que os resultados obtidos podem ser usados para predição, mas não para inferência, pois os efeitos marginais estimados e erro padrão são enviesados. Nesse sentido, sugerem que o método pode ser aplicado para avaliação da política com o reenquadramento da questão para algo como: o método seleciona variáveis da política como preditoras importantes da variável de resultado?

Ballestar et al. (2019a) avaliaram o desempenho de pesquisadores universitários submetidos a um programa de incentivos na região de Madri, Espanha, no período de 2005 a 2010. Os achados dos autores demonstram que a forma de contratação do profissional, na forma de servidor público efetivo ou temporária, influencia no desempenho do pesquisador no programa e que o desempenho profissional diminui ao longo do tempo. Complementarmente, observaram que o impacto do programa é menor para o servidor público. Além disso, o gênero é uma variável relevante para a análise e impacta de forma diferente em cada tipo de contratação.

Na avaliação, foi empregada uma abordagem com aprendizado de máquina não supervisionada chamada *Automated Nested Longitudinal Clustering* a partir de 96 variáveis por pesquisador, relacionadas a: gênero, forma de contratação, área de conhecimento e desempenho profissional. A aplicação do modelo dividiu os pesquisadores pela forma de contratação e cada um desses grupos em três subconjuntos relacionados ao desempenho. Aplicou-se testes de análise de variância (ANOVA) para confirmar a diferença estaticamente significativa entre os dois grupos, primeiro em termos das notas médias dos pesquisadores e, em seguida, entre a variação anual no desempenho dos pesquisadores. A análise descritiva foi realizada sobre esses grupos.

Com fins de avaliar a robustez do modelo, foi treinado um modelo supervisionado de rede neural artificial, usando as mesmas variáveis de agrupamento anteriores, para classificação dos pesquisadores nos seis grupos descobertos pelo modelo não supervisionado. A validação demonstrou a alta convergência de 99,2% entre os dois modelos. Por fim, os autores destacam que o modelo de rede neural pode ser facilmente utilizado para classificação de novos pesquisadores em tempo real.

Ratledge et al. (2021b) usaram aprendizado de máquina na avaliação de impacto da implantação de eletricidade na melhoria dos meios de sustento nas áreas rurais e periurbanas em Uganda. O método foi implementado em dois estágios. No primeiro, foi

utilizada uma Rede Neural Convolutacional treinada a partir de imagens de satélite de 27 mil vilas de 25 países da África subsariana, que inclui a Uganda, a fim de prever uma variável dependente indicativa dos meios de sustento em determinada localidade (comunidade ou vila). A justificativa da predição é por esse indicador ser oriundo de pesquisa demográfica esparsa, não é realizada em todas as localidades e nem nos mesmos locais todos os anos.

O segundo estágio envolveu a aplicação de dois métodos de aprendizado de máquina para isolar o efeito médio de tratamento: *Matrix Completion* (MC) e *Synthetic Controls with Elastic Net* (SC-EN). Os dados sobre a expansão da malha energética se referem a 2005, 2006, 2010 e 2013 a 2018. O grupo de tratamento compreendeu as localidades ao redor de 2 km de linhas de distribuição implantadas em 2011 e 2012. O grupo de controle englobou as localidades que não haviam recebido acesso até 2016.

Os modelos com MC e SC-EN estimaram efeitos similares estatisticamente significantes e comparáveis com uma estimativa de validação utilizando diferença em diferenças. O estudo estimou que o acesso à eletricidade eleva o nível dos ativos de riqueza na área rural de Uganda em 0,17 desvio padrão, mais do que o dobro da taxa de crescimento das regiões não tratadas. Adicionalmente, fornece uma forma generalizável, de baixo custo, para avaliar o impacto de políticas públicas com dados esparsos.

[Kreif et al. \(2022a\)](#) avaliaram os efeitos heterogêneos de programas de seguro de saúde da Indonésia usando modelo de aprendizado de máquina causal. Foram avaliados dois programas com tipo de oferecimento de seguro de saúde distintos, subsidiado e contributivo, e seus efeitos na mortalidade infantil e na assistência qualificada ao parto, se por parteira ou médico.

Os dados utilizados são de cinco pesquisas demográficas da Indonésia referentes a 13 das 27 províncias nos anos de 1993, 1997, 1998, 2000, 2007/2008 e 2014/2015. O tipo de programa a que a mãe estava associada (subsidiado, contributivo ou sem seguro saúde) definiu os dois grupos de tratamento e o de controle. O método utilizado para estimar efeitos individuais foi *Generalized Random Forests*. Quando agregados, possibilitam estimar efeito de tratamento médio para subgrupos de interesse.

As vantagens apontadas pelos autores é que, com o método de aprendizado de máquina, não é necessário especificar os subgrupos previamente. Ademais, são previstos testes estatísticos para avaliar se a heterogeneidade do tratamento explicada pelas covariáveis é significativa, além de indicarem quais variáveis estão mais associadas com a explicação da heterogeneidade. Nesse sentido, em adição aos efeitos heterogêneos associados às variáveis socioeconômicas especificadas inicialmente, foram encontradas outras variáveis relevantes, a exemplo de mulheres residentes em certas províncias que teriam se beneficiado mais que outros subgrupos se estivessem seguradas no modo contributivo.





## 3 Metodologia

O objetivo deste trabalho é mapear evidências de uso de aprendizado de máquina na avaliação de políticas públicas. Desse modo, buscou-se uma metodologia mais adequada a esse propósito. A revisão de escopo é indicada quando se deseja examinar a extensão, alcance e natureza da atividade de pesquisa, determinar o valor de realizar uma revisão sistemática, sumarizar e disseminar evidências e identificar lacunas de pesquisa (ARKSEY; O'MALLEY, 2005). Portanto, é adequada quando ainda não foi feita uma revisão abrangente de um corpo de literatura ou esse exibe uma natureza ampla, complexa ou heterogênea, para a qual ainda não é possível uma revisão sistemática mais precisa (PETERS et al., 2015).

A revisão de escopo se assemelha à revisão sistemática, vez que ambas seguem processo criterioso e previamente definido de revisão de literatura, porém possuem objetivos diferentes. De forma geral, enquanto na análise sistemática busca-se as melhores evidências para confirmar práticas correntes, solucionar conflitos, identificar vieses; na revisão de escopo, a estratégia está mais limitada ao mapeamento e entendimento das evidências existentes. Nesse sentido, a avaliação da qualidade dos estudos incluídos e análises estatísticas mais avançadas, em geral, não estão presentes na revisão de escopo (COELHO et al., 2021). Porém, Arksey e O'Malley (2005) alertam que a revisão de escopo não deve ser vista como uma avaliação rápida ou de menor custo, vez que demanda bastante recurso para sua realização.

Em que pese a revisão de escopo ter uma abrangência mais ampla que a revisão sistemática, é importante que sejam definidos o objetivo e a questão (ou questões) a ser perseguida na pesquisa a fim de aumentar a sua utilidade pelos leitores. Da mesma forma, é importante que se garanta a relação entre os critérios de inclusão dos trabalhos avaliados com o objetivo e questão definidos, de tal sorte a guiar o pesquisador na busca de evidências que respondam a questão elaborada. Salienta-se que a revisão de escopo pode ser realizada de maneira interativa. Isso quer dizer que, com o registro dos devidos esclarecimentos, o protocolo estabelecido para a pesquisa pode ser refinado durante o processo visando a atender melhor o objetivo proposto (PETERS, 2016).

Em comparação com a revisão de literatura tradicional, Munn et al. (2018) apontam as seguintes características relativas à revisão de escopo:

- definição prévia do protocolo;
- sistemática e, em geral, incluem uma pesquisa exaustiva;
- transparência e reprodutibilidade;

- controles para redução de erro e aumento de confiança;
- extração e apresentação de dados de forma estruturada.

Feita essa introdução sobre a revisão de escopo, explica-se a seguir os passos da metodologia para esse estudo. As definições foram embasadas no arcabouço metodológico proposto por [Arksey e O'Malley \(2005\)](#) e contribuições efetuadas por [Levac, Colquhoun e O'Brien \(2010\)](#) e [Peters et al. \(2015\)](#), bem como no Manual elaborado por [Peters et al. \(2017\)](#).

### 3.1 Identificação da questão de pesquisa

Como o objetivo da revisão de escopo é de mapeamento de estudos, recomenda-se utilizar uma questão de pesquisa ampla para não restringir a descoberta de artigos. Eventuais decisões de escopo, caso se obtenha um grande número de referências, podem ser tomadas assim que se obtiver conhecimento do volume ([ARKSEY; O'MALLEY, 2005](#)).

[Peters et al. \(2017\)](#) recomendam a aplicação do mnemônico PCC (População, Conceito e Contexto) para elaboração do título e da questão de pesquisa. Com isso, espera-se que comuniquem bem e facilitem a identificação da abrangência, foco e aplicação da revisão de escopo.

Considerando que se almeja mapear a utilização de aprendizado de máquina na avaliação de políticas públicas, a população objeto dessa revisão compreende as políticas públicas.

O conceito é o uso de aprendizado de máquina. É interessante nesse estágio definir o conceito para clarificar o escopo do estudo, guiar a estratégia de busca e seleção de estudos ([LEVAC; COLQUHOUN; O'BRIEN, 2010](#)). A explicação do conceito pode incluir detalhes da intervenção ou fenômeno que será melhor detalhado na revisão, incluindo classificações, teorias, desenhos de pesquisa, *frameworks*, entre outros. Para esse estudo, a definição consta na [seção 2.1](#).

O contexto pode se referir a localidades, fatores culturais, interesses raciais, comunidades, disciplinas, entre outros ([PETERS et al., 2015](#)). No estudo em tela, o contexto é a avaliação de políticas públicas, ou seja, a configuração onde se deseja obter mais informações sobre a aplicação de aprendizado de máquina, conforme [seções 2.2 e 2.3](#).

Deste modo, considerando o mnemônico PCC, a questão de pesquisa definida foi: Como o aprendizado de máquina tem sido utilizado na avaliação de políticas públicas?

Subquestões não são obrigatórias, mas podem auxiliar no mapeamento das evidências ([PETERS et al., 2017](#)). Isto posto, elaborou-se as subquestões abaixo com relação aos estudos identificados:

- Quais as características gerais dos estudos?
- Quais os objetos de avaliação?
- Quais métodos de aprendizado são aplicados?

Levac, Colquhoun e O'Brien (2010) ressaltam a importância de se estabelecer o objetivo da revisão de escopo no estágio inicial para facilitar as decisões futuras na seleção de estudos e extração de dados. Assim, o objetivo desse estudo é que as evidências possam ser utilizadas como fonte de referência para avaliadores de políticas públicas, seja no âmbito acadêmico, de consultores privados e servidores da administração pública, mais especificamente de órgãos de controle e de avaliação de políticas públicas.

## 3.2 Identificação dos estudos relevantes

A revisão de escopo deve identificar estudos publicados e não publicados que subsidiem a resposta da questão de pesquisa. Além das bases de pesquisa devem ser definidas língua, período de publicação e estratégias de pesquisa. Deve-se avaliar ainda a disponibilidade e necessidade de ajuda técnica para elaboração dos termos de consulta. A estratégia pode ser refinada após uma busca piloto e mais familiaridade com o corpo de conhecimento. Na verdade, trata-se de um processo interativo e incremental, aperfeiçoado à medida que se obtém mais conhecimento sobre o assunto (ARKSEY; O'MALLEY, 2005). Os autores devem justificar e dar conhecimento das potenciais limitações de escopo (LEVAC; COLQUHOUN; O'BRIEN, 2010).

Decidiu-se não restringir inicialmente o período de publicação dos trabalhos. Caso sejam incluídos muitos estudos, esse critério pode ser revisto. Quanto à linguagem de escrita, serão adotadas as línguas portuguesa e inglesa adotado como critério de inclusão. Quanto aos termos de busca, inicialmente foi definida a expressão do [Quadro 1](#), baseada nos três núcleos da questão de pesquisa (população, conceito e contexto). Procedimento semelhante é exemplificado por [Galvão e Ricarte \(2019\)](#) para elaboração da estratégia de busca.

Os parênteses indicam a ordem de precedência de avaliação da expressão. As aspas duplas delimitam um termo composto, de forma que os termos compreendidos devem ocorrer juntos e nessa sequência. O conector OR define que a ocorrência de pelo menos um dos termos conectados por ele é suficiente para satisfazer a expressão, enquanto que o conector AND estabelece que ambos os termos devem ocorrer no texto pesquisado. Observa-se que a intenção é que ocorra ao menos um termo relacionado a aprendizado de máquina, um de avaliação e um de política pública.

(PETERS et al., 2015) recomendam que o primeiro passo é uma pesquisa limitada nos bancos de dados mais relevantes com o objetivo de extrair termos dos títulos, resumo

Quadro 1 – Expressão inicial

| Língua    | Expressão de Busca   |
|-----------|--|
| Português | (“aprendizado de máquina” OR “aprendizado profundo” OR “inteligência artificial”)<br>AND (avaliação)<br>AND (“política pública”) |
| Inglês    | (“machine learning” OR “deep learning” OR “artificial intelligence”)<br>AND (evaluation)<br>AND (“public policy” )               |

Fonte: Elaborado pelo autor.

dos artigos e palavras-chave indexadas. A seguir, deve ser realizada busca completa em todas as bases com os termos incluídos. Por fim, a lista de referência dos artigos e relatórios selecionados podem ser analisados em busca de estudos adicionais, técnica conhecida como bola de neve.

### 3.3 Seleção de estudos

Em seu trabalho que definiu originariamente a revisão de escopo, [Arksey e O'Malley \(2005\)](#) exemplificaram a estrutura metodológica com um caso prático de aplicação. Nele, foram obtidos um total de 3.867 referências na etapa identificação que, ao final da seleção de estudos, resultaram em 204.

Como estratégia, foram utilizados critérios de inclusão e exclusão baseados na questão de pesquisa. Eles devem ser documentados desde o início para garantir a consistência ao longo do trabalho, mas podem ser revisados à medida que se adquire familiaridade com o assunto. Os critérios foram aplicados no resumo dos trabalhos e, quando não estava claro a sua ligação ou não com a questão de pesquisa, foi realizada a leitura do texto completo.

Recomenda-se que a revisão de escopo seja realizada de forma cegada, por no mínimo dois participantes independentes, para evitar potenciais ambiguidades e garantir a relevância dos estudos. Eventuais conflitos na análise podem ser resolvidos por consenso ou por um terceiro revisor ([LEVAC; COLQUHOUN; O'BRIEN, 2010](#); [PETERS et al., 2015](#); [PETERS et al., 2017](#)). Nesse contexto, a etapa de seleção de estudos será realizada de forma cegada com a colaborações de servidores da instituição do autor.

Finalmente, o número de estudos selecionados e o processo decisório para seleção devem ser descritos de forma narrativa e por meio de um fluxograma de decisão de busca ([PETERS et al., 2015](#)).

## 3.4 Mapeamento dos dados

Nesta etapa, os dados relevantes de cada estudo primário serão extraídos para permitir o mapeamento, categorização e comparação das evidências na etapa seguinte. É ideal que se tenha um formulário padronizado para extração dos dados, em que pese nem sempre os dados procurados estejam disponíveis em todos os estudos ([ARKSEY; O'MALLEY, 2005](#)).

[Levac, Colquhoun e O'brien \(2010\)](#) recomendam que esse processo seja iterativo, possibilitando ao pesquisador atualizar o formulário durante o estudo. Ademais, como passo inicial, aconselham que seja feita a extração de 5 a 10 estudos, por dois pesquisadores do time de pesquisa, e avaliar se a extração está consistente com a questão de pesquisa e objetivo. Contudo, conforme ressaltado em [Peters et al. \(2017\)](#), o mais importante é que os autores sejam transparentes nos métodos adotados para extração de dados. Assim, o passo-a-passo da execução do projeto será registrado de forma a assegurar a reprodutibilidade da pesquisa.

## 3.5 Agrupamento, sumarização e reporte dos resultados

[Levac, Colquhoun e O'brien \(2010\)](#) recomendam organizar esta etapa em três passos: análise dos dados, reporte e agregação de significado ao resultado. A partir da planilha elaborada na etapa anterior, pretende-se ter uma visão da distribuição dos estudos no tempo, por localidade e demais campos categorizáveis do mapeamento. Além disso, pretende-se realizar a análise dos estudos de forma atenta a outras codificações e classificações que puderem ser realizadas.

Na sequência, faremos o reporte dos resultados considerando o objetivo e as questões da pesquisa, principalmente por meio de gráficos de frequência. Por fim, consignaremos nossa visão dos resultados e possíveis aplicações nas considerações finais.

Como exemplo da identificação de lacunas, em revisão de escopo sobre a aplicação de inteligência artificial na gestação, [Oprescu et al. \(2020\)](#) identificaram que havia poucos estudos sobre análise automatizada envolvendo a situação emocional da gestante, apesar de esse ser um fator de risco relevante.



## 4 Identificação, seleção e mapeamento

Neste capítulo, iniciamos a aplicação da metodologia estabelecida no [Capítulo 3](#). Dado que o objetivo, questão e subquestões de pesquisa foram delimitados na [seção 3.1](#), nas seções seguintes discorreremos sobre a identificação de estudos, seleção de estudos e mapeamentos dos dados.

### 4.1 Identificação dos estudos relevantes

A fim de identificar os estudos relevantes, conforme estratégia elaborada na [seção 3.2](#), primeiramente realizamos o refinamento incremental da expressão de pesquisa, demonstrado na primeira subseção. Na subseção seguinte, utilizamos a expressão com as adaptações necessárias para cada base de dados pesquisada.

#### 4.1.1 Refinamento da expressão de busca

A fim de testar a pesquisa do [Quadro 1](#) da [seção 3.2](#), utilizamos a Scopus<sup>1</sup>, uma base de dados multidisciplinar com mais de 84 milhões de documentos indexados. A aplicação da expressão de busca acima, com adaptações para essa base para busca no título, resumo e palavras-chave indexadas, retornou 80 resultados<sup>2</sup>.

Quadro 2 – Primeiro teste no Scopus

|  |  |
|--|--|
| Inglês<br>80 resultados  | TITLE-ABS-KEY ( ( “machine learning” OR “deep learning” OR “artificial intelligence” ) AND evaluation AND “public policy” )  |
| Português<br>Nenhum resultado  | TITLE-ABS-KEY ( ( “aprendiza* de máquina” OR “aprendiza* profund*” OR “inteligência artificial” ) AND avaliaç* AND “política* pública*” )  |
| 15 palavras-chave com mais ocorrência nos resultados da pesquisa em inglês | <b>Public Policy</b> (40), <b>Artificial Intelligence</b> (35), <b>Machine Learning</b> (28), Learning Systems (15), Decision Making (12), Humans (9), Human (8), Sustainable Development (7), Learning Algorithms (7), Big Data (7), Forecasting (6), Article (6), Risk Assessment (5), Policy Making (5), <b>Policy Evaluation</b> (5) |

Fonte: Elaborado pelo autor.

Nota: A tabela ilustra a expressão de busca utilizada e as 15 palavras-chave mais frequentes nos estudos retornados. O número entre parênteses reflete a quantidade de trabalhos que continham a respectiva palavra-chave indexada.

<sup>1</sup> <http://www.scopus.com>

<sup>2</sup> As pesquisas realizadas nessa seção foram conduzidas em 11/03/2023

A avaliação de programa também é um termo bastante utilizado pois, em geral, é como a ação governamental é organizada. Dos artigos citados na [seção 2.3](#), observamos dois em que “programa” ocorre no título: “*program evaluation*” e “*mid-day meal program*”. Na tentativa de incluir o termo “programa” junto com os termos de política pública, a pesquisa na mesma base retornou 5.702 documentos. Observando-se os estudos retornados, conforme [Quadro 3](#), depreende-se, da análise das palavras-chave mais frequentes, que a pesquisa ficou muito abrangente e a predominância passou a ser de termos relacionados a programas da computação.

Quadro 3 – Teste Scopus incluindo o termo programa

|  |   |
|--|---|
| Inglês<br>5.741 resultados                           | TITLE-ABS-KEY ( ( “machine learning” OR “deep learning” OR “artificial intelligence” ) AND evaluation AND “public policy” )   |
| 15 palavras-chave com mais ocorrência nos resultados | <b>Artificial Intelligence</b> (2,512), <b>Machine Learning</b> (1,465), Article (1,057), Learning Systems (985), Human (984), <b>Deep Learning</b> (852), Humans (821), Application Programs (701), Computer Program (580), Algorithms (575), Program Processors (526), Algorithm (484), Software (478), Learning Algorithms (452), Priority Journal (357) |

Fonte: Elaborado pelo autor.

Nota: A tabela ilustra a expressão de busca utilizada e as 15 palavras-chave mais frequentes nos estudos retornados. O número entre parênteses reflete a quantidade de trabalhos que continham a respectiva palavra-chave indexada.

Nessa esteira, testamos a inclusão do termo composto “avaliação de programa”, em substituição a apenas programa, obtendo um resultado de 246 artigos. A expressão de busca utilizada e as palavras-chave mais frequentes estão ilustradas no [Quadro 4](#):

Quadro 4 – Teste Scopus incluindo o termo “avaliação de programa”

|  |  |
|--|--|
| Inglês<br>246 resultados   | TITLE-ABS-KEY ( ( “machine learning” OR “deep learning” OR “artificial intelligence” ) AND evaluation AND (“public policy” OR “program evaluation” OR “policy evaluation”) )   |
| Português<br>Nenhum resultado  | TITLE-ABS-KEY ( ( “aprendiza* de máquina” OR “aprendiza* profund*” OR “inteligência artificial” ) AND avaliaç* AND (“política* pública*” OR “avaliac* de programa*” ) )  |
| 15 palavras-chave com mais ocorrência nos resultados da pesquisa em inglês | <b>Program Evaluation</b> (112), <b>Artificial Intelligence</b> (110), Human (106), Humans (93), <b>Machine Learning</b> (78), Article (66), <b>Public Policy</b> (40), Female (37), Male (34), Procedures (31), Adult (26), Learning Systems (25), Education (25), Priority Journal (24), <b>Deep Learning</b> (22) |

Fonte: Elaborado pelo autor.

Nota: A tabela ilustra a expressão de busca utilizada e as 15 palavras-chave mais frequentes nos estudos retornados. O número entre parênteses reflete a quantidade de trabalhos que continham a respectiva palavra-chave indexada.



Dos cinco estudos da [seção 2.3](#), identificamos que estão incluídos apenas o de [Ballestar et al. \(2019a\)](#), incluído com a expressão inicial, e o de [Linden e Yarnold \(2016a\)](#), que passou a constar após a última expressão com “*program evaluation*”. O trabalho de [Kreif et al. \(2022a\)](#) não estava entre os resultados. Assim, investigando as palavras-chave do artigo, verificamos que consta entre elas a “*policy evaluation*”. Editamos novamente a expressão de pesquisa, que passou a conter o referido estudo e elevou a quantidade de resultados na base Scopus para 493. A nova versão da expressão de pesquisa está consignada no [Quadro 5](#).

Quadro 5 – Teste Scopus incluindo o termo “avaliação de política”

|  |  |
|--|--|
| Inglês<br>493 resultados                                       | TITLE-ABS-KEY ( ( “machine learning” OR “deep learning” OR “artificial intelligence” ) AND evaluation AND (“public policy” OR “program evaluation” OR “policy evaluation”))  |
| Português<br>Nenhum resultado                                  | TITLE-ABS-KEY ( ( “aprendiza* de máquina” OR “aprendiza* profund*” OR “inteligência artificial” ) AND avaliaç* AND (“política* pública*” OR “avaliaç* de programa*” OR “avaliaç* de política*” ) )   |
| 15 palavras-chave com mais ocorrência nos resultados em inglês | <b>Artificial Intelligence</b> (197), <b>Machine Learning</b> (166), <b>Policy Evaluation</b> (146), Human (119), Reinforcement Learning (112), <b>Program Evaluation</b> (112), Humans (101), Article (83), Learning Algorithms (67), Learning Systems (64), <b>Deep Learning</b> (54), <b>Public Policy</b> (40), Female (40), Decision Making (38), Markov Processes (37) |

Fonte: Elaborado pelo autor.

Nota: A tabela ilustra a expressão de busca utilizada e as 15 palavras-chave mais frequentes nos estudos retornados. O número entre parênteses reflete a quantidade de trabalhos que continham a respectiva palavra-chave indexada.

Objetivando incluir o termo “programa”, mas sem deixar os resultados tão amplos quanto no [Quadro 3](#), incluímos uma nova condição de que o termo “política” deve ocorrer em conjunto “(programa e política)” quando não for um termo composto, como em “avaliação de programa”. O resultado está no [Quadro 6](#).

Verifica-se que a quantidade de artigos encontrados passou de 493 para 640. Apesar de palavras-chave relacionadas à política pública não terem sofrido aumento nos resultados (*Policy Evaluation*(146), *Program Evaluation*(112) e *Public Policy*(40)), em uma rápida leitura dos títulos dos 147 novos estudos, encontramos alguns potencialmente relevantes para o mapeamento:

- *Has the Three-Year Action Plan improved the air quality in the Fenwei Plain of China? Assessment based on a machine learning technique;*
- *The impact of institutional delivery on neonatal and maternal health outcomes: Evidence from a road upgrade programme in India;*

Quadro 6 – Teste Scopus incluindo o termo “programa AND política”

|  |  |
|--|--|
| Inglês<br>640 resultados                                       | TITLE-ABS-KEY ( ( “machine learning” OR “deep learning” OR “artificial intelligence” ) AND evaluation AND (“public policy” OR “program evaluation” OR “policy evaluation” OR ( policy AND program ) ) )  |
| Português<br>Nenhum resultado                                  | TITLE-ABS-KEY ( ( “aprendiza* de máquina” OR “aprendiza* profund*” OR “inteligência artificial” ) AND avaliaç* AND (“política* pública*” OR “avaliaç* de programa*” OR “avaliaç* de política*” OR ( política* AND programa* ) ) )  |
| 15 palavras-chave com mais ocorrência nos resultados em inglês | <b>Artificial Intelligence</b> (250), <b>Machine Learning</b> (203), <b>Human</b> (146), <b>Policy Evaluation</b> (146), Reinforcement Learning (127), <b>Humans</b> (122), <b>Program Evaluation</b> (112), <b>Article</b> (100), Learning Systems (82), <b>Deep Learning</b> (81), Learning Algorithms (74), Decision Making (38), Female (43), <b>Public Policy</b> (40), Male (39) |

Fonte: Elaborado pelo autor.

Nota: A tabela ilustra a expressão de busca utilizada e as 15 palavras-chave mais frequentes nos estudos retornados. O número entre parênteses reflete a quantidade de trabalhos que continham a respectiva palavra-chave indexada.

- *Machine learning for causal inference: Estimating heterogeneous treatment effects;*
- *Impact of the Qali Warma school feeding program on chronic child malnutrition [Impacto del programa de alimentación escolar Qali Warma sobre la anemia y la desnutrición crónica infantil];*
- *Coupling data science with community crowdsourcing for urban renewal policy analysis: An evaluation of Atlantas Anti-Displacement Tax Fund.*

Retornando aos artigos iniciais da seção 2.3, o de Brahma e Mukherjee (2018) foi publicado no Volume 18-1 do Jornal *Applied Econometrics and International Development*, que é indexado pelo Scopus. Contudo, não identificamos esse estudo nos resultados, apesar de haver outros publicados no mesmo volume do jornal que estão indexados. Recebemos a mensagem de que o artigo foi incluído na plataforma em 13/05/2023. Como a resposta foi posterior à etapa de identificação, ele foi incluído manualmente na etapa de seleção.

Já o estudo de Ratledge et al. (2021b) foi publicado no arXiv<sup>3</sup>. Trata-se de uma plataforma de publicação de artigos que passa por um processo de moderação mas não é revisada por pares, sendo as publicações de responsabilidade dos respectivos autores. Engloba as áreas de física, matemática, ciência da computação, matemática, biologia quantitativa, finança quantitativa, ciência de sistemas e engenharia elétrica, economia. Por conseguinte, essa base de dados será incluída na revisão de escopo como integrante da literatura cinzenta.

<sup>3</sup> <<http://www.arxiv.org>>

### 4.1.2 Buscas nas bases de dados

Utilizando a expressão refinada da subseção anterior, e com o objetivo de ser bastante abrangente, realizamos a pesquisa em seis bases de dados multidisciplinares entre os dias 13 e 15 de março de 2023:

- Scopus: base de dados multidisciplinar;
- Web of Science: base de dados multidisciplinar;
- ScienceDirect: base de dados multidisciplinar;
- IEEE Xplore: base de dados voltada para áreas de engenharia elétrica, ciência da computação e eletrônica;
- ACM: base de dados voltada para computação e tecnologia da informação;
- ArXiv: plataforma de publicação de artigos que passa por um processo de moderação mas não é revisada por pares. Engloba as áreas de física, matemática, ciência da computação, matemática, biologia quantitativa, finança quantitativa, ciência de sistemas e engenharia elétrica, economia.

A [Tabela 1](#) apresenta o número de estudos extraídos e as expressões com os ajustes necessários de cada base de dados.

Tabela 1 – Quantidade de estudos encontrados por pesquisa realizada

| Base                  | Quantidade | Expressão  |
|-----------------------|------------|--|
| Scopus<br>(inglês)    | 640        | TITLE-ABS-KEY ( ( "machine learning"OR "deep learning"OR "artificial intelligence") AND evaluation AND ( "public policy"OR "program evaluation"OR "policy evaluation"OR ( policy AND program ) ) )                           |
| Scopus<br>(português) | 0          | TITLE-ABS-KEY ( ( "aprendiza* de máquina"OR "aprendiza* profund*"OR "inteligência artificial") AND avaliaç* AND ( "política* pública*"OR "avaliaç* de programa*"OR "avaliaç* de política*"OR ( política* AND programa* ) ) ) |
| ArXiv<br>(inglês)     | 568        | (all:"machine learning"OR all:"deep learning"OR all:"artificial intelligence") AND (all:evaluation) AND (all:"public policy"OR all:"program evaluation"OR all:"policy evaluation")   |
| ArXiv<br>(inglês)     | 135        | (all:"machine learning"OR all:"deep learning"OR all:"artificial intelligence") AND (all:evaluation) AND (all:policy AND all:program)   |

| Base                            | Quantidade | Expressão   |
|---------------------------------|------------|---|
| ArXiv<br>(português)            | 0          | (all:"aprendiza* de máquina"OR all:"aprendiza* profund*"OR all:"inteligência artificial") AND (all:avaliaç*) AND (all:"política* pública*"OR all:"avaliaç* de programa*"OR all:"avaliaç* de política*")   |
| ArXiv<br>(português)            | 0          | (all:"aprendiza* de máquina"OR all:"aprendiza* profund*"OR all:"inteligência artificial") AND (all:avaliaç*) AND (all:política* AND all:programa*)  |
| Web of Science<br>(inglês)      | 609        | TS=(( "machine learning"OR "deep learning"OR "artificial intelligence") AND evaluation AND ( "public policy"OR "program evaluation"OR "policy evaluation"OR ( policy AND program ) ))   |
| Web of Science<br>(português)   | 1          | TS=(( "aprendiza* de máquina"OR "aprendiza* profund*"OR "inteligência artificial") AND avaliaç* AND ( "política* pública*"OR "avaliaç* de programa*"OR "avaliaç* de política*"OR ( política* AND programa* ) ) )  |
| IEEE Xplore<br>(inglês)         | 409        | ( "machine learning"OR "deep learning"OR "artificial intelligence") AND evaluation AND ( "public policy"OR "program evaluation"OR "policy evaluation"OR ( policy AND program ) )  |
| IEEE Xplore<br>(português)      | 0          | ( "All Metadata":"aprendizado de máquina"OR "All Metadata":"aprendizagem de máquina"OR "All Metadata":"aprendizagem profunda"OR "All Metadata":"aprendizado profundo"OR "All Metadata":"inteligência artificial") AND "All Metadata":avaliação AND ( "All Metadata":"política pública"OR "All Metadata":"políticas públicas"OR "All Metadata":"avaliação de programa*"OR "All Metadata":"avaliação de política*"OR ( "All Metadata":política* AND "All Metadata":programa* ) )  |
| ACM Digital Library<br>(inglês) | 282        | (Title:( "machine learning"OR "deep learning"OR "artificial intelligence") OR Abstract:( "machine learning"OR "deep learning"OR "artificial intelligence") OR Keyword:( "machine learning"OR "deep learning"OR "artificial intelligence")) AND (Title:(evaluation) OR Abstract:(evaluation) OR Keyword:(evaluation)) AND (Title:( "public policy"OR "program evaluation"OR "policy evaluation"OR ( policy AND program ) ) OR Abstract:( "public policy"OR "program evaluation"OR "policy evaluation"OR ( policy AND program ) ) OR Keyword:( "public policy"OR "program evaluation"OR "policy evaluation"OR ( policy AND program ) )) |

| Base                            | Quantidade   | Expressão   |
|---------------------------------|--------------|---|
| ACM Digital Library (português) | 0            | [[Title: "aprendiza* de máquina"] OR [Title: "aprendiza* profund*"] OR [Title: "inteligência artificial"] OR [Abstract: "aprendiza* de máquina"] OR [Abstract: "aprendiza* profund*"] OR [Abstract: "inteligência artificial"] OR [Keywords: "aprendiza* de máquina"] OR [Keywords: "aprendiza* profund*"] OR [Keywords: "inteligência artificial"]] AND [[Title: avaliaç*] OR [Abstract: avaliaç*] OR [Keywords: avaliaç*]] AND [[Title: "política* pública*"] OR [Title: "avaliaç* de programa*"] OR [Title: "avaliaç* de política*"] OR [[Title: política*] AND [Title: programa*]] OR [Abstract: "política* pública*"] OR [Abstract: "avaliaç* de programa*"] OR [Abstract: "avaliaç* de política*"] OR [[Abstract: política*] AND [Abstract: programa*]] OR [Keywords: "política* pública*"] OR [Keywords: "avaliaç* de programa*"] OR [Keywords: "avaliaç* de política*"] OR [[Keywords: política*] AND [Keywords: programa*]]] |
| Science Direct                  | 81           | ( "machine learning"OR "deep learning"OR "artificial intelligence") AND evaluation AND ( "public policy"OR "program evaluation"OR "policy evaluation"OR ( policy AND program ) )  |
| Science Direct                  | 0            | ("aprendizado de máquina"OR "aprendizagem de máquina"OR "aprendizagem profunda"OR "aprendizado profundo"OR "inteligência artificial") AND avaliação   |
| <b>Total</b>                    | <b>2.725</b> |   |

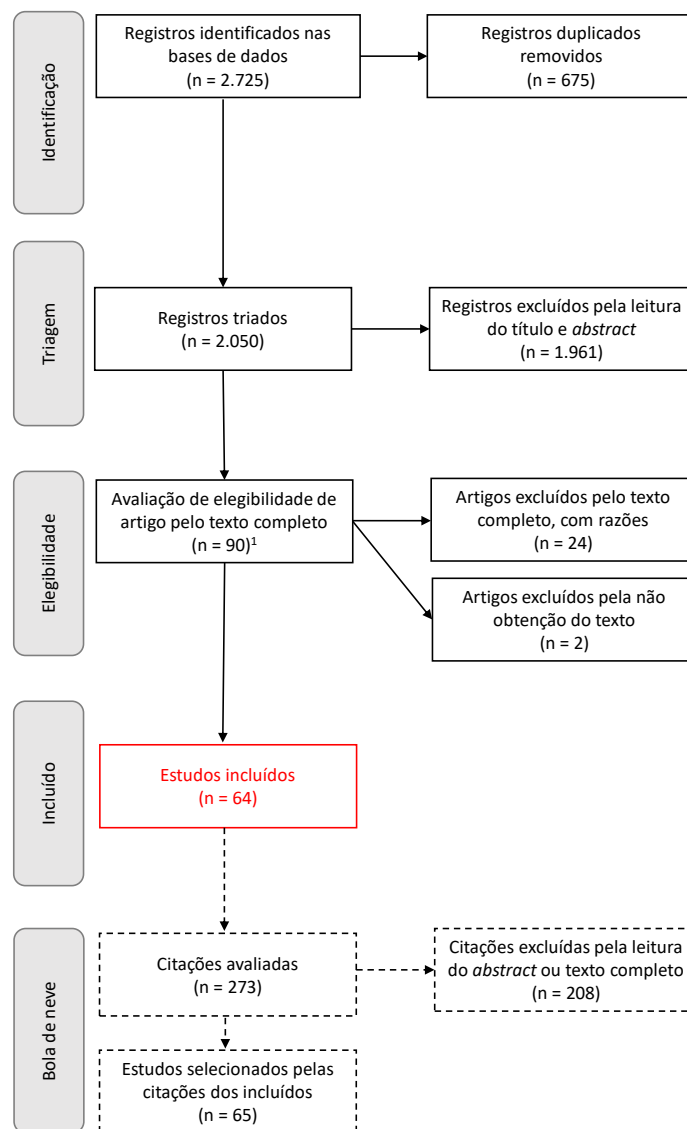
Fonte: Elaborado pelo autor.

No total foram identificados 2.725 estudos. Registra-se que podem haver duplicações decorrente, por exemplo, de indexação do mesmo trabalho por diferentes bases de dados ou ainda versões de um mesmo estudo em diferentes meios de publicação. Todos os resultados foram exportados em arquivos no formato RIS, padrão de armazenamento de citações e suportado por diversos gerenciadores de referências.

## 4.2 Seleção de estudos

A seguir, descrevemos os passos executados para a seleção dos estudos, resumidos no fluxograma da [Figura 1](#).

Figura 1 – Fluxograma de decisão de seleção de estudos



Fonte: Elaborado pelo autor com base em [Peters et al. \(2015\)](#).

Nota: O total de estudos resultantes da triagem foi 89. Contudo, conforme registrado na [4.1](#), o trabalho de [Brahma e Mukherjee \(2018\)](#) foi adicionado manualmente por uma inconsistência identificada na base de dados e corrigida posteriormente pelo fornecedor, passando o número para 90.

Utilizamos a ferramenta Rayyan ([OUZZANI et al., 2016](#)) para apoiar essa etapa do trabalho. Trata-se de uma aplicação disponibilizada na internet<sup>4</sup> com a finalidade de ajudar na triagem de estudos de revisão, com recursos que incluem filtragem por termos

<sup>4</sup> <<https://rayyan.ai/>>

de interesse, agrupamento por características do estudo, fluxo de revisão, detecção de duplicados, definição de rótulos, colaboração e revisão cegada. Passamos, agora, a detalhar o fluxo da [Figura 1](#).

**Identificação.** Inicialmente, foi criado um projeto de revisão no Rayyan com a importação dos 2.725 estudos identificados na seção anterior. A ferramenta agrupa possíveis estudos duplicados para apoiar a remoção de trabalhos repetidos. Realizamos esse passo com a remoção de 675 artigos duplicados. De tal sorte, seguiram 2.050 artigos para a triagem.

**Triagem.** A triagem foi realizada pela leitura de título, *abstract* e palavras-chave indexadas, com o registro de decisão para inclusão ou remoção na ferramenta. Esse processo foi realizado de forma cegada. Dessa maneira, os estudos foram analisados por ao menos dois revisores. O autor revisou todos os trabalhos, enquanto que o segundo revisor foi um entre 10 servidores da mesma instituição. Para cada um, foram designados conjuntos com aproximadamente 205 estudos distintos.

A decisão final pela inclusão ou exclusão de cada trabalho foi dada pela maioria. Os 76 casos de empate entre revisores pela inclusão e exclusão foram avaliados por uma terceira revisora, papel desempenhado pela orientadora desta dissertação. Seu voto, portanto, foi o de minerva. Sintetizando, do total de 2.050 estudos incluídos na ferramenta, 89 foram selecionados para verificação de elegibilidade. A [Figura 2](#) apresenta um esquema dos passos da revisão cegada.

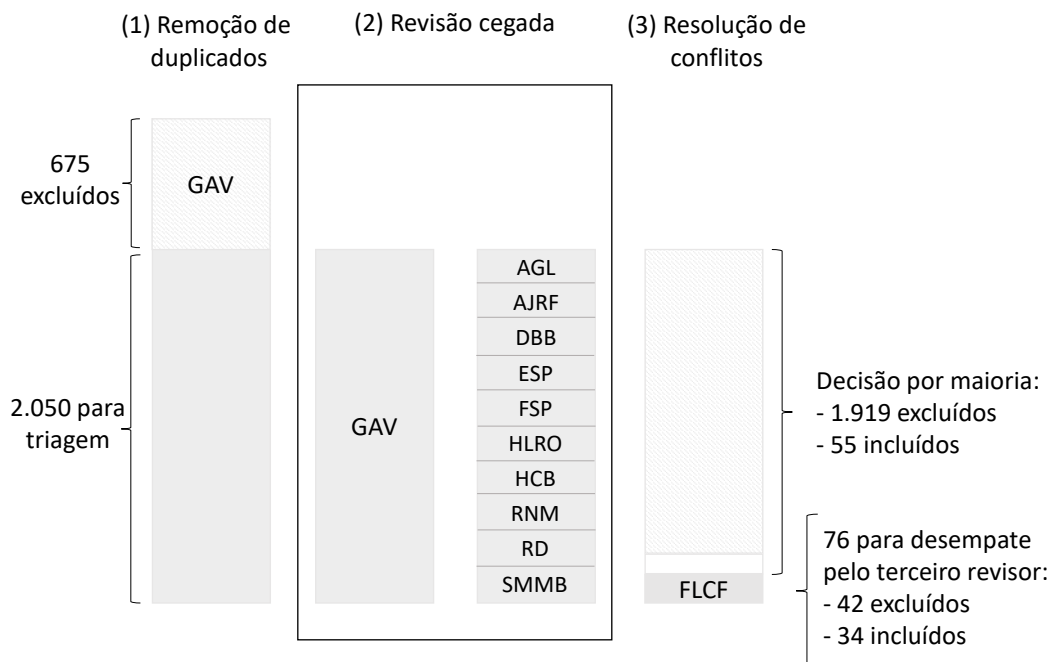
Os critérios de inclusão e exclusão tanto da triagem quanto da elegibilidade estão contidos no [Quadro 7](#). Eles foram definidos em atendimento ao objetivo da dissertação, cujo enfoque principal é de que os estudos selecionados sejam de aplicações de aprendizado de máquina na avaliação de políticas públicas, ilustradas pelo uso de dados reais.

**Elegibilidade.** Os dados bibliográficos dos estudos marcados para inclusão na ferramenta foram exportados do Rayyan e importados no sistema de gerenciamento de referências Mendeley para facilitar a navegação e leitura dos trabalhos. Adicionalmente, carregamos os arquivos contendo o conteúdo completo de cada publicação nessa ferramenta. Registra-se que não conseguimos acesso a dois estudos por meio dos convênios disponibilizados para o Programa de Pós-Graduação e nem em contato com as autoras ([KUMAR; SHARMA, 2019](#); [KUMAR; SHARMA, 2021](#)).

A partir daí, passamos à leitura dos textos avaliando a pertinência do conteúdo com os critérios do [Quadro 7](#). Do total de 90 estudos avaliados<sup>5</sup>, 2 foram excluídos pela não disponibilidade do texto e 24 pela análise do texto completo, cujas razões para exclusão foram elencadas no [Apêndice D](#). Logo, ao final, 64 estudos foram incluídos.

<sup>5</sup> O total de estudos resultantes da triagem foi 89. Contudo, conforme registrado na [4.1](#), o trabalho de [Brahma e Mukherjee \(2018\)](#) foi adicionado manualmente por uma inconsistência identificada na base de dados e corrigida posteriormente pelo fornecedor, passando o número para 90.

Figura 2 – Esquema da etapa de seleção com revisão cegada



Fonte: Elaborado pelo autor.

Nota: A figura destaca três momentos da etapa de seleção. (1) Remoção de estudos duplicados, tarefa realizada pelo autor (GAV). (2) Na revisão cegada, os trabalhos foram divididos em 10 grupos. O autor triou todos os estudos. Adicionalmente, cada grupo foi designado a um segundo revisor, dentre um grupo de 10 servidores do mesmo órgão do autor. Dessa forma, cada estudo foi triado por ao menos duas pessoas, de forma cegada, ou seja, sem tomar conhecimento da avaliação do outro colega. (3) Ao final, o cegamento foi desligado e as avaliações foram decididas por maioria. Os casos de empate de decisão foram encaminhados a uma terceira revisora, papel desempenhado pela orientadora desta dissertação (FLCF), que proferiu o voto de minerva.

**Bola de neve.** Para os estudos incluídos, durante a leitura, coletamos citações ao longo do texto que poderiam estar relacionadas com o tema dessa dissertação, observando principalmente o contexto em que a citação foi referenciada. Realizamos a pesquisa do *abstract* e, em alguns casos, quando necessário, fizemos uma breve análise do conteúdo completo para verificar indicativos de que poderia se tratar de aprendizado de máquina com demonstração de aplicação na avaliação de política pública, sem, contudo, concluir por uma análise definitiva. Consequentemente, foram coletadas 65 referências com potencial vinculação ao tema. Todas foram listadas no [Apêndice E](#) como contribuição para eventual trabalho futuro de expansão do mapeamento, vez que o prazo deste trabalho não permitia a inclusão no momento.



Quadro 7 – Critérios de inclusão e exclusão

| Critérios de inclusão   | Critérios de exclusão  |
|---|--|
| <ul style="list-style-type: none"> <li>• texto escrito em inglês ou português;</li> <li>• contém ao menos uma das palavras-chave de cada um dos grupos usados na expressão de busca (aprendizado de máquina, avaliação, política pública). Esse critério é aplicado na própria expressão de pesquisa;</li> <li>• aplicação de aprendizado de máquina na avaliação de política pública.</li> </ul> | <ul style="list-style-type: none"> <li>• apenas sugere como pode ser aplicado e não apresenta um caso prático de aplicação de aprendizado de máquina no âmbito da avaliação de política pública;</li> <li>• dados são simulados ou apenas baseados nos dados reais, ou seja, não são de fato dados de políticas públicas;</li> <li>• aplicação de aprendizado de máquina em avaliação de processos ou políticas de organização privada, ou seja, não é de política pública;</li> <li>• avaliação de tratamentos clínicos de saúde;</li> <li>• uso de aprendizado de máquina para implementação da política pública, e não para avaliação.</li> </ul> |

Fonte: Elaborado pelo autor.

### 4.3 Mapeamento dos dados

Esta etapa foi realizada em paralelo com o passo de verificação elegibilidade da seleção de estudos. Assim, à medida que cada estudo era selecionado para inclusão pelo texto completo lido, o mesmo era mapeado. A catalogação foi realizada em uma planilha no formato MS Excel contendo as seguintes colunas:

1. Referência: referência interna para o sistema gerenciador de referências;
2. Título: indicado no documento;
3. Palavras-chave: indicadas pelos autores no documento, separadas por ponto-e-vírgula;
4. Autores: conforme descrito no documento. Início pelo último nome, separado por vírgula dos demais nomes. Múltiplos autores separados por ponto-e-vírgula;
5. Ano: ano da publicação, conforme indicado na base de dados;
6. País dos Autores: país informado na afiliação dos autores, separados por ponto-vírgula existirem diferentes países;
7. Revista: meio de publicação do estudo, a exemplo de revista, jornal, conferência;
8. Objetivo resumido: síntese do objetivo obtido após leitura do texto;
9. Codificação do objetivo: codificação própria para agrupamento dos objetivos;
10. Intervenção: nome da política, programa, projeto avaliado;
11. Função de Governo: classificação da intervenção em uma das funções de governo estabelecidas na Portaria SOF/ME nº 2.520, de 21 de março de 2022 ([Brasil, 2023](#));
12. Causal: se o estudo propõe inferência causal (sim/não);
13. Dados - Registros: quantidade de registros no conjunto de dados;
14. Dados - Variáveis: quantidade de variáveis no conjunto de dados;
15. Dados - Fonte: descrição resumida da origem dos dados;
16. Métodos: métodos de aprendizado de máquina utilizados, separados por ponto-e-vírgula quando envolver mais de um;
17. Tipo de aprendizado: supervisionado, não supervisionado, semi-supervisionado ou aprendizado por reforço. Separados por ponto-e-vírgula quando houver mais de um algoritmo de tipos diferentes;
18. Aplicação: descrição de como o aprendizado de máquina foi aplicado;

19. Codificação da aplicação: codificação própria da aplicação;
20. Linguagem de programação: linguagem de programação utilizada;
21. Pacotes de software: pacotes mencionados.

Os dados coletados são apresentados nos Apêndices [A](#), [B](#) e [C](#), e analisados no [Capítulo 5](#).



## 5 Resultados

Neste capítulo realizamos a análise e reporte descrita na [seção 3.4](#). O escopo são os 64 estudos identificados, selecionados e incluídos, conforme relatado no [Capítulo 4](#). Os dados detalhados estão listados nos Apêndices [A](#), [B](#) e [C](#). As seções estão estruturadas de acordo com as subquestões de pesquisa, as quais recordamos a seguir:

1. Como o aprendizado de máquina tem sido utilizado na avaliação de políticas públicas?
  - a) Quais as características gerais dos estudos?
  - b) Quais os objetos de avaliação?
  - c) Quais métodos de aprendizado são aplicados?

Primeiramente, apresentamos uma visão das características gerais dos estudos em termos de tempo, localidade, assunto e objetivo. Na sequência, passamos para os objetos das avaliações, especificamente as intervenções avaliadas e conjuntos de dados envolvidos. Finalmente, exploramos a aplicação de aprendizado de máquina, com olhar para o objetivo de sua utilização, métodos empregados e linguagens e recursos de programação.

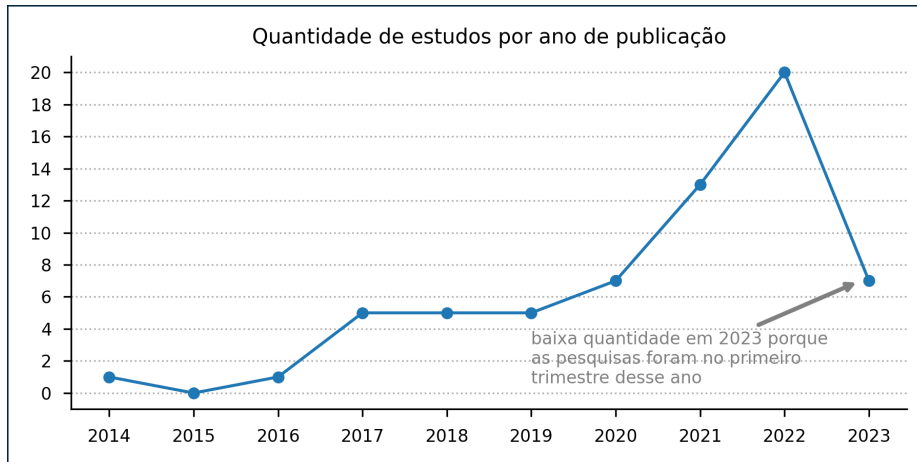
### 5.1 Características dos estudos

A primeira análise que fizemos foi sobre a frequência de publicação ao longo do tempo. Da leitura do gráfico da [Figura 3](#), se observa que o escopo de estudos publicados sobre avaliação de políticas públicas utilizando aprendizado de máquina em dados reais delimita um corpo de literatura recente e em ascensão. O estudo mais antigo é de 2014 ([MA; ZHANG, 2014](#)), com aumento no número de publicações a partir de 2017 e a maioria concentrada nos últimos cinco anos.

A etapa de busca nas bases bibliográficas foi realizada entre 13 e 15 de março de 2023, por isso há um decréscimo abrupto para esse ano ([HA; GRUBERT, 2023; YANG et al., 2023; MCFOWLAND; SOMANCHI; NEILL, 2023; VALENTE, 2023; HATAMYAR; KREIF, 2023; VIVIANO, 2023; COCKX; LECHNER; BOLLENS, 2023](#)). Para fins de comparação, a quantidade do início de 2023 é maior que a registrada para todo o ano de 2019 ([BALLESTAR et al., 2019b; KALLUS; ZHOU, 2019; GORI et al., 2019; CARTER; TJERNSTROM; TOLEDO, 2019; KREIF; DIAZORDAZ, 2019](#)).

Como foi dito no capítulo anterior, identificamos novos estudos por meio da exploração das citações bibliográfica dos estudos, o método conhecido como bola de neve.

Figura 3 – Quantidade de estudos por ano de publicação



Fonte: Elaborado pelo autor.

Nota: As pesquisas foram realizadas entre 13 e 15 de março de 2013. Não considera os potenciais estudos descobertos por meio de revisão das citações referenciadas e que não foram processados (mais detalhes sobre o processo de seleção na [seção 4.2](#)).

Em que pese não termos analisado-os completamente, extraímos o ano de publicação e combinamos com os dos estudos selecionados para montar o gráfico da [Figura 4](#).

Figura 4 – Quantidade de estudos por ano de publicação (selecionados e bola de neve)



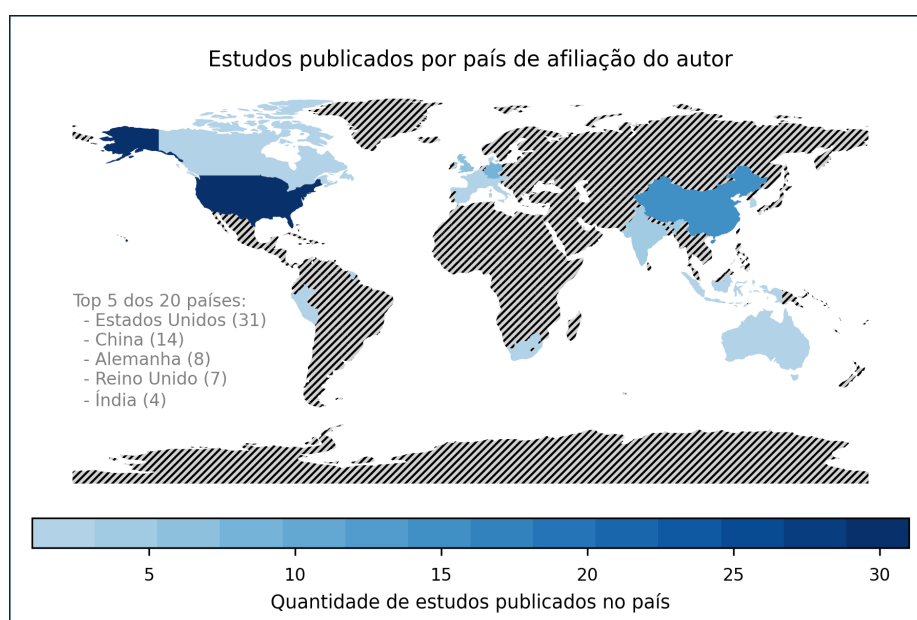
Fonte: Elaborado pelo autor.

Nota-se 5 estudos prévios a 2014, ano mais antigo entre os estudos selecionados. Todavia, são bem poucos em relação ao conjunto. De fato, verifica-se que a concentração

de publicações nos últimos anos se mantém, bem como o aclave da curva, apesar do breve decréscimo em 2018 e manutenção em 2021 da quantidade no ano anterior. Depreende-se, portanto, que mesmo contabilizando os estudos de bola de neve, o período de publicações para o escopo delimitado é bastante recente. Proporcionalmente, 96% estão compreendidos na última década (a partir 2014) e 74% nos últimos 5 anos (a partir de 2019).

A seguir, com o objetivo de verificar onde as pesquisas estão sendo realizadas, criamos a visualização de frequência por país de afiliação dos autores mostrada no gráfico da Figura 5. Na contagem, para cada estudo, países com mais de um autor foram considerados apenas uma vez.

Figura 5 – Quantidade de estudos por país de afiliação do autor



Fonte: Elaborado pelo autor.

Nota: Total de 20 países. Os países hachurados não tiveram nenhuma publicação. Para os demais, a intensidade da cor indica a quantidade de publicações em cada país, conforme legenda na base da figura. Países com mais de um autor no mesmo estudo foram considerados apenas uma vez por estudo.

Ao todo são 20 países, sendo que 8 estão relacionados a um único estudo. Depreende-se que praticamente metade dos 64 estudos estão concentrados nos Estados Unidos. E, quando somados com a China, representam 70% do total. Por outro lado, vale dizer que do ponto de vista dos autores, não há uma concentração. São 221 autores envolvidos, dos quais 13 participam de duas publicações, e um é autor em quatro delas. Todos os demais só estão relacionados a um dos trabalhos.

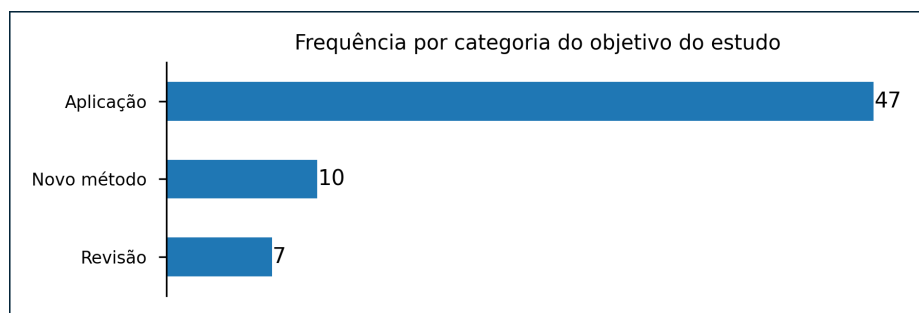
O escopo dos estudos selecionados é de aplicação prática de aprendizado de máquina na avaliação de política pública. Entretanto, nem todos tem esse entre os objetivos precípuos,

ainda que no conteúdo seja apresentado um caso de aplicação. Feito esse enquadramento, agrupamos os trabalhos em:

- novo método: caso tenha sido apresentado um novo método de aprendizado de máquina;
- revisão: se o objetivo era demonstrar a aplicação de métodos de aprendizagem de máquina ou realizar uma análise com método conhecido;
- aplicação: os demais casos. Vale ressaltar que todos apresentam uma aplicação, porém consideramos a especificidade do objetivo do estudo.

O resultado com a frequência de estudos por objetivo é apresentado no gráfico da [Figura 6](#). O quadro com a lista completa, contendo as codificações apresentadas, está no [Apêndice A](#).

Figura 6 – Quantidade de estudos por objetivo



Fonte: Elaborado pelo autor.

Todos os estudos agregam conhecimento, seja pela abordagem utilizada, seja pelas informações produzidas na avaliação. Contudo, os trabalhos de revisão e de novo método trazem mais explicação sobre os métodos empregados e, assim, podem ser de mais interesse para aqueles que querem conhecer detalhes sobre seu funcionamento.

Nessas duas categorias, predominam os métodos de aprendizado de máquina para inferência causal. Quanto aos novos métodos, [Stoffi e Gnecco \(2020\)](#) estendem um método para permitir tratar variáveis instrumentais. [Colangelo e Lee \(2020\)](#) estendem um método para tratar efeitos contínuos de tratamento. [Knaus \(2022\)](#) revisa e propõe um novo método duplamente robusto. [Heiler e Knaus \(2021\)](#) desenvolvem duas estimativas que permitem identificar efeitos heterogêneos quando analisadas em conjunto. [McFowland, Somanchi e Neill \(2023\)](#) propõem um método para descobrir a subpopulação mais afetada pelo tratamento em um experimento aleatório. [Singh, Xu e Gretton \(2020\)](#) propõem estimadores baseados no modelo de regressão de *kernel ridge* para funções causais não paramétricas de curvas de resposta. [Viviano \(2023\)](#) propõe método de alocação ótima de



tratamento de uma política na presença de *spillover*. [Grecov et al. \(2022\)](#) apresenta método para estimar os efeitos causais sobre múltiplas unidades tratadas. [Ballestar et al. \(2019b\)](#) propõem uma nova abordagem de agrupamento multinível de grupo de pesquisadores. [Belloni et al. \(2017\)](#) fornecem estimadores eficientes e faixas de confiança honestas para uma variedade de efeitos de tratamento, incluindo efeitos médios de tratamento local (LATE) e efeitos de tratamento local por quantil (LQTE) em ambientes ricos em dados.

No conjunto de revisão, [Shah, Kreif e Jones \(2021\)](#) proveem conhecimento e percepções sobre aprendizado de máquina causal. [Kreif e DiazOrdaz \(2019\)](#) revisam algoritmos populares de aprendizado de máquina. [Hu e Huang \(2018\)](#) tratam os desafios da extensão de AutoML aplicada à análise de impacto causal de avaliação de política. [Kallus e Zhou \(2019\)](#) demonstram a identificação de impacto discrepante de regras personalizadas. [Goller et al. \(2020a\)](#) investigam se algoritmos de aprendizado de máquina para estimar escore de propensão aprimoram a estimação. [Colangelo \(2021\)](#) estuda algoritmos de transferência de aprendizado e avalia a efetividade do estimador duplamente robusto para tratamento contínuo. [Hatamyar e Kreif \(2023\)](#) analisam diferentes métodos de aprendizado de máquina na tarefa de aprender regras de alocação que otimizem os efeitos de uma política.

Registramos também outras referências que, apesar de não terem sido selecionadas para o escopo desta dissertação, podem interessar ao leitor que deseja explorar mais o tema de inferência causal com aprendizado de máquina. [Yao et al. \(2021\)](#) realizaram uma revisão abrangente de métodos de inferência causal sob a estrutura de resultados potenciais. Além de métodos estatísticos tradicionais, também são abordados métodos de aprendizado de máquina. Nessa mesma linha, [Huber \(2021\)](#) apresentam uma visão de diferentes abordagens para avaliação de efeitos, incluindo a aplicação de aprendizado de máquina para inferência causal.

[Athey e Imbens \(2017\)](#) discutem desenvolvimentos recentes na econometria, incluindo o uso de aprendizado de máquina para inferência causal. A intersecção entre econometria e aprendizado de máquina tem gerado muito material com foco nos economistas. Nessa linha, se enquadram os trabalhos de [Varian \(2014\)](#), [Athey \(2019\)](#), [Athey e Imbens \(2019\)](#) e [Harding e Lamarche \(2021\)](#).

Outra visualização que realizamos foi a partir das palavras presentes nos títulos. A nuvem de palavras da [Figura 7](#) foi construída com os termos de cada um. Uma breve análise dos termos mais frequentes, que ocorrem mais de três vezes, mostram que os títulos estão relacionados ao escopo delimitado para este trabalho. Basicamente se verifica os seguintes termos mais frequentes relacionados a:

- avaliação: *evaluation, causal, effect, impact, heterogeneous, analysis, approach, treatment, policy evaluation, inference, model, data, study, framework, estimation*;

Figura 7 – Nuvem de palavras dos títulos



Fonte: Elaborado pelo autor.

Nota: Gerado utilizando o pacote *Wordcloud* para a linguagem Python (MUELLER, 2020). Foram excluídos termos avaliados como pouco informativos: ‘based’, ‘using’, ‘use’, ‘s’.

- de políticas públicas: *program*, *policy*, *intervention*, *public*, *policies*;
- usando aprendizado de máquina: *machine learning*;
- outros: *sustainable*, *case*, *china*, *health*, *network*, *system*.

## 5.2 Objetos de avaliação

### 5.2.1 Intervenção

Visando identificar o objeto das avaliações, realizamos a descrição das intervenções avaliadas em cada um dos estudos, conforme pode ser observado na coluna “Intervenção” do quadro do Apêndice B. A fim de facilitar a visualização de padrões, apresentamos o gráfico de nuvem de palavras mais frequentes nessa coluna na Figura 8.

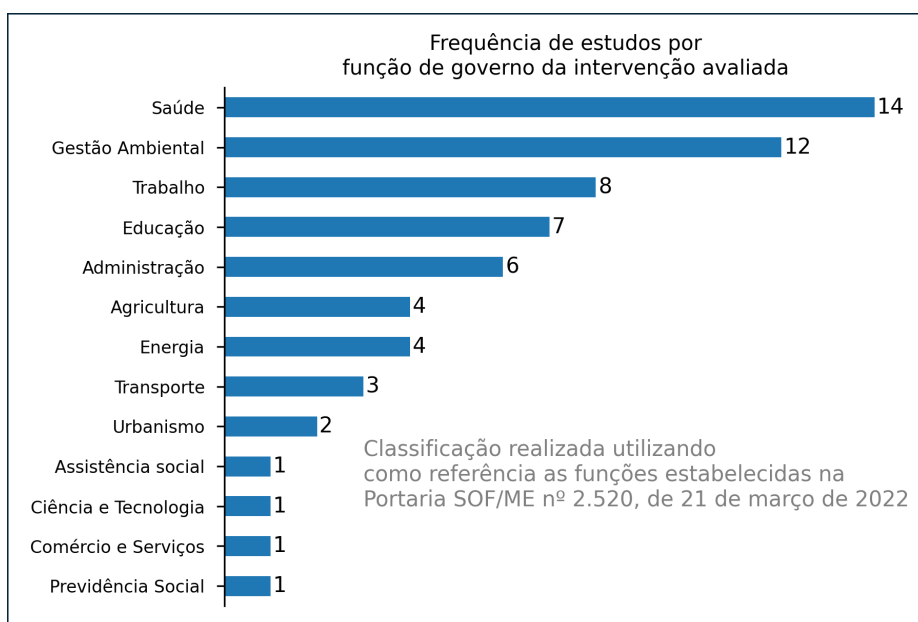
Destaca-se o termo programa como o mais citado seguido de política. Ainda com menos frequência, verifica-se plano, lei e *program*, todos também relacionados à formalização de ações de autoridade públicas. Outra perspectiva capturada, é a da localidade das ações, como Estados Unidos (terceiro termo mais frequente no geral), China, Alemanha, Suíça, Indonésia, Itália, Índia, Houston, Austrália, Londres, Pequim, *Beijing*, França, Toscana, África.



lockdown, resposta e pandemia referente a estudos que abordaram esse tema (WOZNY, 2022; WANG et al., 2021).

Como se visualiza na nuvem de palavras, são diversos os assuntos abordados. Assim, a fim de dar uma visão mais objetiva, codificamos as intervenções avaliadas utilizando como referência a classificação por função de governo presente na Portaria SOF/ME nº 2.520, de 21 de março de 2022 (Brasil, 2023). O resultado pode ser observado no gráfico da Figura 9.

Figura 9 – Quantidade de estudos por função de governo da intervenção avaliada



Fonte: Elaborado pelo autor.

Nota: A classificação foi realizada pelo autor de acordo com a proximidade da temática com as funções de governo presentes na Portaria SOF/ME nº 2.520, de 21 de março de 2022 (Brasil, 2023)

A leitura do gráfico indica que as funções de governo com mais aplicações são saúde e gestão ambiental. Na primeira, as intervenções compreendem serviços relacionados à seguro de saúde (CHEN et al., 2021; HATAMYAR; KREIF, 2023; KREIF; DIAZORDAZ, 2019; KREIF et al., 2022b; SHAH; KREIF; JONES, 2021), acesso a medicamento e meios de tratamento (CHANG, 2020; RUDOLPH et al., 2022; MARTINS et al., 2021; PAN et al., 2017), Covid-19 (WANG et al., 2021), combate ao tabagismo (CHEW et al., 2018; LINDEN; YARNOLD, 2018), atenção domiciliar (LINDEN; YARNOLD, 2016a) e medidas de acompanhamento (ZARE et al., 2021). A gestão ambiental abrange ações sobre poluição (ABRELL; KOSCH; RAUSCH, 2022; DAI et al., 2022; GUO et al., 2022; WANG et al., 2020; ZHANG et al., 2021), conservação (LU et al., 2022; RANA et al., 2022; SHARMA; SHEKHAR, 2021; YIN et al., 2018), resíduos sólidos (ALI; SHIRAZI, 2022; VALENTE, 2023) e controle de inundação (GORI et al., 2019).

As diferentes ações relacionadas à função trabalho tratam de capacitação para emprego (COCKX; LECHNER; BOLLENS, 2023; COLANGELO, 2021; COLANGELO; LEE, 2020; GOLLER et al., 2020a; HEILER; KNAUS, 2021; KALLUS; ZHOU, 2019; KNAUS, 2022; SINGH; XU; GRETTON, 2020).

Estudos de administração envolveram licença para comércio de bebidas alcoólicas (GRECOV et al., 2022), seguro de propriedade rural (VIVIANO, 2023), regulamento de licitação (MA; CHEN; LEI, 2022), demonetização (RAY et al., 2018) e desenvolvimento econômico (HU; HUANG, 2018; DUTHIL; IMOUSSATEN; MONTMAIN, 2017).

Na área de educação foram avaliadas ações quanto ao resultado do ensino (AGASISTI et al., 2022; BALLESTAR; SAINZ; SANZ, 2022; MCFOWLAND; SOMANCHI; NEILL, 2023), programa de pesquisa (BALLESTAR et al., 2019b), alimentação escolar (FRANCKE; ACOSTA, 2021; BRAHMA; MUKHERJEE, 2018) e financiamento (MA; ZHANG, 2014).

Agricultura e energia envolveram três assuntos cada. Na primeira, foram desenvolvimento rural (CARTER; TJERNSTROM; TOLEDO, 2019; STETTER; MENNIG; SAUER, 2022), regulação de preço (LI et al., 2021) e política agrícola (KUMAR; SHARMA, 2020). A segunda, energia limpa (HA; GRUBERT, 2023; MATENGA, 2022), transição energética (YANG et al., 2023) e malha de energia elétrica (RATLEDGE et al., 2021a).

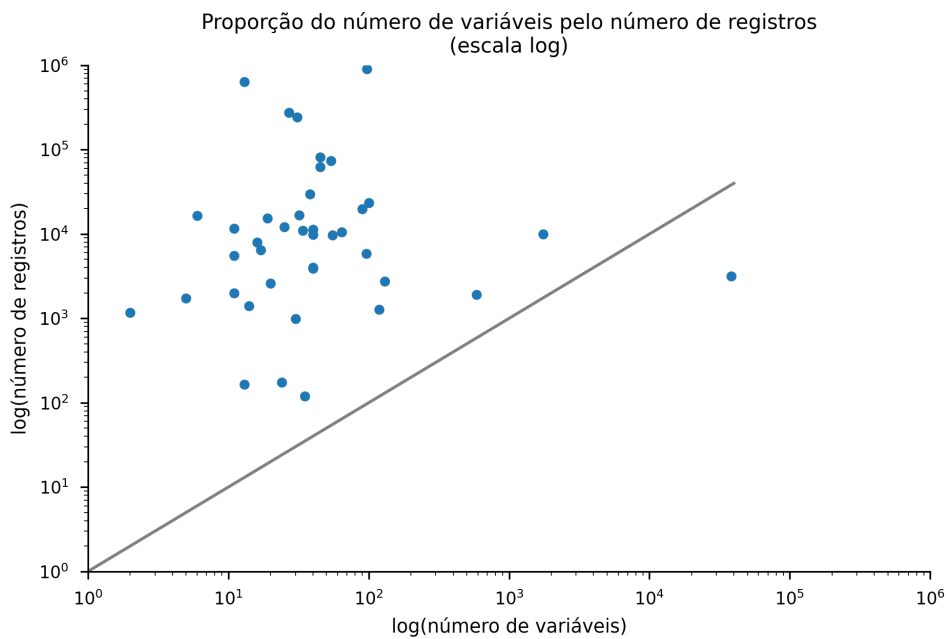
Na função de transporte, as intervenções estão associadas a tarifas de uso de transporte (WOZNY, 2022; CHEN; ZHOU, 2022) e controle de velocidade (ZHANG; LI; REN, 2022). Na de urbanismo, a prevenção de deslocamento de moradores (AUERBACH et al., 2020) e plano de mobilidade urbana (BRONDEEL; KESTENS; CHAIX, 2017). Para as demais, identificamos uma ação por assunto. Na de ciência e tecnologia, foi plano de trabalho de comunidade de inovação (GAO; LI; WANG, 2021). Em previdência social, uma programa previdência (BELLONI et al., 2017). Em comércio e serviços, acesso a crédito (STOFFI; GNECCO, 2020). E, finalmente, em assistência social, suporte a cuidadores (SHEPHERD-BANIGAN et al., 2022).

### 5.2.2 Conjunto de dados

Ainda com relação ao objeto de avaliação, consolidamos no gráfico da [Figura 10](#) as informações mapeadas sobre os conjuntos de dados. Registramos que nem sempre essa informação está clara nos estudos. Logo, para alguns deles não foi possível quantificar as variáveis e/ou registros precisamente. Há casos em que os autores apresentam de forma clara o número exato de variáveis e registros, há outros em que se inferiu um montante a partir da contabilização de menções ao longo do texto. E outros que a informação não é mencionada.

Na avaliação de política pública por meios econométricos a partir de dados ob-

Figura 10 – Quantidade de estudos pelo tamanho do conjunto de dados



Fonte: Elaborado pelo autor.

Nota: Nem sempre a quantidade de registros e variáveis está clara nas publicações. Em alguns casos não foi possível quantificar as variáveis ou registros precisamente. Não foi possível identificar o número de registros em 11 estudos e de variáveis em 16. O Gráfico contabiliza os 40 estudos em que ambas as informações estavam presentes. A linha perpassa os pontos em que x e y coincidem, ou seja, o número de registros seria igual ao de variáveis.

servados com grande número de covariáveis, é comum ser aplicado alguma estratégia de balanceamento dos grupos de tratamento e controle, como score de propensão, pois o erro na estimação do impacto das covariáveis leva a erros muito grandes (ATHEY; IMBENS, 2017). Para um estimador baseado em mínimos quadrados ordinários, quando o número de variáveis é maior que o número de registros, pode haver um número infinito de soluções e a estimação ser ruim por conta do *overfitting* (HARDING; LAMARCHE, 2021; BÜHLMANN; GEER, 2011).

Uma vantagem geralmente apontada para os modelos de aprendizado de máquina em relação a modelos econométricos é a possibilidade de lidar com grandes conjuntos de dados pela utilização de técnicas de regularização e validação cruzada (MULLAINATHAN; SPIESS, 2017). No gráfico da Figura 12, observa-se um estudo que se localiza abaixo da linha cinza traçada sobre os pontos em que o número de variáveis é igual ao de registros. Trata-se de avaliação que tinha como objetivo identificar e entender as comunidades seguidoras de campanha pública realizada por meio da Plataforma Twitter (CHEW et al., 2018). Um modelo de aprendizado de máquina foi treinado, usando os termos mais relevantes de postagens de usuários seguidores da conta da campanha, para classificar o

público de acordo com sua faixa etária. Os termos, utilizados como variáveis, totalizaram 38.200, enquanto o número de registros era 3.184.

Nesse cenário, métodos de aprendizado de máquina abrem caminho para se trabalhar com diferentes tipos de dados que, naturalmente, possuem muitas dimensões para métodos de estimação tradicionais da econometria, como regressões lineares (MULLAINATHAN; SPIESS, 2017). O tratamento de textos, como visto acima, tende a ter alta dimensão. Além do exemplo visto acima, alguns dos estudos apresentaram análises de opinião pública por meio do texto de postagem em redes sociais (KUMAR; SHARMA, 2020; SHARMA; SHEKHAR, 2021; RAY et al., 2018; CHEW et al., 2018). Ainda na linha de análise de sentimento, há aplicação em dados de audiência pública onde o objetivo inicial era verificar um método híbrido de codificação qualitativa com aprendizado de máquina (HA; GRUBERT, 2023). Outra abordagem com dados textuais é o processamento de documentos oficiais, normas e artigos científicos com a finalidade de identificar termos relevantes para subsidiar processo de avaliação (GAO; LI; WANG, 2021; ALI; SHIRAZI, 2022; LI et al., 2021).

Outro tipo de dado trabalhado é o de imagens de satélite, principalmente em avaliações relacionadas à gestão ambiental (WANG et al., 2020; ZHANG et al., 2021; YIN et al., 2018; LU et al., 2022; GORI et al., 2019). Em outra frente, Ratledge et al. (2021a) utilizaram imagens de satélite para treinar um modelo de aprendizado de máquina com o objetivo de imputar valores faltantes de um indicador de meios de sustento para regiões da África Subsaariana, que era a variável de resultado para a qual se pretendia verificar os efeitos da expansão da malha energética na região. Trata-se de uma estratégia econômica para avaliação de impacto em regiões com dados esparsos.

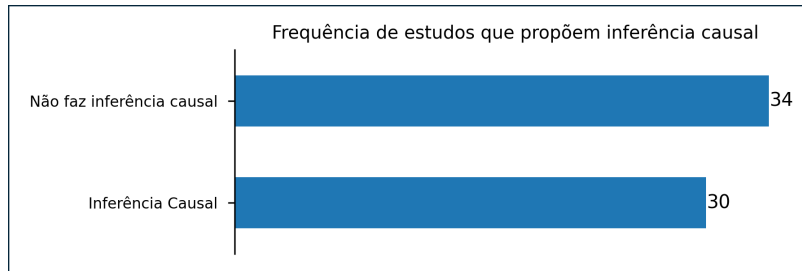
## 5.3 Métodos de aprendizado de máquina

### 5.3.1 Objetivo da aplicação de aprendizado de máquina

Métodos de aprendizado de máquina se destacam pela sua capacidade em aprender modelos preditivos dirigidos a dados com alta dimensionalidade, devido a sua capacidade de inferir formas funcionais flexíveis e estratégias de regularização. Todavia, apesar de tentador, o modelo gerado não deve ser usado diretamente para inferência causal, pois das mesmas vantagens decorrem desvantagens, como inconsistência dos coeficientes (MULLAINATHAN; SPIESS, 2017).

Por outro lado, novos métodos, apresentados adiante, têm sido propostos e aplicados na inferência causal. Nesse contexto, registramos que em torno da metade dos estudos selecionados buscam inferir causalidade, enquanto a outra não, conforme gráfico da Figura 11.

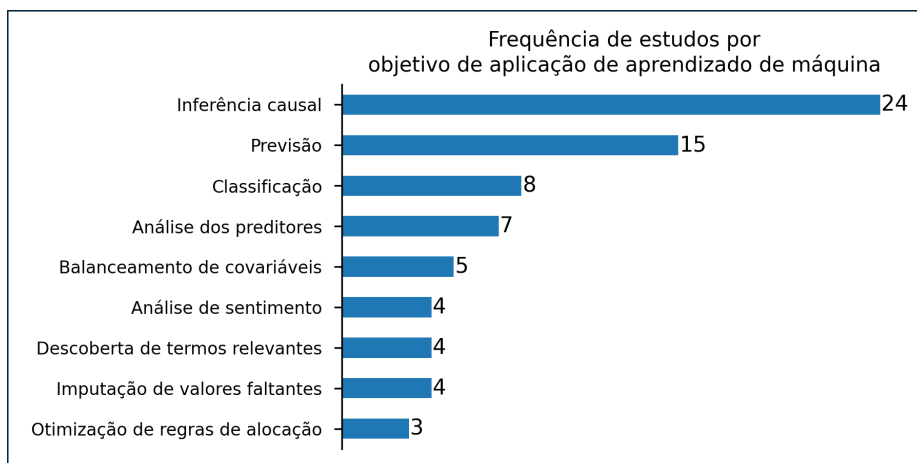
Figura 11 – Quantidade de estudos pelo tipo de análise (causal ou não)



Fonte: Elaborado pelo autor.

A fim de detalhar melhor os objetivos de aplicação, classificamos os estudos selecionados de acordo com o objetivo de aplicação de aprendizado de máquina segundo uma codificação própria. As frequências podem ser observadas no gráfico da [Figura 12](#) e a categorização individual no [Apêndice C](#).

Figura 12 – Quantidade de estudos por objetivo da aplicação de aprendizado de máquina



Fonte: Elaborado pelo autor.

Nota: Codificação criada pelo autor a partir da análise do objetivo da aplicação de aprendizado de máquina em cada estudo. Um estudo pode estar classificado em mais de um objetivo.

Para **inferência causal**, os estudos exploram estimação de efeitos médios de tratamento ([LINDEN; YARNOLD, 2016a](#); [LINDEN; YARNOLD, 2018](#); [RATLEDGE et al., 2021a](#); [KREIF; DIAZORDAZ, 2019](#); [GRECOV et al., 2022](#); [RUDOLPH et al., 2022](#)), incluindo função média de resposta à dose de tratamento de variável contínua ([COLANGELO; LEE, 2020](#); [COLANGELO, 2021](#)) e função quantil ([GRECOV et al., 2022](#); [BELLONI et al., 2017](#)).

A quantidade de dados disponível cada vez maior fez com que o problema de estimação de efeitos causais heterogêneos por subgrupos populacionais passasse a ser mais



explorado nos últimos tempos. Com a utilização de métodos econométricos tradicionais, os subgrupos precisavam ser especificados com antecedência. Utilizando modelos de aprendizado de máquina os subgrupos são selecionados automaticamente pelos próprios algoritmos (STOFFI; GNECCO, 2020).

Colocando de outra forma, o ônus de estabelecer quais hipóteses testar deixa de ser do pesquisador e passa a ser da máquina em gerar novas hipóteses. Parafraseando, Tukey (1980), “ideias surgem mais de explorações prévias do que na forma de relâmpagos” (MCFOWLAND; SOMANCHI; NEILL, 2023). O questionamento deixa de ser se a política é efetiva e passa a ser para quem ela é efetiva (CHEN et al., 2021).

Verifica-se vários estudos que avaliam a presença de efeitos heterogêneos por meio da estimação de efeitos de tratamento médio condicional (CATE) por grupo (GATE) e/ou individualizado (IATE) (KALLUS; ZHOU, 2019; CHEN et al., 2021; SHAH; KREIF; JONES, 2021; VALENTE, 2023; KREIF et al., 2022b; CHANG, 2020; ZHANG; LI; REN, 2022; CARTER; TJERNSTROM; TOLEDO, 2019; STETTER; MENNIG; SAUER, 2022; STOFFI; GNECCO, 2020; KNAUS, 2022; SHAH; KREIF; JONES, 2021; KNAUS, 2022; SHAH; KREIF; JONES, 2021; COCKX; LECHNER; BOLLENS, 2023), efeitos médios de tratamento naturais (nATE) em conjunto com efeitos médios de tratamento aleatórios (rATE) (HEILER; KNAUS, 2021), curva de resposta heterogênea (SINGH; XU; GRETTON, 2020), efeitos médios de tratamento não-paramétrico (NATE)(MCFOWLAND; SOMANCHI; NEILL, 2023).

A **otimização de regras de alocação** tem enfoque semelhante a de efeitos heterogêneos, porém com o objetivo de buscar a melhor política de atribuição ao tratamento, ou seja, quais regras podem maximizar os efeitos desejados (HATAMYAR; KREIF, 2023). Knaus (2022) faz uma revisão de várias aplicações de *Double Machine Learning*, incluindo a ilustração de otimização de regra de alocação. Hatamyar e Kreif (2023) analisa essa tarefa em uma configuração de resultados raros, a exemplo do impacto de um programa de seguro de saúde na mortalidade infantil. Viviano (2023) trata o desafio em um cenário com presença de *spillover* na avaliação.

Na inferência causal, uma das tarefas comumente realizadas é o balanceamento entre os grupos de controle e tratamento. Nesse sentido, um objetivo de aplicação que identificamos foi no **balanceamento de covariáveis**. Estudos demonstraram que métodos de aprendizado de máquina superaram métodos de regressão logística e de balanceamento de propensão de covariáveis em modelos complexos com relações não lineares (SETODJI et al., 2017; MCCAFFREY; RIDGEWAY; MORRAL, 2004). Chang (2020), Goller et al. (2020a) e Rudolph et al. (2022) utilizam aprendizado de máquina como estimador do escore de propensão. Já Linden e Yarnold (2016a) aplicam o modelo de aprendizado de máquina para analisar o balanceamento após a estimação do escore de propensão de pareamento. Linden e Yarnold (2018) novamente utilizam o mesmo método de balanceamento, porém

sem aplicar escore de propensão dessa vez. Além desses casos, com objetivo específico de balanceamento, avaliações que utilizam *Double Machine Learning* também utilizam escore de propensão como parte do processo, permitindo a utilização de modelos de aprendizado de máquina para isso (KNAUS, 2022).

A partir daqui, os objetivos de aplicação podem estar compreendidos tanto no âmbito de uma avaliação de política pública que busca inferir a causalidade ou não.

A aplicação do modelo de aprendizado de máquina pode ocorrer ainda na etapa de preparação do conjunto de dados. Um caso interessante é a **imputação de valores faltantes** da variável resultado de meios de sustento de vida obtida por Ratledge et al. (2021a), pois essa variável, originalmente obtida por meio de pesquisa de campo com moradores da África subsaariana, foi imputada utilizando dados externos de imagens de satélite. Os demais casos foram imputados com base em dados do próprio conjunto de dados (WANG et al., 2020; WOZNY, 2022; STETTER; MENNIG; SAUER, 2022).

Ainda na linha de dados faltantes, mas sem recorrer à imputação de valores, vale destacar a aplicação de transferência de conhecimento realizada por Colangelo (2021). Sua estratégia foi utilizar informações de estimação de uma variável auxiliar de resultado, correlacionada com o resultado alvo, porém que possui mais observações, para aprimorar a estimação dos efeitos médios do resultado desejado.

Na **classificação**, agregamos os estudos que tiveram como objetivo subsidiar a avaliação com a rotulação da população, sendo que a avaliação em si seria realizada por algum outro método além do aprendizado de máquina. Destacam-se nessa categoria os métodos não supervisionados para divisão da população avaliada em grupos. Matenga (2022) aplicou para agrupamento de países, (BALLESTAR et al., 2019b) para pesquisadores universitários, Ma e Zhang (2014) para agentes de escola e Chen e Zhou (2022) para usuários de metrô. Todavia, há aplicações também de classificação supervisionada. Exemplos incluem a classificação de usuários de rede social de acordo com sua faixa etária (CHEW et al., 2018), de imagens de satélite quanto ao uso e cobertura da terra (YIN et al., 2018; LU et al., 2022), de indicadores relacionados a textos de documentos oficiais (GAO; LI; WANG, 2021).

Numa abordagem semelhante, porém com finalidade de prever os valores de variáveis a serem analisada por outros métodos, agrupamos os estudos em uma categoria chamada de **previsão**. Importante dizer que, caso a finalidade da previsão seja das categorias específicas, o objetivo de aplicação não foi enquadrado nessa mais geral. Por exemplo, não entram aqui previsões para imputação de valores faltantes, balanceamento de covariáveis ou análise de preditores. Destacamos que um subgrupo desses estudos simulam cenários futuros ou com variações nas características da população. Em que pese estudos baseados em simulação sofrerem por conta de variáveis desconhecidas, como adaptação dos participantes, efeitos de longo prazo, mudanças não intencionais que impactam o resultado, eles são um

complemento aos estudos empíricos e propiciam a comparação de diferentes cenários com baixo custo (BRONDEEL; KESTENS; CHAIX, 2017).

Brondeel, Kestens e Chaix (2017) simulam cenários de modais de transporte. Rana et al. (2022) projetam desperdícios de recursos de reforestamento. Wang et al. (2021) projetam mortalidade decorrente de políticas de lockdown para o enfrentamento da Covid-19. Gori et al. (2019) preveem o desenvolvimento de setores de uma bacia hidrográfica para projeção de planície de inundação. Auerbach et al. (2020) projetam imposto futuro sobre propriedades para estimar cenários de custos do programa. Yang et al. (2023) preveem cenários de tendência na transição para energia renovável. Abrell, Kosch e Rausch (2022) simulam cenários de abatimento e impacto decorrentes de diferentes níveis de imposto de carbono. Esses são alguns exemplos, recomendamos ao leitor verificar a lista completa de aplicações desta categoria no quadro do Apêndice C.

A **análise de preditores** compreende casos de avaliação das covariáveis mais relevantes para o resultado desejado. Zare et al. (2021), Ballestar, Sainz e Sanz (2022), Brahma e Mukherjee (2018) e Agasisti et al. (2022) avaliaram programas na área da educação relacionados com a identificação de fatores que contribuem para obesidade infantil, cobertura de crianças na alimentação escolar e desempenho escolar em intervenções específicas. Agasisti et al. (2022) destacam que, utilizar o aprendizado de forma complementar ao método dos mínimos quadrados ordinários para inferência causal, permitiu explorar outras relações entre as variáveis ao relaxar as restrições de linearidade exigidas por aquele método. Outros exemplos incluem a identificação de cláusulas legais relacionadas à alta taxa de dispensação de opioides (MARTINS et al., 2021), de variáveis importantes para o recebimento de resultados de testes de HIV por usuários de substância (PAN et al., 2017) e de impulsionadores de sintomas depressivos, de sobrecarga subjetiva e de resolução de problema em cuidadores de um programa de suporte a veteranos de guerra (SHEPHERD-BANIGAN et al., 2022).

Já explorando dados textuais, temos a **análise de sentimento** e descoberta de termos relevantes. Ha e Grubert (2023) utilizaram dados de comentários transcritos dos participantes de audiência pública para avaliar o suporte a um plano de energia limpa nos Estados Unidos. Contudo, a fonte de dados mais comum nesse caso são as redes sociais, especificamente a Plataforma Twitter<sup>1</sup> (KUMAR; SHARMA, 2020; SHARMA; SHEKHAR, 2021; RAY et al., 2018). Sharma e Shekhar (2021) avaliam que, enquanto outras plataformas são mais restritas, o Twitter promove a comunicação aberta entre os usuários. Ademais, à época de sua avaliação de programa da Índia, era a única utilizada para questões governamentais. Ray et al. (2018), que também realizaram avaliação na Índia, frisam que, apesar de os dados cobrirem apenas um segmento da população, existiam 292 milhões de usuários de redes sociais no país. Portanto, de alguma forma, capturaria o

---

<sup>1</sup> <<https://www.twitter.com>>

sentimento da população.

A aplicação de aprendizado de máquina com o objetivo de **descoberta de termos relevantes** utiliza predominantemente dados não estruturados e, de forma geral, objetiva a construção de indicador comparativo utilizando como critérios os termos descobertos (GAO; LI; WANG, 2021; ALI; SHIRAZI, 2022; LI et al., 2021; DUTHIL; IMOUSSATEN; MONTMAIN, 2017). Gao, Li e Wang (2021) e Li et al. (2021) utilizam os resultados do modelo para montar um indicador utilizando a abordagem *Policy Modeling Consistency Index* (ESTRADA, 2011). No caso de Ali e Shirazi (2022), os termos compuseram o componente de saída de um modelo de análise envoltória de dados. Por fim, Duthil, Imoussaten e Montmain (2017) utilizaram segmentos de texto relevantes como insumo para aplicação da teoria de possibilidade na avaliação da priorização de ações relacionadas ao desenvolvimento sustentável por municípios franceses.

Com o objetivo de facilitar o acesso do leitor aos estudos obtidos por meio de bola de neve, incluímos no [Apêndice E](#) uma coluna descrevendo quais aplicações de aprendizado de máquina foram identificadas no estudo de onde a referência foi extraída originalmente. Dessa forma, o leitor que deseja mais material sobre as formas de aplicações pode consultar essa listagem. Vale lembrar que os estudos não foram analisados profundamente para além da identificação como potencial relação com o escopo deste trabalho. Ademais, a aplicação de aprendizado de máquina dada no estudo de origem pode não ser a mesma dada no trabalho referenciado. Outra ressalva é que a referência pode ter sido citada em outros trabalhos, porém só foi registrada a citação originária, ou seja, do primeiro estudo em que foi vista.

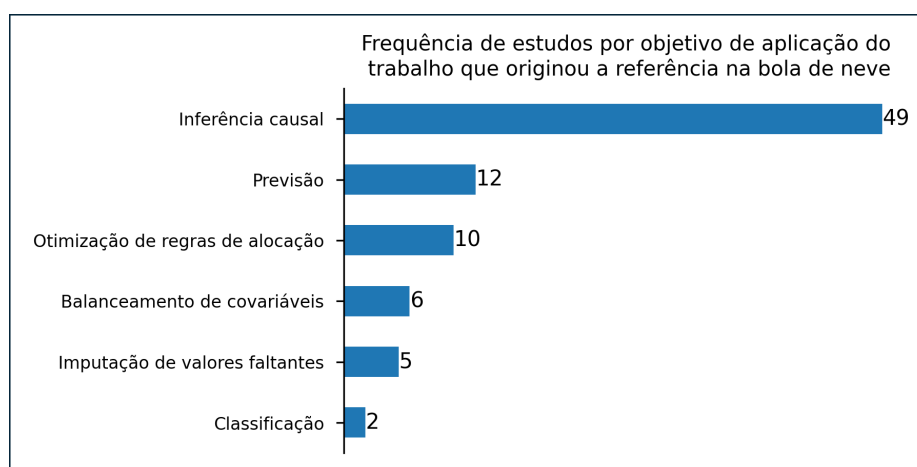
Tendo essas considerações em mente, apresentamos, no gráfico da [Figura 13](#), a frequência do objetivo de aplicação dos estudos que originaram as referências obtidas por meio de bola de neve.

Nota-se a predominância de estudos decorrentes de inferência causal. A despeito de ser o esperado, pois também foi o objetivo de aplicação mais frequente nos estudos selecionados, verifica-se ser um método que está sendo bastante explorado na avaliação de políticas públicas. Uma possível explicação é o fato de a inferência causal ser historicamente um dos métodos econométricos utilizados na avaliação de políticas públicas.

### 5.3.2 Métodos de aprendizado de máquina

Os métodos de aprendizado de máquina mais utilizados nos estudos selecionados são os supervisionados, os quais predizem um valor com base em um histórico conhecido da variável dependente. Juntamente, com uma frequência menor, observamos a utilização de métodos não supervisionados, aqueles para os quais não há rotulação prévia dos dados. Finalmente, há uma aplicação do aprendizado por reforço, em que é estabelecida uma

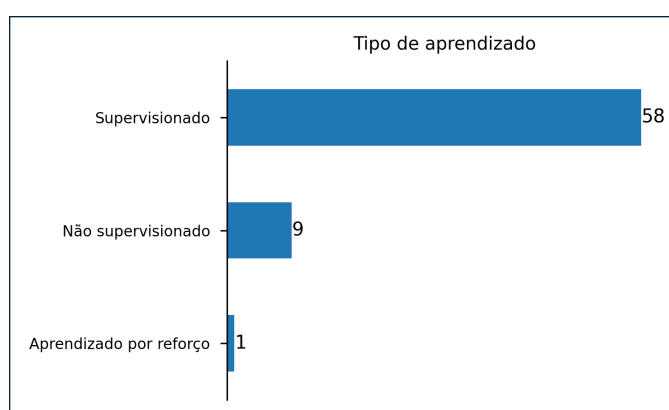
Figura 13 – Quantidade de estudos (bola de neve) por objetivo da aplicação de aprendizado de máquina da referência originária



Fonte: Elaborado pelo autor.

recompensa para a maximização do aprendizado. A Figura 14 apresenta a frequência de estudos por caso de uso. Ressalta-se que há casos em que mais de um tipo foi empregado no mesmo estudo (MA; ZHANG, 2014; BALLESTAR et al., 2019b; GAO; LI; WANG, 2021; COCKX; LECHNER; BOLLENS, 2023).

Figura 14 – Quantidade de estudos por tipo de aprendizado do método utilizado



Fonte: Elaborado pelo autor.

Nota: A quantidade somada é maior que a quantidade de estudos selecionados, pois alguns aplicaram mais de um tipo de método.

Identificamos apenas uma aplicação para o aprendizado por reforço. Ele foi utilizado por Ma e Zhang (2014) em um modelo hierárquico definido com uma política de recompensa como função do desempenho dos alunos a fim de verificar a relação entre os recursos empregados e o desempenho dos alunos.

O aprendizado não supervisionado é aplicado em dados textuais na descoberta de termos relevantes em documentos (GAO; LI; WANG, 2021; ALI; SHIRAZI, 2022; LI et al., 2021; DUTHIL; IMOUSSATEN; MONTMAIN, 2017). Além disso, na divisão da amostra de análise. Há divisão de grupos de países (MATENGA, 2022), agentes de ensino (MA; ZHANG, 2014), pesquisadores universitários (BALLESTAR et al., 2019b), usuários de metrô (CHEN; ZHOU, 2022), efeitos médios de tratamento individualizados (COCKX; LECHNER; BOLLENS, 2023).

Sublinha-se a aplicação no agrupamento de usuários de metrô por Chen e Zhou (2022). Eles aplicaram o método *Latent Dirichlet Allocation* (BLEI; NG; JORDAN, 2003) comumente aplicado em documentos de texto, como em Li et al. (2021). Os usuários de metrô foram considerados como documentos para o modelo. As características dos indivíduos foram consideradas como palavras, em um modelo foram consideradas as características relacionadas ao tempo de viagem e em outro as geográficas. A combinação do tópico mais relevante de cada modelo foi utilizada para estabelecer o perfil espaço-temporal de cada usuário, gerando 16 grupos. Os grupos foram utilizados para as análises estatísticas de objetivo, padrões e receitas.

Os demais estudos aplicam o aprendizado supervisionado, que se assemelha a regressões econométricas empregadas tradicionalmente na avaliação de políticas públicas. Convidamos o leitor a verificar a classificação completa no Apêndice C. A Figura 15 apresenta a frequência dos métodos identificados em ao menos dois estudos e a Figura 16 relaciona os mencionados uma vez.

Os métodos de aprendizado de máquina são tradicionalmente otimizados para predição e, se aplicados com esse objetivo, não são capazes de responder questões que envolvem causalidade. Questões de causalidade são o cerne das avaliações de impacto, onde quando se busca atribuir o resultado de determinada intervenção, ou seja, se as mudanças no bem-estar dos indivíduos podem ser atribuídas à política pública (GERTLER et al., 2018).

Um bom exemplo dado por Athey (2019) para ilustrar a diferença entre predição e causalidade é o da ocupação de quartos de hotel. Se uma rede de hotéis quer estimar a taxa de ocupação dos seus competidores com base nos preços anunciados, o modelo preditivo é a opção. Intuitivamente, espera-se que os preços aumentem conforme aumente a ocupação. No entanto, se a gestão deseja saber qual o efeito na ocupação se o preço for intencionalmente acrescido, essa é uma questão de inferência causal que não pode ser respondida por um modelo tradicional de aprendizado de máquina.

Kleinberg et al. (2015) apresentam dois casos hipotéticos que ilustram essa distinção e a aplicabilidade das duas respostas. No primeiro, um gestor está passando por um período prolongado de seca e precisa decidir se investe em dança da chuva para aumentar a chance de chover. Esse é um problema de causalidade. Qual o efeito da dança na probabilidade

Figura 15 – Quantidade de estudos por métodos utilizados



Fonte: Elaborado pelo autor.

Nota: São apresentadas as frequências para os métodos que apareceram em mais de um estudo. Em alguns casos foram usados vários algoritmos e selecionado o modelo que obteve melhor desempenho, a exemplo de [Dai et al. \(2022\)](#) e [Kumar e Sharma \(2020\)](#).

de chover? Por outro lado, imagine a situação em que o gestor está visualizando nuvens escuras no céu e deve decidir se carregará um guarda-chuva para o trabalho para evitar se molhar no retorno para casa. Nesse caso, basta saber se a previsão de chuva é alta o suficiente.

A exploração de aprendizado de máquina para inferência causal tem agregado novos métodos que alteram métodos preditivos ou que permitem os acoplar como componente de um estimador causal. Por exemplo, estimadores causais podem envolver modelos intermediários de parâmetros necessários para estimação do parâmetro alvo, como escores de propensão e regressão do resultado. Esses modelos intermediários, chamados de perturbadores, podem se valer de aprendizado de máquina para agregar vantagens, como a seleção automática de modelos e transparência do processo de pesquisa ([KREIF; DIAZORDAZ, 2019](#)).

Na linha de métodos que agregam outros métodos, notamos a aplicação de *Meta Learners*, formado pela família de *T*-, *S*- e *X-Learner*, e *R-Learner* para estimação de efeitos de tratamento heterogêneos condicionais ([SHAH; KREIF; JONES, 2021](#); [VALENTE, 2023](#)).

Figura 16 – Outros métodos utilizados

|   |   |
|---|---|
|   | Logistic Regression                             |
| AdaBoost Regression                                     | Long Short-Term Memory Recurrent Neural Network |
| Agglomerative Hierarchical Clustering                   | MAXQ Value Function Decomposition               |
| Artificial Neural Network                               | Matrix Completion                               |
| Automated Nested Longitudinal Clustering                | Meta-Learners                                   |
| BP Neural Network Optimized by Annealing Algorithm      | Modified Causal Forest                          |
| Bidirectional Encoder Representations from Transformers | Multiple Additive Regression Splines            |
| Causal Tree with Instrumental Variable                  | Multiple Linear Regression                      |
| Chi-square Automatic Interaction Detector               | Recursive Partitioning                          |
| Conv2LSTM   | Super Learner                                   |
| Elastic Net Regression                                  | Support Vector Regression                       |
| Feedforward Artificial Neural Network                   | Synopsis  |
| Generalized Linear Models                               | Synthetic Controls with Elastic Net             |
| Gradient Boosted Decision Trees                         | Treatment Effect Subset Scan                    |
| Honest Causal Tree with Instrumental Variable           | U-Net Convolutional Neural Network              |
| K Deep Neural Network                                   |   |
| Kernel Ridge Regression                                 |   |
| LASSO hierárquico                                       |   |
| LDA2Vec   |   |
| Least Absolute Selection and Shrinkage Operator         |   |

Fonte: Elaborado pelo autor.

Nota: A relação de métodos que foram mencionados em apenas um estudo.

Destaca-se também um conjunto de estimadores duplamente robustos que permitem usar aprendizado de máquina na estimação do escore de propensão e regressão do resultado, baseados em *Double Machine Learning*, que podem ser utilizados para estimação de efeitos médios para tratamentos binários ou contínuos, estimação de efeitos heterogêneos e estimação de regras de otimização de alocação (COLANGELO; LEE, 2020; KNAUS, 2022; COLANGELO, 2021; KREIF; DIAZORDAZ, 2019; HATAMYAR; KREIF, 2023). *Random Forest* e *Least Absolute Shrinkage and Selection Operator* são dois algoritmos frequentemente agregados por esses métodos, o que contribuiu para a popularidade de ambos no gráfico da Figura 15.

Quanto a novos métodos decorrentes de adaptações de modelos de aprendizado de máquina, identificamos um conjunto de métodos baseados em árvores, como *Random Forest* (BREIMAN, 2001), para estimação de efeitos heterogêneos. Destacam-se *Causal Forest* (CHEN et al., 2021; SHAH; KREIF; JONES, 2021; VALENTE, 2023; HATAMYAR; KREIF, 2023; COCKX; LECHNER; BOLLENS, 2023; STETTER; MENNIG; SAUER,



2022), *Generalized Random Forest* (KALLUS; ZHOU, 2019; COLANGELO; LEE, 2020; COLANGELO, 2021; KREIF et al., 2022b; ZHANG; LI; REN, 2022; CARTER; TJERNSTROM; TOLEDO, 2019) e extensões para tratamento de variáveis instrumentais (STOFFI; GNECCO, 2020).

Alguns estudos propõem novos métodos que não estão compreendidos entre os anteriores. Heiler e Knaus (2021) desenvolvem duas estimativas que permitem identificar efeitos heterogêneos quando analisadas em conjunto. McFowland, Somanchi e Neill (2023) propõem um método para descobrir a subpopulação mais afetada pelo tratamento em um experimento aleatório. Singh, Xu e Gretton (2020) propõem estimadores baseados no modelo de regressão de kernel ridge para funções causais não paramétricas de curvas de resposta. Viviano (2023) propõe método de alocação ótima de tratamento de uma política na presença de spillover. Grecov et al. (2022) apresentam método para estimar os efeitos causais sobre múltiplas unidades tratadas. Belloni et al. (2017) fornecem estimadores eficientes e faixas de confiança honestas para uma variedade de efeitos de tratamento, incluindo efeitos médios de tratamento local (LATE) e efeitos de tratamento local por quantil (LQTE) em ambientes ricos em dados.

Ratledge et al. (2021a) utilizam dois métodos de aprendizado de máquina para inferência causal na mesma avaliação para estimar efeitos médios de tratamento, são eles *Matrix Completion* (MC) e *Synthetic Controls with Elastic Net*. Além desses, os autores compararam o desempenho com um estimador de diferença em diferenças. O MC foi eleito o melhor entre os três para o caso analisado.

Constata-se que mais de um método pode ser adequado para um mesmo desenho de avaliação. Linden e Yarnold (2018) lecionam que o verdadeiro efeito de tratamento nunca é conhecido em experimentos não aleatórios. Dito isso, uma análise de sensibilidade que pode ser conduzida por pesquisadores é a aplicação de diversos métodos para verificar a robustez dos resultados. Linden e Yarnold (2016a), por exemplo, que aplicaram *Optimal Discriminant Analysis* (ODA) no balanceamento de covariáveis e estimação de efeitos de tratamento, afirmam que se os resultados comparados com mínimos quadrados ordinários for consistente, o pesquisador pode confiar que a estimativa é possivelmente um reflexo do efeito de tratamento verdadeiro. Por outro lado, se forem conflitantes, eles recomendam considerar a estimativa do ODA como mais robusta pelo fato de usar permutação de p-valores, que não requerem suposições de distribuição e são sempre válidas. De maneira semelhante, Ballestar et al. (2019b) treinam um modelo supervisionado para validar a acurácia da classificação descoberta por um modelo não supervisionado sobre os mesmos dados.

No que diz respeito à inferência causal, a automação propiciada pelos métodos de aprendizado de máquina não são uma bala de prata para as suposições e cuidados necessários à estimação. Por exemplo, se os efeitos causais são realmente identificados a

partir dos dados, se todos confundidores foram mensurados, quais as estratégias aplicáveis, quais considerações devem ser incorporadas no caso concreto, quais as métricas de resultado refletem os objetivos e definição do intervalo de confiança. Por outro lado, a intersecção das duas disciplinas traz vantagens para ambas. A seleção automatizada de modelos traz eficiência e consistência documentada para o pesquisador, e as preocupações da econometria provocam novas evoluções na área de aprendizado de máquina (ATHEY, 2019).

Finalmente, Hu e Huang (2018) relacionam uma série de desafios para aplicação de AutoML na inferência causal. AutoML é a automatização do processo analítico, compreendendo, por exemplo, pré-processamento dos dados, seleção de algoritmos para construção do modelo, otimização de parâmetros e hiperparâmetros, interpretação de resultado. Essa, portanto, é uma linha de pesquisa promissora dentro do campo de aprendizado de máquina para inferência causal visando otimizar os processos de avaliação.

### 5.3.3 Linguagens e recursos de programação

Analisamos quais linguagens de programação e pacotes foram utilizados nos estudos. Importante dizer que nem sempre esses recursos são mencionados. Na prática, conseguimos mapear a linguagem e pacotes, respectivamente, em 42 e 35 dos 64 trabalhos. O resultado está apresentado no gráfico de frequência de linguagem de programação à esquerda e nuvem de palavras dos pacotes à direita da Figura 17.

É evidente que a linguagem R é a mais utilizada no subconjunto de estudos observado, seguida de Python. A soma das demais não chega à metade da segunda colocada. Interessante observar que alguns trabalhos se valem de mais de uma linguagem de programação: R e Matlab (BELLONI et al., 2017), R e SAS (SHEPHERD-BANIGAN et al., 2022), R e Stata (CHANG, 2020), Python e Stata (ALI; SHIRAZI, 2022) e Python e R (DAI et al., 2022; COLANGELO; LEE, 2020; COLANGELO, 2021; SINGH; XU; GRETTON, 2020; GRECOV et al., 2022; KALLUS; ZHOU, 2019).

Destaca-se que, na Figura 17, as cores da nuvem de palavras dos pacotes estão relacionadas com as linguagens de programação apresentadas no gráfico de frequência de estudos por linguagem de programação. A cor preta é utilizada para softwares não relacionados a uma linguagem especificamente.

Desta feita, chama a atenção que, assim como nas frequências, os pacotes são predominantemente das linguagens R e Python. Há apenas um pacote para a linguagem Stata, ITSAMATCH, para pareamento em análise de série temporal interrompida. Quanto aos demais, organizamos em uma classificação simplificada exposta no Quadro 8. Cabe dizer que, mesmo havendo sobreposição, fizemos a divisão entre inferência causal, processamento de linguagem natural, aprendizado de máquina e cálculo e estatística, sempre considerando a especialidade nessa ordem.



Quadro 8 – Categorização de pacotes

| <b>Categoria</b>                     | <b>Python</b>                                      | <b>R</b>   |
|--------------------------------------|--|--|
| algoritmos de aprendizado de máquina | fastcluster, keybert, pytorch, sklearn, tensorflow | bartmachine, caret, dbarts, e1071, gbm, glmnet, h2oensemble, iml, randomforest, randomforestsrc, ranger, rlearner, rpart, subsemble, superlearner, xgboost |
| cálculo e estatística                | gurobipy, scipy                                    | crs, np  |
| inferência causal                    | mcf  | arco, causaldml, causalimpact, causaltoolbox, grf, hdm, lmt, matching, npcausal, policytree, tmle, twang   |
| manipulação e obtenção de dados      | numpy, pandas, pypdf2                              | normalweatherr, sas7bdat, twitter, worldmet  |
| processamento de linguagem natural   | gensim, nltk, nrclex, textblob                     | corenlp, nlp, tm   |
| visualização de gráficos             | matplotlib   | ggplot2  |

Fonte: Elaborado pelo autor.

Nota: Apesar de haver sobreposição, considerando a finalidade didática, foi realizada a divisão dos pacotes entre inferência causal, processamento de linguagem natural, aprendizado de máquina e cálculo e estatística, sempre considerando a especialidade nessa ordem.

## 6 Considerações finais

Este trabalho pretendeu mapear evidências de utilização de aprendizado de máquina na avaliação de políticas públicas. Dessa forma, a metodologia aplicada foi baseada na revisão de escopo, indicada quando ainda não existe revisão abrangente de um corpo de literatura.

Foram realizadas pesquisas sistemáticas, entre os dias 13 e 15 de março, em seis bases de dados de publicações multidisciplinares e voltadas à tecnologia, a saber: Scopus, Web of Science, ScienceDirect, IEEE Xplore, ACM e ArXiv. Em seguida, os trabalhos passaram por critérios de seleção para que fossem analisados apenas aqueles que tratassem de aplicação de aprendizado de máquina para avaliação de intervenções reais. Ainda que não tenha sido feita restrição quanto ao tipo de avaliação, a delimitação do escopo em aplicações com dados reais de políticas públicas, naturalmente, fez as avaliações de impacto e resultados preponderarem.

Com o objetivo de responder como o aprendizado de máquina tem sido utilizado na avaliação de políticas públicas, foram definidas três subquestões voltadas a contextualizar as características gerais dos estudos, conhecer os objetos de avaliação e descobrir quais métodos são aplicados.

De forma geral, dentro do escopo delimitado de aplicações em dados reais de políticas públicas e das buscas realizadas, verifica-se que são estudos recentes, sendo 96% compreendidos na última década (a partir 2014) e 74% nos últimos 5 anos. Quanto à localidade, se destacam Estados Unidos e China, vez que ao menos um dos autores de 70% dos trabalhos são afiliados a uma instituição desses países. Não foi identificado nenhum autor afiliado à instituição brasileira dentre os estudos selecionados. Portanto, a publicação de estudos com aplicação prática de aprendizado de máquina para avaliação de intervenções no país é uma lacuna identificada e oportunidade de trabalho futuro.

Nem todos os estudos têm o objetivo principal de avaliação da intervenção. Dessa forma, do ponto de vista do método de aprendizado de máquina, os estudos foram agrupados em aplicação, revisão e novo método. Essa categorização visa facilitar a navegação nos estudos, principalmente para o leitor que deseja mais detalhes sobre os métodos, comumente mais detalhados nos estudos de revisão e que apresentam novos métodos.

Quanto aos objetos, são abordadas intervenções diversas associadas a 13 das 28 funções de governo da Portaria SOF/ME n 2.520, de 21 de março de 2022. A maior concentração é na área de saúde e gestão ambiental. Assistência social e previdência social, que são áreas relevantes e de grande materialidade orçamentária, contêm um estudo relacionado apenas. Assim, há lacuna em muitas áreas que podem ser exploradas e

disseminadas, sendo oportunidades de trabalho futuro.

Outra característica dos objetos analisados é a presença de dados não estruturados. Aprendizado de máquina abre a possibilidade de utilizar imagens e documentos de textos com mais facilidade. Exemplos nessa linha incluem a predição de indicador de riqueza e concentração de poluidores no ar, ambos a partir de imagens de satélite.

Com relação a dados não estruturados do tipo texto, e já passando a relatar os diferentes objetivos de aplicação do aprendizado de máquina na avaliação de políticas públicas, uma aplicação comum é a análise de sentimentos a partir de comentários públicos para avaliar a opinião pública sobre uma intervenção. Outra abordagem geralmente utilizada com textos é a descoberta dos termos mais relevantes no conteúdo para subsidiar outras etapas da avaliação.

A análise de sentimentos e descoberta de termos relevantes produzem insumo para outros métodos de análise que não envolvem necessariamente o aprendizado de máquina. Nessa mesma linha, outras aplicações intermediárias incluem a classificação dos conjuntos de dados em diferentes grupos e previsão de variáveis, que são tarefas naturais dos algoritmos de aprendizado de máquina.

Na etapa de preparação dados, uma forma de uso do aprendizado de máquina é a imputação de valores faltantes, do qual um exemplo é o preenchimento de dados ausentes a partir de imagens de satélite. Outra é no balanceamento de covariáveis dos grupos de tratamento e controle.

Na análise dos preditores, os resultados do modelo são empregados para verificar as variáveis independentes mais relevantes para o resultado avaliado. Aqui se deve atentar para aquilo que se quer obter como resposta, pois no caso de métodos tradicionais de aprendizado de máquina os parâmetros obtidos não podem ser interpretados como efeitos de causalidade.

Sobre inferência causal, há uma grande quantidade de estudos voltados a esse tema, incluindo a apresentação de novos métodos que unem econometria e aprendizado de máquina. A otimização de regras de alocação decorre da inferência causal, mais especificamente dos efeitos heterogêneos decorrentes do tratamento da intervenção, com a finalidade de encontrar subpopulações ótimas para o tratamento.

Com o fito de contribuir na disseminação de ferramentas, foram mapeados os pacotes de software utilizados no desenvolvimento das avaliações. Adicionalmente, foram identificados códigos-fonte relacionados que podem ser úteis na apropriação do conhecimento pelos interessados.

Diante do exposto, a questão e subquestões de pesquisa foram respondidas, dando cumprimento ao objetivo proposto. Dito isso, a contribuição da dissertação é fornecer evidências de aprendizado de máquina na avaliação de políticas públicas, incluindo codificações

---

realizadas pelo autor.

Dos métodos aplicados, destaca-se que os de inferência causal trazem vantagens em relação aos métodos econométricos, a exemplo de lidar com dados de alta dimensão, automação e transparência na seleção de modelo. Todavia, a automação não exime o avaliador de aplicar as teorias de avaliação de políticas públicas e da econometria. Considerando que é uma área recente, outra frente de trabalho futuro é uma proposta de trilha de aprendizado que possibilite os interessados internalizarem e aplicarem esses métodos de aprendizado causal na avaliação de políticas públicas. Foram listados estudos de revisão, código-fonte e tecnologias, além de outras referências que podem apoiar esse processo de construção.

Apesar de as buscas terem sido realizadas em seis bases de dados com a finalidade de ser o mais abrangente possível, cabe dizer que a delimitação das fontes de pesquisa é uma limitação desse trabalho. É possível que estudos relacionados ao escopo do trabalho não tenham sido indexados por elas ou, ainda, que as expressões de busca não tenham capturado todos os artigos relacionados. Acrescenta-se que organizações e outros interessados podem desenvolver trabalhos que sejam publicados em outros meios não indexados, como *blogs*, páginas institucionais, entre outros. Portanto, a leitura das análises expostas neste trabalho deve levar essa limitação em consideração.

Outra limitação desse estudo é o fato de novos trabalhos identificados pelo método de bola de neve não terem sido processados. Isso não seria possível dentro do cronograma para realização da dissertação. Contudo, todos esses estudos constam nos apêndices como contribuição para novos trabalhos. Nesse contexto, outro trabalho futuro é a complementação dessa avaliação a partir dos estudos decorrentes da exploração das citações bibliográficas. Um desafio interessante seria o uso dos grandes modelos de linguagem, também conhecidos como modelos de inteligência artificial generativa, para automação das tarefas de seleção e mapeamento dos dados. O sucesso dessa tarefa poderia permitir a automação de novas atualizações e inclusão de novas fontes de publicação ao longo do tempo, algo bem interessante frente a uma área em franca ascensão.





## Referências

ABRELL, J.; KOSCH, M.; RAUSCH, S. How effective is carbon pricing? A machine learning approach to policy evaluation. *Journal of Environmental Economics and Management*, v. 112, 2022. ISSN 00950696 (ISSN). Citado 6 vezes nas páginas 58, 65, 101, 111, 127 e 162.

AGASISTI, T. et al. Financial Education during COVID-19 - Assessing the effectiveness of an online programme in a high school. *Applied Economics*, v. 54, n. 35, p. 4006–4029, 2022. ISSN 00036846 (ISSN). Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85122233377&doi=10.1080%2f00036846.2021.2016586&partnerID=40&md5=78dcbdf242fa1f6e6a09070dec98e8e>>. Citado 5 vezes nas páginas 59, 65, 101, 111 e 127.

ALI, S.; SHIRAZI, F. A Transformer-Based Machine Learning Approach for Sustainable E-Waste Management: A Comparative Policy Analysis between the Swiss and Canadian Systems. *Sustainability (Switzerland)*, v. 14, n. 20, 2022. ISSN 20711050 (ISSN). Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85140791461&doi=10.3390%2fsu142013220&partnerID=40&md5=28cb3450be5468d7b842cd139a7ccb94>>. Citado 8 vezes nas páginas 58, 61, 66, 68, 72, 101, 112 e 128.

AMARASINGHE, K. et al. Explainable Machine Learning for Public Policy: Use Cases, Gaps, and Research Directions. *Data & Policy*, n. 5, 8 2023. Disponível em: <<https://www.cambridge.org/core/journals/data-and-policy/article/explainable-machine-learning-for-public-policy-use-cases-gaps-and-research-directions/B5B66B3C3B16196482984E878D795161>>. Citado na página 155.

ANDINI, M. et al. Machine learning in the service of policy targeting: The case of public credit guarantees. *Journal of Economic Behavior and Organization*, v. 198, p. 434–475, 2022. ISSN 01672681 (ISSN). Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85129291431&doi=10.1016%2fj.jebo.2022.04.004&partnerID=40&md5=18b9a4fb473dc7edb845faa45693e507>>. Citado 2 vezes nas páginas 155 e 156.

ANDINI, M. et al. Targeting with machine learning: An application to a tax rebate program in Italy. *Journal of Economic Behavior and Organization*, v. 156, p. 86–102, 2018. ISSN 01672681 (ISSN). Citado 2 vezes nas páginas 155 e 156.

ARKSEY, H.; O'MALLEY, L. Scoping studies: towards a methodological framework. *The International Journal of Social Research Methodology*, v. 8, p. 19–32, 2005. Disponível em: <<https://doi.org/10.1080/1364557032000119616>>. Citado 5 vezes nas páginas 31, 32, 33, 34 e 35.

ATHEY, S. The Impact of Machine Learning on Economics. In: AGRAWAL, A.; GANS, J.; GOLDFARB, A. (Ed.). *The Economics of Artificial Intelligence: An Agenda*. 1. ed. [S.l.]: University of Chicago Press, 2019. v. 1, cap. 21, p. 507–547. ISBN 9780226613338. Citado 3 vezes nas páginas 55, 68 e 72.

- ATHEY, S.; IMBENS, G. W. The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives*, American Economic Association, v. 31, n. 2, p. 3–32, 3 2017. ISSN 08953309. Disponível em: <<https://pubs.aeaweb.org/doi/pdfplus/10.1257/jep.31.2.3>>. Citado 2 vezes nas páginas 55 e 60.
- ATHEY, S.; IMBENS, G. W. Machine Learning Methods That Economists Should Know About. *Annual Review of Economics*, v. 11, n. 1, p. 685–725, 2019. Disponível em: <<https://doi.org/10.1146/annurev-economics-080217-053433>>. Citado na página 55.
- ATHEY, S.; IMBENS, G. W.; WAGER, S. Approximate Residual Balancing: Debiased Inference of Average Treatment Effects in High Dimensions. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, v. 80, n. 4, p. 597–623, 7 2018. ISSN 1369-7412. Disponível em: <<https://doi.org/10.1111/rssb.12268>>. Citado na página 161.
- ATHEY, S.; WAGER, S. *Policy Learning with Observational Data*. 2020. Citado na página 161.
- AUERBACH, J. et al. Coupling data science with community crowdsourcing for urban renewal policy analysis: An evaluation of Atlantas Anti-Displacement Tax Fund. *Environment and Planning B: Urban Analytics and City Science*, v. 47, n. 6, p. 1081–1097, 2020. ISSN 23998083 (ISSN). Citado 5 vezes nas páginas 59, 65, 101, 112 e 128.
- BAIARDI, A.; NAGHI, A. A. *The Value Added of Machine Learning to Causal Inference: Evidence from Revisited Studies*. 2021. Citado na página 161.
- BALLESTAR, M. T. et al. A novel machine learning approach for evaluation of public policies: An application in relation to the performance of university researchers. *Technological Forecasting and Social Change*, Elsevier Inc., v. 149, 12 2019. ISSN 00401625. Citado 2 vezes nas páginas 28 e 39.
- BALLESTAR, M. T. et al. A novel machine learning approach for evaluation of public policies: An application in relation to the performance of university researchers. *Technological Forecasting and Social Change*, v. 149, p. 119756, 2019. ISSN 0040-1625. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0040162519311217>>. Citado 10 vezes nas páginas 51, 55, 59, 64, 67, 68, 71, 101, 112 e 129.
- BALLESTAR, M. T.; SAINZ, J.; SANZ, I. An economic evaluation of educational interventions in the LOMLOE\*: Proposals for improvement with Artificial Intelligence. *Revista Espanola de Pedagogia*, v. 80, n. 281, p. 133–154, 2022. ISSN 00349461 (ISSN). Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85124830089&doi=10.22550%2fREP80-1-2022-09&partnerID=40&md5=2397878584780130c1f693e828fe6230>>. Citado 5 vezes nas páginas 59, 65, 102, 112 e 129.
- BANDARA, K. et al. Towards Accurate Predictions and Causal What-if Analyses for Planning and Policy-making: A Case Study in Emergency Medical Services Demand. In: *2020 International Joint Conference on Neural Networks (IJCNN)*. [S.l.: s.n.], 2020. p. 1–10. Citado na página 161.
- BEITEL, S. E. Applying Artificial Intelligence Data Mining Tools to the Challenges of Program Evaluation. 2005. Citado na página 156.

BELLONI, A. et al. Program Evaluation and Causal Inference With High-Dimensional Data. *Econometrica*, v. 85, n. 1, p. 233–298, 2017. ISSN 00129682 (ISSN). Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85010868595&doi=10.3982%2fECTA12723&partnerID=40&md5=c8db27b2d5d357f435c88563bf037e81>>. Citado 8 vezes nas páginas 55, 59, 62, 71, 72, 102, 113 e 130.

BELLONI, A.; CHERNOZHUKOV, V.; HANSEN, C. High-Dimensional Methods and Inference on Structural and Treatment Effects. *Journal of Economic Perspectives*, v. 28, n. 2, p. 29–50, 5 2014. Disponível em: <<https://www.aeaweb.org/articles?id=10.1257/jep.28.2.29>>. Citado na página 161.

BERTRAND, M. et al. *Contemporaneous and Post-Program Impacts of a Public Works Program*. World Bank, 2017. Disponível em: <<https://elibrary.worldbank.org/doi/abs/10.1596/28460>>. Citado na página 161.

BI, Q. et al. What is machine learning? A primer for the epidemiologist. *American Journal of Epidemiology*, Oxford University Press, v. 188, n. 12, p. 2222–2239, 12 2019. ISSN 14766256. Citado na página 25.

BLASIO, G. de; DIGNAZIO, A. Regional policy in the machine learning era: Opportunities and challenges. *Scienze Regionali*, Societa Editrice Il Mulino, Banca d'Italia, Via Nazionale 91, Roma, 00184, Italy, v. 20, n. 2, p. 205–220, 2021. ISSN 17203929 (ISSN). Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85109003517&doi=10.14650%2f99817&partnerID=40&md5=cdb2790159818c5dd7c507d23bd71928>>. Citado na página 156.

BLEI, D. M.; NG, A. Y.; JORDAN, M. I. Latent dirichlet allocation. *Journal of machine Learning research*, v. 3, n. Jan, p. 993–1022, 2003. Citado na página 68.

BODORY, H.; BUSSHOFF, H.; LECHNER, M. High Resolution Treatment Effects Estimation: Uncovering Effect Heterogeneities with the Modified Causal Forest. *Entropy*, v. 24, n. 8, 2022. ISSN 1099-4300. Disponível em: <<https://www.mdpi.com/1099-4300/24/8/1039>>. Citado na página 161.

BRAHMA, D.; MUKHERJEE, D. IndiaS Mid-Day Meal Program And Schooling: An Evaluation Based On Machine Learning. *Applied Econometrics and International Development*, v. 18, n. 1, p. 141–152, 2018. Disponível em: <<https://www.usc.gal/econom-et/reviews/aeid18110.pdf>>. Citado 9 vezes nas páginas 27, 40, 44, 45, 59, 65, 102, 113 e 131.

Brasil. *Portaria SOF/ME nº 2.520, de 21 de março de 2022. Divulga a Portaria SOF/SETO/ME nº 42, de 14 de abril de 1999, atualizada, e revoga os atos que menciona*. Brasília, DF: [s.n.], 2023. Disponível em: <<https://www.in.gov.br/en/web/dou/-/portaria-sof/me-n-2.520-de-21-de-marco-de-2022-387374808>>. Citado 2 vezes nas páginas 48 e 58.

BREIMAN, L. Random Forests. *Machine Learning*, v. 45, n. 1, p. 5–32, 2001. ISSN 1573-0565. Disponível em: <<https://doi.org/10.1023/A:1010933404324>>. Citado na página 70.

BRONDEEL, R.; KESTENS, Y.; CHAIX, B. An evaluation of transport mode shift policies on transport-related physical activity through simulations

based on random forests. *International Journal of Behavioral Nutrition and Physical Activity*, v. 14, n. 1, 2017. ISSN 14795868 (ISSN). Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85031922752&doi=10.1186%2f12966-017-0600-1&partnerID=40&md5=735323968df1b0586a044f11bd781125>>. Citado 5 vezes nas páginas 59, 65, 102, 113 e 131.

BUHL-WIGGERS, J. et al. Some children left behind: Variation in the effects of an educational intervention. *Journal of Econometrics*, 2022. ISSN 0304-4076. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0304407622000355>>. Citado na página 161.

BÜHLMANN, P.; GEER, S. van de. *Statistics for High-Dimensional Data: Methods, Theory and Applications*. 1. ed. Springer Berlin, Heidelberg, 2011. Disponível em: <<http://www.springer.com/series/692>>. Citado na página 60.

BURLIG, F. et al. Machine Learning from Schools about Energy Efficiency. *Journal of the Association of Environmental and Resource Economists*, v. 7, n. 6, p. 1181–1217, 2020. Disponível em: <<https://doi.org/10.1086/710606>>. Citado na página 161.

CARTER, M. R.; TJERNSTROM, E.; TOLEDO, P. Heterogeneous impact dynamics of a rural business development program in Nicaragua. *JOURNAL OF DEVELOPMENT ECONOMICS*, v. 138, p. 77–98, 2019. ISSN ["0304-3878", "1872-6089"]. Citado 7 vezes nas páginas 51, 59, 63, 71, 103, 114 e 131.

CHAMPAGNE, F. et al. A avaliação no campo da saúde: conceitos e métodos. In: *Avaliação Conceitos e Métodos*. [S.l.: s.n.], 2011. p. 41–60. Citado na página 26.

CHANG, W. Decision-Making Power for Women and Girls: Evaluating Interventions in Sexual and Reproductive Health in Sub-Saharan Africa. 2020. Citado 7 vezes nas páginas 58, 63, 72, 103, 114, 132 e 163.

CHEN, H. et al. Heterogeneous effects of health insurance on rural childrens health in China: A causal machine learning approach. *International Journal of Environmental Research and Public Health*, v. 18, n. 18, 2021. ISSN 16617827 (ISSN). Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85114740701&doi=10.3390%2fijerph18189616&partnerID=40&md5=1e9182ec21b691ef0ad10287920ecec2>>. Citado 9 vezes nas páginas 58, 63, 70, 71, 103, 114, 132, 161 e 163.

CHEN, J.-e.; HSIANG, C.-W. Causal Random Forests Model Using Instrumental Variable Quantile Regression. *Econometrics*, v. 7, n. 4, 2019. ISSN 2225-1146. Disponível em: <<https://www.mdpi.com/2225-1146/7/4/49>>. Citado na página 161.

CHEN, J.-e.; HUANG, C.-H.; TIEN, J.-J. Debiased/Double Machine Learning for Instrumental Variable Quantile Regressions. *Econometrics*, v. 9, n. 2, 2021. ISSN 2225-1146. Disponível em: <<https://www.mdpi.com/2225-1146/9/2/15>>. Citado na página 161.

CHEN, R.; ZHOU, J. Fare adjustment's impacts on travel patterns and farebox revenue: An empirical study based on longitudinal smartcard data. *Transportation Research Part A: Policy and Practice*, v. 164, p. 111–133, 2022. ISSN 09658564 (ISSN). Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85135985015&doi=10.1016%2fj.tara.2022.08.003&partnerID=40&md5=a9e967793f3653574bda3c5ac8ca598b>>. Citado 8 vezes nas páginas 59, 64, 68, 103, 114, 133, 162 e 163.

- CHEN, T.-S. Evaluating Conditional Cash Transfer Policies with Machine Learning Methods. *ArXiv*, 2018. Disponível em: <<http://arxiv.org/pdf/1803.06401v1>>. Citado na página 156.
- CHERNOZHUKOV, V. et al. Double/Debiased/Neyman Machine Learning of Treatment Effects. 2017. Citado na página 161.
- CHERNOZHUKOV, V. et al. Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, v. 21, n. 1, p. C1–C68, 7 2018. ISSN 1368-4221. Disponível em: <<https://doi.org/10.1111/ectj.12097>>. Citado na página 161.
- CHERNOZHUKOV, V. et al. *Fischer-Schultz Lecture: Generic Machine Learning Inference on Heterogenous Treatment Effects in Randomized Experiments, with an Application to Immunization in India*. 2023. Citado na página 161.
- CHERNOZHUKOV, V.; NEWHEY, W. K.; SINGH, R. *Automatic Debiased Machine Learning of Causal and Structural Effects*. 2022. Citado na página 161.
- CHERNOZHUKOV, V.; NEWHEY, W. K.; SINGH, R. Debiased machine learning of global and local parameters using regularized Riesz representer. *The Econometrics Journal*, v. 25, n. 3, p. 576–601, 7 2022. ISSN 1368-4221. Disponível em: <<https://doi.org/10.1093/ectj/utac002>>. Citado na página 161.
- CHEW, R. F. et al. Assessing target audiences of digital public health campaigns: A computational approach. In: *International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction conference and Behavior Representation in Modeling and Simulation, SBP-BRiMS*. ["RTI International, Research Triangle Park, Durham, NC, United States", "Northwestern University, Evanston, IL, United States", "Imangi Studios, Raleigh, NC, United States"]: Springer Verlag, 2018. v. 10899, p. 286–291. ISBN 03029743 (ISSN); 9783319933719 (ISBN). Disponível em: <[https://www.scopus.com/inward/record.uri?eid=2-s2.0-85049770128&doi=10.1007%2f978-3-319-93372-6\\_32&partnerID=40&md5=e2b91ecbbb1a4010ee95d9f9a21a2109](https://www.scopus.com/inward/record.uri?eid=2-s2.0-85049770128&doi=10.1007%2f978-3-319-93372-6_32&partnerID=40&md5=e2b91ecbbb1a4010ee95d9f9a21a2109)>. Citado 7 vezes nas páginas 58, 60, 61, 64, 103, 115 e 133.
- CHIN, S.; KAHN, M. E.; MOON, H. R. Estimating the Gains from New Rail Transit Investment: A Machine Learning Tree Approach. *Real Estate Economics*, v. 48, n. 3, p. 886–914, 2020. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1111/1540-6229.12249>>. Citado na página 161.
- CHRISTENSEN, P. et al. Decomposing the Wedge between Projected and Realized Returns in Energy Efficiency Programs. *The Review of Economics and Statistics*, v. 105, n. 4, p. 798–817, 7 2023. ISSN 0034-6535. Disponível em: <[https://doi.org/10.1162/rest\\_a\\_01087](https://doi.org/10.1162/rest_a_01087)>. Citado na página 162.
- COCKX, B.; LECHNER, M.; BOLLENS, J. Priority to unemployed immigrants? A causal machine learning evaluation of training in Belgium. *Labour Economics*, v. 80, 2023. ISSN 09275371 (ISSN). Citado 13 vezes nas páginas 51, 59, 63, 67, 68, 70, 71, 103, 115, 134, 161, 162 e 163.
- COELHO, T. P. et al. Comparação e análise do uso de revisão sistemática e revisão de escopo na área do cuidado ao paciente na Farmácia. *Research, Society and Development*, Research, Society and Development, v. 10, n. 12, p. e08101219915, 9 2021. Disponível em: <<https://rsdjournal.org/index.php/rsd/article/view/19915/17912>>. Citado na página 31.

COLANGELO, K. *Essays in Econometrics and Labor Economics*. Tese (Doutorado) — University of California, Irvine, 2021. Disponível em: <<https://escholarship.org/uc/item/3940b3fs>>. Citado 13 vezes nas páginas 55, 57, 59, 62, 64, 70, 71, 72, 73, 103, 115, 135 e 161.

COLANGELO, K.; LEE, Y.-Y. Double Debiased Machine Learning Nonparametric Inference with Continuous Treatments. *ArXiv*, 2020. Disponível em: <<http://arxiv.org/pdf/2004.03036v7>>. Citado 14 vezes nas páginas 54, 57, 59, 62, 70, 71, 72, 73, 104, 116, 135, 161, 162 e 164.

CORIA, S.; JUKIC, T.; DECMAN, M. Open data for human development in Mexico: A data science perspective. *Proceedings of the European Conference on e-Government, ECEG*, v. 2016, p. 38–44, 2016. ISSN 20491034 (ISSN); 9781910810989 (ISBN). Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-84979518618&partnerID=40&md5=5e07a664ed5fb47c241dec7f85fa0a3c>>. Citado na página 156.

CUNHA, C. G. S. d. Avaliação de Políticas Públicas e Programas Governamentais: tendências recentes e experiências no Brasil. *Revista Estudos de Planejamento*, v. 12, p. 27–57, 12 2018. Citado na página 26.

DAI, X. et al. Has the Three-Year Action Plan improved the air quality in the Fenwei Plain of China? Assessment based on a machine learning technique. *Atmospheric Environment*, v. 286, 2022. ISSN 13522310 (ISSN). Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85132781501&doi=10.1016%2fj.atmosenv.2022.119204&partnerID=40&md5=ff62011029b84377bac9ff32082f491d>>. Citado 10 vezes nas páginas 57, 58, 69, 72, 104, 116, 136, 162, 163 e 164.

DAOUD, A.; JOHANSSON, F. *Estimating Treatment Heterogeneity of International Monetary Fund Programs on Child Poverty with Generalized Random Forest*. [S.l.], 2019. Disponível em: <<https://EconPapers.repec.org/RePEc:osf:socarx:awfjt>>. Citado na página 162.

DAVIS, J. M. V.; HELLER, S. B. Using Causal Forests to Predict Treatment Heterogeneity: An Application to Summer Jobs. *American Economic Review*, v. 107, n. 5, p. 546–550, 5 2017. Disponível em: <<https://www.aeaweb.org/articles?id=10.1257/aer.p20171000>>. Citado na página 162.

DAVIS, J. M. V.; HELLER, S. B. Rethinking the Benefits of Youth Employment Programs: The Heterogeneous Effects of Summer Jobs. *The Review of Economics and Statistics*, v. 102, n. 4, p. 664–677, 7 2020. ISSN 0034-6535. Disponível em: <[https://doi.org/10.1162/rest\\_a\\_00850](https://doi.org/10.1162/rest_a_00850)>. Citado na página 162.

DENTE, B.; SUBIRATS, J. *Decisiones públicas: Análisis y estudio de los procesos de decisión en políticas públicas*. [S.l.: s.n.], 2014. 1–368 p. ISBN 978-8434409965. Citado na página 26.

DU, Y. et al. Leveraging Computational Tools to Inform Public Policy Process for Social Good. 2022. Citado na página 160.

DUEÑAS, M. et al. *Assessing the Impact of COVID-19 on Trade: a Machine Learning Counterfactual Analysis*. 2021. Citado na página 162.

DUTHIL, B.; IMOUSSATEN, A.; MONTMAIN, J. A text-mining and possibility theory based model using public reports to highlight the sustainable development strategy of a city. *IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications*, Institute of Electrical and Electronics Engineers Inc., ["Laboratoire L3I, Université de la Rochelle, Avenue Michel Crépeau, La-Rochelle Cedex 01, 17042, France", "Centre de Recherche LGI2P, École des Mines d'Alès, Site ERIEE, Parc Scientifique G. Besse, Nîmes Cedex 01, 30035, France"], p. 36–41, 2017. ISSN 9781509042524 (ISBN). Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85028511890&doi=10.1109%2fCIVEMSA.2017.7995298&partnerID=40&md5=1b96c997c91670ca8c5b7549a65320d5>>. Citado 6 vezes nas páginas 59, 66, 68, 104, 116 e 136.

ESTRADA, M. A. R. Policy modeling: Definition, classification and evaluation. *Journal of Policy Modeling*, v. 33, n. 4, p. 523–536, 2011. ISSN 0161-8938. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0161893811000196>>. Citado na página 66.

FARRELL, M. H. Robust inference on average treatment effects with possibly more covariates than observations. *Journal of Econometrics*, v. 189, n. 1, p. 1–23, 2015. ISSN 0304-4076. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0304407615001864>>. Citado na página 162.

FRANCKE, P.; ACOSTA, G. Impact of the qali warma school feeding program on chronic child malnutrition. *Apuntes*, v. 48, n. 88, p. 141–178, 2021. ISSN 02521865 (ISSN). Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85109892637&doi=10.21678%2fapuntes.88.1228&partnerID=40&md5=a6ae8ce4dcc5dc43fe3a9b60aabb2070>>. Citado 4 vezes nas páginas 59, 104, 116 e 136.

FURTADO, B. A.; ANDREÃO, G. O. Machine Learning Simulates Agent-Based Model Towards Policy. *ArXiv*, 2022. Disponível em: <<http://arxiv.org/pdf/2203.02576v2>>. Citado na página 157.

GALVÃO, M. C. B.; RICARTE, I. L. M. Revisão sistemática da literatura: conceituação, produção e publicação. *Logeion: Filosofia da Informação*, v. 6, n. 1, p. 57–73, 9 2019. Disponível em: <<https://revista.ibict.br/fiinf/article/view/4835>>. Citado na página 33.

GAO, Y.; LI, Y. y.; WANG, Y. Modular Policy Evaluation System: A Policy Evaluation Framework Based on Text Mining. In: *6th IEEE International Conference on Big Data Analytics, ICBDA 2021*. Institute of Electrical and Electronics Engineers Inc., 2021. p. 204–209. ISBN 9780738131672 (ISBN). Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85105250444&doi=10.1109%2fICBDA51983.2021.9403142&partnerID=40&md5=aa37cb14397cdabde39c850ec0f559c5>>. Citado 9 vezes nas páginas 59, 61, 64, 66, 67, 68, 104, 116 e 137.

GERTLER, P. J. et al. *Avaliação de Impacto na Prática, segunda edição*. Washington DC: World Bank Publications, 2018. Disponível em: <<https://publications.iadb.org/pt/avaliacao-de-impacto-na-pratica-segunda-edicao>>. Citado na página 68.

GOLLER, D. et al. Does the estimation of the propensity score by machine learning improve matching estimation? The case of Germany's programmes for long term

unemployed. *Labour Economics*, v. 65, 2020. ISSN 09275371 (ISSN). Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85086147183&doi=10.1016%2fj.labecon.2020.101855&partnerID=40&md5=6b1b3df56f70141e6e324e03dfae7f4b>>. Citado 8 vezes nas páginas 55, 59, 63, 104, 117, 137, 162 e 163.

GOLLER, D. et al. Does the estimation of the propensity score by machine learning improve matching estimation? The case of Germany's programmes for long term unemployed. *Labour Economics*, v. 65, p. 101855, 2020. ISSN 0927-5371. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0927537120300592>>. Citado na página 162.

GORI, A. et al. Characterizing urbanization impacts on floodplain through integrated land use, hydrologic, and hydraulic modeling. *Journal of Hydrology*, v. 568, p. 82–95, 2019. ISSN 0022-1694. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0022169418308242>>. Citado 7 vezes nas páginas 51, 58, 61, 65, 104, 117 e 138.

GRANGE, S. K.; CARSLAW, D. C. Using meteorological normalisation to detect interventions in air quality time series. *Science of The Total Environment*, v. 653, p. 578–588, 2019. ISSN 0048-9697. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S004896971834244X>>. Citado na página 162.

GRECOV, P. et al. Causal Inference Using Global Forecasting Models for Counterfactual Prediction. In: CHENG, H. K. K. et al. (Ed.). *Advances in Knowledge Discovery and Data Mining*. Cham: Springer International Publishing, 2021. p. 282–294. ISBN 978-3-030-75765-6. Citado na página 162.

GRECOV, P. et al. Probabilistic Causal Effect Estimation With Global Neural Network Forecasting Models. *IEEE Transactions on Neural Networks and Learning Systems*, p. 1–15, 2022. ISSN 2162-2388 VO -. Citado 12 vezes nas páginas 55, 59, 62, 71, 72, 104, 117, 138, 161, 162, 163 e 164.

GUO, Y. et al. Evaluating the real changes of air quality due to clean air actions using a machine learning technique: Results from 12 Chinese mega-cities during 20132020. *Chemosphere*, v. 300, 2022. ISSN 00456535 (ISSN). Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85128580727&doi=10.1016%2fj.chemosphere.2022.134608&partnerID=40&md5=1d15abed1b10e987739ff1a6a89119b1>>. Citado 6 vezes nas páginas 57, 58, 104, 117, 139 e 164.

HA, S.; GRUBERT, E. Hybridizing qualitative coding with natural language processing and deep learning to assess public comments: A case study of the clean power plan. *Energy Research & Social Science*, v. 98, p. 103016, 2023. ISSN 2214-6296. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2214629623000762>>. Citado 7 vezes nas páginas 51, 59, 61, 65, 105, 117 e 139.

HALVORSEN, A. et al. Reducing Subway Crowding: Analysis of an Off-Peak Discount Experiment in Hong Kong. *Transportation Research Record*, v. 2544, n. 1, p. 38–46, 2016. Disponível em: <<https://doi.org/10.3141/2544-05>>. Citado na página 162.

HARDING, M. C.; LAMARCHE, C. Small steps with big data: Using machine learning in energy and environmental economics. *Annual Review of Resource Economics*, v. 13, p. 469–488, 2021. ISSN 19411340 (ISSN). Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85117145366&doi=10.1146%2fannurev>>



-resource-100920-034117&partnerID=40&md5=c51dad35663dee389acfa915bca1971c>. Citado 2 vezes nas páginas 55 e 60.

HATAMI, F. et al. Non-Linear Associations Between the Urban Built Environment and Commuting Modal Split: A Random Forest Approach and SHAP Evaluation. *IEEE Access*, v. 11, p. 12648–12661, 2023. ISSN 21693536 (ISSN). Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85148430431&doi=10.1109%2fACCESS.2023.3241627&partnerID=40&md5=48ca0d79cb0fd573a6bf99a389c3377f>>. Citado na página 157.

HATAMYAR, J.; KREIF, N. Policy Learning with Rare Outcomes. *ArXiv*, 2023. Disponível em: <<http://arxiv.org/pdf/2302.05260v1>>. Citado 10 vezes nas páginas 51, 55, 57, 58, 63, 70, 71, 105, 118 e 139.

HEILER, P.; KNAUS, M. C. Effect or Treatment Heterogeneity? Policy Evaluation with Aggregated and Disaggregated Treatments. *ArXiv*, 2021. Disponível em: <<http://arxiv.org/pdf/2110.01427v2>>. Citado 10 vezes nas páginas 54, 57, 59, 63, 71, 73, 105, 118, 140 e 161.

HILL, J.; WEISS, C.; ZHAI, F. Challenges With Propensity Score Strategies in a High-Dimensional Setting and a Potential Alternative. *Multivariate Behavioral Research*, Routledge, v. 46, n. 3, p. 477–513, 2011. Disponível em: <<https://doi.org/10.1080/00273171.2011.570161>>. Citado na página 162.

HILL, J. L. Bayesian Nonparametric Modeling for Causal Inference. *Journal of Computational and Graphical Statistics*, Taylor & Francis, v. 20, n. 1, p. 217–240, 2011. Disponível em: <<https://doi.org/10.1198/jcgs.2010.08162>>. Citado na página 162.

HOFFMAN, I. L. Essays in Public Economics. 2018. Citado na página 157.

HOZOURI, N. T.; DONG, Y.; FREEDMAN, M. Advancing National Health Through Investigations of Nutrition, Medical Expenditure, and Public Insurance Programs. University of California, Irvine, 2020. Citado na página 158.

HSU, Y.-C. et al. *Testing Monotonicity of Mean Potential Outcomes in a Continuous Treatment with High-Dimensional Data*. 2022. Citado na página 162.

HU, Y.-J.; HUANG, S.-W. Challenges of automated machine learning on causal impact analytics for policy evaluation. In: *2nd International Conference on Telecommunication and Networks, TEL-NET*. Institute of Electrical and Electronics Engineers Inc., 2018. v. 2018, p. 1–6. ISBN 9781509067107 (ISBN). Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85049112944&doi=10.1109%2fTEL-NET.2017.8343571&partnerID=40&md5=ba11923d06b2abde2785905e33d1a705>>. Citado 6 vezes nas páginas 55, 59, 72, 105, 118 e 140.

HUANG, L. Y.; HSIANG, S.; GONZALEZ-NAVARRO, M. Using Satellite Imagery and Deep Learning to Evaluate the Impact of Anti-Poverty Programs. *ArXiv*, 2021. Disponível em: <<http://arxiv.org/pdf/2104.11772v1>>. Citado na página 158.

HUBER, M. An introduction to flexible methods for policy evaluation. In: *Handb. of Research Methods and Applications in Empirical Microeconomics*. University of Fribourg, Switzerland: Edward Elgar Publishing Ltd., 2021. p. 82–111. ISBN 9781788976480 (ISBN); 9781788976473 (ISBN). Disponível em: <<https://www.scopus.com/inward/record.uri?e>

id=2-s2.0-85129817378&partnerID=40&md5=4263386bd96c800907621aaf8e164db6>. Citado na página 55.

HUBER, M. et al. Direct and indirect effects of continuous treatments based on generalized propensity score weighting. *Journal of Applied Econometrics*, v. 35, n. 7, p. 814–840, 2020. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/jae.2765>>. Citado na página 162.

IMAI, K.; RATKOVIC, M. Estimating treatment effect heterogeneity in randomized program evaluation. *The Annals of Applied Statistics*, Institute of Mathematical Statistics, v. 7, n. 1, p. 443–470, 2013. Disponível em: <<https://doi.org/10.1214/12-AOAS593>>. Citado na página 162.

JANIESCH, C.; ZSCHECH, P.; HEINRICH, K. Machine learning and deep learning. *Electronic Markets*, v. 31, n. 3, p. 685–695, 9 2021. Disponível em: <<https://doi.org/10.1007/s12525-021-00475-2>>. Citado 2 vezes nas páginas 25 e 26.

JARVIS, S.; DESCHENES, O.; JHA, A. The Private and External Costs of Germanys Nuclear Phase-Out. *Journal of the European Economic Association*, v. 20, n. 3, p. 1311–1346, 7 2022. ISSN 1542-4766. Disponível em: <<https://doi.org/10.1093/jeea/jvac007>>. Citado na página 162.

KALLUS, N. Treatment Effect Risk: Bounds and Inference - 2022 ACM Conference on Fairness, Accountability, and Transparency. p. 213, 2022. Disponível em: <<https://doi.org/10.1145/3531146.3533087>>. Citado na página 158.

KALLUS, N.; MAO, X.; UEHARA, M. *Localized Debiased Machine Learning: Efficient Inference on Quantile Treatment Effects and Beyond*. 2022. Citado na página 162.

KALLUS, N.; ZHOU, A. Assessing Disparate Impacts of Personalized Interventions: Identifiability and Bounds. In: *Conference on Neural Information Processing Systems*. [s.n.], 2019. Disponível em: <<http://arxiv.org/pdf/1906.01552v1>>. Citado 12 vezes nas páginas 51, 55, 59, 63, 71, 72, 73, 105, 118, 140, 162 e 163.

KLEINBERG, J. et al. Prediction policy problems. *American Economic Review*, American Economic Association, v. 105, n. 5, p. 491–495, 5 2015. ISSN 19447981. Citado na página 68.

KNAUS, M. C. Double machine learning-based programme evaluation under unconfoundedness. *The Econometrics Journal*, v. 25, n. 3, p. 602–627, 2022. ISSN ["1368-4221", "1368-423X"]. Citado 13 vezes nas páginas 54, 59, 63, 64, 70, 73, 105, 118, 141, 161, 162, 163 e 164.

KNAUS, M. C.; LECHNER, M.; STRITTMATTER, A. Machine learning estimation of heterogeneous causal effects: Empirical Monte Carlo evidence. *The Econometrics Journal*, Oxford University Press (OUP), v. 24, n. 1, p. 134–161, 6 2020. Disponível em: <<https://doi.org/10.1093%2Fectj%2Futaa014>>. Citado na página 162.

KNAUS, M. C.; LECHNER, M.; STRITTMATTER, A. Heterogeneous Employment Effects of Job Search Programs. *Journal of Human Resources*, University of Wisconsin Press, v. 57, n. 2, p. 597–636, 2022. ISSN 0022-166X. Disponível em: <<https://jhr.uwpress.org/content/57/2/597>>. Citado na página 163.

KOMOROWSKI, J. et al. Rough sets: A tutorial. *Rough fuzzy hybridization: A new trend in decision-making*, Springer Singapore, p. 3–98, 1999. Citado na página 159.

KREIF, N.; DIAZORDAZ, K. Machine learning in policy evaluation: new tools for causal inference. *ArXiv*, 2019. Disponível em: <<http://arxiv.org/pdf/1903.00402v1>>. Citado 12 vezes nas páginas 51, 55, 57, 58, 62, 69, 70, 73, 105, 118, 142 e 164.

KREIF, N. et al. Estimating heterogeneous policy impacts using causal machine learning: a case study of health insurance reform in Indonesia. *Health Services and Outcomes Research Methodology*, Springer, v. 22, n. 2, p. 192–227, 6 2022. ISSN 15729400. Citado 2 vezes nas páginas 29 e 39.

KREIF, N. et al. Estimating heterogeneous policy impacts using causal machine learning: a case study of health insurance reform in Indonesia. *Health Services and Outcomes Research Methodology*, v. 22, n. 2, p. 192–227, 2022. ISSN 13873741 (ISSN). Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85118618112&doi=10.1007%2fs10742-021-00259-3&partnerID=40&md5=9ad65fc8cf624133290d408f68881510>>. Citado 7 vezes nas páginas 57, 58, 63, 71, 105, 118 e 142.

KREIF, N. et al. *Who benefits from health insurance? Uncovering heterogeneous policy impacts using causal machine learning*. York, UK, 2020. (CHE Research Paper). Disponível em: <<https://eprints.whiterose.ac.uk/166393/>>. Citado na página 163.

KUBE, A.; DAS, S.; FOWLER, P. J. Allocating Interventions Based on Predicted Outcomes: A Case Study on Homelessness Services. *Proceedings of the AAAI Conference on Artificial Intelligence*, v. 33, n. 01, p. 622–629, 7 2019. Disponível em: <<https://ojs.aaai.org/index.php/AAAI/article/view/3838>>. Citado na página 163.

KUMAR, A.; SHARMA, A. SOOP: A swarm-optimized opinion prediction model for S-health governance. *Recent Patents on Computer Science*, v. 12, n. 4, p. 280–292, 2019. ISSN 22132759 (ISSN). Citado na página 45.

KUMAR, A.; SHARMA, A. Socio-Sentic framework for sustainable agricultural governance. *Sustainable Computing: Informatics and Systems*, v. 28, 2020. ISSN 22105379 (ISSN). Citado 7 vezes nas páginas 59, 61, 65, 69, 106, 118 e 142.

KUMAR, A.; SHARMA, A. Decision making using machine learning based opinion prediction model for smart governance. *Recent Advances in Computer Science and Communications*, v. 14, n. 5, p. 1402–1411, 2021. ISSN 26662558 (ISSN). Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85114682438&doi=10.2174%2f213275912666191026123414&partnerID=40&md5=ab7ccacad83607baac3f5a11c33c8b80>>. Citado na página 45.

LECHNER, M. *Modified Causal Forests for Estimating Heterogeneous Causal Effects*. 2019. Citado na página 163.

LECHNER, M.; MARECKOVA, J. *Modified Causal Forest*. 2022. Citado na página 163.

LEVAC, D.; COLQUHOUN, H.; O'BRIEN, K. K. Scoping studies: advancing the methodology. *Implementation Science*, v. 5, n. 1, p. 69, 9 2010. ISSN 1748-5908. Disponível em: <<https://doi.org/10.1186/1748-5908-5-69>>. Citado 4 vezes nas páginas 32, 33, 34 e 35.

LI, Y. et al. Quantitative Evaluation of Chinas Pork Industry Policy: A PMC Index Model Approach. *Agriculture*, v. 11, n. 2, 2021. ISSN 20770472 (ISSN). Disponível em: <<https://www.mdpi.com/2077-0472/11/2/86>>. Citado 7 vezes nas páginas 59, 61, 66, 68, 106, 119 e 143.

LINDEN, A.; YARNOLD, P. R. Combining machine learning and matching techniques to improve causal inference in program evaluation. *Journal of Evaluation in Clinical Practice*, v. 22, n. 6, p. 864–870, 2016. ISSN 13561294 (ISSN). Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-84978532216&doi=10.1111%2fjep.12592&partnerID=40&md5=4e87c47951eb896c674ab58b738a9d6f>>. Citado 10 vezes nas páginas 27, 39, 58, 62, 63, 71, 106, 119, 143 e 163.

LINDEN, A.; YARNOLD, P. R. Using data mining techniques to characterize participation in observational studies. *Journal of Evaluation in Clinical Practice*, John Wiley & Sons, Ltd, v. 22, n. 6, p. 839–847, 12 2016. ISSN 1356-1294. Disponível em: <<https://doi.org/10.1111/jep.12515>>. Citado na página 163.

LINDEN, A.; YARNOLD, P. R. Using machine learning to assess covariate balance in matching studies. *Journal of Evaluation in Clinical Practice*, v. 22, n. 6, p. 848–854, 2016. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1111/jep.12538>>. Citado na página 163.

LINDEN, A.; YARNOLD, P. R. Using machine learning to identify structural breaks in single-group interrupted time series designs. *Journal of Evaluation in Clinical Practice*, v. 22, n. 6, p. 855–859, 2016. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1111/jep.12544>>. Citado na página 163.

LINDEN, A.; YARNOLD, P. R. Using machine learning to evaluate treatment effects in multiple-group interrupted time series analysis. *Journal of Evaluation in Clinical Practice*, v. 24, n. 4, p. 740–744, 2018. ISSN 13561294 (ISSN). Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85050814914&doi=10.1111%2fjep.12966&partnerID=40&md5=e067acb6aa0cb069006913c51dfc1ade>>. Citado 8 vezes nas páginas 58, 62, 63, 71, 106, 119, 143 e 163.

LÓPEZ-ANDREU, F. J. et al. Sentinel-2 images and machine learning as tool for monitoring of the common agricultural policy: Calasparra rice as a case study. *Agronomy*, v. 11, n. 4, 2021. ISSN 20734395 (ISSN). Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85108796238&doi=10.3390%2fagronomy11040621&partnerID=40&md5=72250c013df6cedcc3842e4243b79d1e>>. Citado na página 159.

LOUIZOS, C. et al. Causal Effect Inference with Deep Latent-Variable Models. In: GUYON, I. et al. (Ed.). *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2017. v. 30. Disponível em: <[https://proceedings.neurips.cc/paper\\_files/paper/2017/file/94b5bde6de888ddf9cde6748ad2523d1-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2017/file/94b5bde6de888ddf9cde6748ad2523d1-Paper.pdf)>. Citado na página 163.

LOVRIĆ, M. et al. Understanding the true effects of the COVID-19 lockdown on air pollution by means of machine learning. *Environmental Pollution*, v. 274, p. 115900, 2021. ISSN 0269-7491. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0269749120365891>>. Citado na página 163.

LU, C. et al. The national nature reserves in China: Are they effective in conserving mangroves? *Ecological Indicators*, v. 142, 2022. ISSN ["1470-160X", "1872-7034"]. Citado 6 vezes nas páginas 58, 61, 64, 106, 119 e 144.

MA, C.; CHEN, Y.; LEI, Z. Evaluation of the Effectiveness of Collusion Control Policy Implementation by BP Neural Network Based on Annealing Algorithm Optimization. *Mathematical Problems in Engineering*, v. 2022, 2022. ISSN 1024123X (ISSN). Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85128178600&doi=10.1155%2f2022%2f9238838&partnerID=40&md5=8346f05294b4baed5bc85f06905f3b79>>. Citado 4 vezes nas páginas 59, 106, 119 e 144.

MA, L.; ZHANG, Y. Hierarchical social network analysis using multi-agent systems: A school system case. In: *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. [S.l.: s.n.], 2014. p. 1412–1419. ISBN 1062-922X VO -. Citado 8 vezes nas páginas 51, 59, 64, 67, 68, 106, 119 e 144.

MA, Z. et al. Behavioral response to promotion-based public transport demand management: Longitudinal analysis and implications for optimal promotion design. *Transportation Research Part A: Policy and Practice*, v. 141, p. 356–372, 2020. ISSN 0965-8564. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0965856420307382>>. Citado na página 163.

MARTINS, S. S. et al. Prescription Opioid Laws and Opioid Dispensing in US Counties Identifying Salient Law Provisions With Machine Learning. *EPIDEMIOLOGY*, v. 32, n. 6, p. 868–876, 2021. ISSN ["1044-3983", "1531-5487"]. Citado 5 vezes nas páginas 58, 65, 107, 121 e 148.

MARTIRE, R. L. *Sickness Absence Among Patients With Chronic Pain in Swedish Specialist Healthcare*. Tese (Doutorado), 2022. Citado na página 158.

MATENGA, Z. Assessment of energy market's progress towards achieving Sustainable Development Goal 7: A clustering approach. *Sustainable Energy Technologies and Assessments*, v. 52, 2022. ISSN 22131388 (ISSN). Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85129527324&doi=10.1016%2fj.seta.2022.102224&partnerID=40&md5=05eec2340f1db2398ab7dd4e65454fb7>>. Citado 6 vezes nas páginas 59, 64, 68, 106, 119 e 145.

MCCAFFREY, D. F.; RIDGEWAY, G.; MORRAL, A. R. Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychological methods*, American Psychological Association, v. 9, n. 4, p. 403, 2004. Citado 2 vezes nas páginas 63 e 163.

MCFOWLAND, E.; SOMANCHI, S.; NEILL, D. B. Efficient Discovery of Heterogeneous Quantile Treatment Effects in Randomized Experiments via Anomalous Pattern Detection. *ArXiv*, 2023. Disponível em: <<http://arxiv.org/pdf/1803.09159v2https://arxiv.org/abs/1803.09159>>. Citado 9 vezes nas páginas 51, 54, 59, 63, 71, 107, 120, 145 e 162.

MILLER, S. Causal forest estimation of heterogeneous and time-varying environmental policy effects. *Journal of Environmental Economics and Management*, v. 103, p. 102337, 2020. ISSN 0095-0696. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0095069620300607>>. Citado na página 163.

MUELLER, A. *WordCloud for Python*. 2020. Disponível em: <[https://amueller.github.io/word\\_cloud/](https://amueller.github.io/word_cloud/)>. Citado 2 vezes nas páginas 56 e 57.

MULLAINATHAN, S.; SPIESS, J. Machine Learning: An Applied Econometric Approach. *Journal of Economic Perspectives*, v. 31, n. 2, p. 87–106, 5 2017. Disponível em: <<https://www.aeaweb.org/articles?id=10.1257/jep.31.2.87>>. Citado 3 vezes nas páginas 60, 61 e 163.

MULLALLY, C.; CHAKRAVARTY, S. Are matching funds for smallholder irrigation money well spent? *Food Policy*, v. 76, p. 70–80, 2018. ISSN 0306-9192. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0306919217305729>>. Citado na página 163.

MUNN, Z. et al. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology*, BioMed Central, v. 18, n. 1, 11 2018. ISSN 14712288. Disponível em: <<https://doi.org/10.1186/s12874-018-0611-x>>. Citado na página 31.

NIJKAMP, P.; BURCH, M. van der; VINDIGNI, G. A comparative institutional evaluation of public-private partnerships in Dutch urban land-use and revitalisation projects. *Urban Studies*, v. 39, n. 10, p. 1865–1880, 2002. ISSN 00420980 (ISSN). Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-0036059314&doi=10.1080%2f0042098022000002993&partnerID=40&md5=0372d0065ace2fea18c95e8438f942cf>>. Citado na página 159.

O'NEILL, E.; WEEKS, M. *Causal Tree Estimation of Heterogeneous Household Response to Time-Of-Use Electricity Pricing Schemes*. 2019. Citado na página 163.

OPRESCU, A. M. et al. Artificial intelligence in pregnancy: A scoping review. *IEEE Access*, Institute of Electrical and Electronics Engineers Inc., v. 8, p. 181450–181484, 2020. ISSN 21693536. Citado na página 35.

OUZZANI, M. et al. Rayyana web and mobile app for systematic reviews. *Systematic Reviews*, v. 5, n. 1, p. 210, 2016. ISSN 2046-4053. Disponível em: <<http://dx.doi.org/10.1186/s13643-016-0384-4>>. Citado na página 44.

PAL, L. A. Policy Analysis: Concepts and Practice. In: *Beyond policy analysis: Public issue management in turbulent times*. [S.l.: s.n.], 2005. p. 1–42. Citado na página 26.

PAN, Y. et al. Factors Associated with HIV Testing Among Participants from Substance Use Disorder Treatment Programs in the US: A Machine Learning Approach. *AIDS and Behavior*, v. 21, n. 2, p. 534–546, 2 2017. ISSN 1090-7165. Citado 5 vezes nas páginas 58, 65, 107, 120 e 146.

PETERS, M. D. *In no uncertain terms: The importance of a defined objective in scoping reviews*. [S.l.], 2016. v. 14, n. 2, 1–4 p. Disponível em: <[https://journals.lww.com/jbisrir/Fulltext/2016/02000/In\\_no\\_uncertain\\_terms\\_the\\_importance\\_of\\_a\\_defined.1.aspx](https://journals.lww.com/jbisrir/Fulltext/2016/02000/In_no_uncertain_terms_the_importance_of_a_defined.1.aspx)>. Citado na página 31.

PETERS, M. D. et al. Scoping reviews. In: *Joanna Briggs Institute reviewers manual*. The Joanna Briggs Institute Australia, 2017. v. 2015, p. 1–24. Disponível em: <<https://jbi-global-wiki.refined.site/space/MANUAL/4687342/Chapter+11%3A+Scoping+reviews>>. Citado 3 vezes nas páginas 32, 34 e 35.

PETERS, M. D. et al. Guidance for conducting systematic scoping reviews. *International Journal of Evidence-Based Healthcare*, Lippincott Williams and Wilkins, v. 13, n. 3, p. 141–146, 9 2015. ISSN 17441609. Citado 5 vezes nas páginas 31, 32, 33, 34 e 44.

POULOS, J.; ZENG, S. RNN-Based Counterfactual Prediction, With an Application to Homestead Policy and Public Schooling. *Journal of the Royal Statistical Society Series C: Applied Statistics*, v. 70, n. 4, p. 1124–1139, 7 2021. ISSN 0035-9254. Disponível em: <<https://doi.org/10.1111/rssc.12511>>. Citado na página 163.

PREST, B. C. Peaking interest: How awareness drives the effectiveness of time-of-use electricity pricing. *Journal of the Association of Environmental and Resource Economists*, The University of Chicago Press Chicago, IL, v. 7, n. 1, p. 103–143, 2020. Citado na página 163.

QU, L. et al. Evaluating the meteorological normalized PM2.5 trend (20142019) in the 2+26 region of China using an ensemble learning technique. *Environmental Pollution*, v. 266, p. 115346, 2020. ISSN 0269-7491. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0269749120360346>>. Citado na página 164.

RANA, P. et al. Predicting wasteful spending in tree planting programs in Indian Himalaya. *World Development*, v. 154, 2022. ISSN 0305750X (ISSN). Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85125243671&doi=10.1016%2fj.worlddev.2022.105864&partnerID=40&md5=69d75c5452437b0f8ca712208818b17d>>. Citado 7 vezes nas páginas 58, 65, 107, 120, 146, 163 e 164.

RANA, P.; MILLER, D. C. Machine learning to analyze the social-ecological impacts of natural resource policy: insights from community forest management in the Indian Himalaya. *Environmental Research Letters*, IOP Publishing, v. 14, n. 2, p. 24008, 2 2019. Disponível em: <<https://dx.doi.org/10.1088/1748-9326/aafa8f>>. Citado na página 164.

RATLEDGE, N. et al. Using Satellite Imagery and Machine Learning to Estimate the Livelihood Impact of Electricity Access. *ArXiv*, 2021. Disponível em: <<http://arxiv.org/pdf/2109.02890v1>>. Citado 9 vezes nas páginas 59, 61, 62, 64, 71, 107, 120, 146 e 159.

RATLEDGE, N. et al. *Using Satellite Imagery and Machine Learning to Estimate the Livelihood Impact of Electricity Access*. [S.l.], 2021. Disponível em: <<https://arxiv.org/abs/2109.02890https://www.nber.org/papers/w29237>>. Citado 2 vezes nas páginas 28 e 40.

RATLEDGE, N. et al. Using machine learning to assess the livelihood impact of electricity access. *Nature*, v. 611, n. 7936, p. 491–495, 2022. ISSN 00280836 (ISSN). Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85141914413&doi=10.1038%2f541586-022-05322-8&partnerID=40&md5=8b52afbfb6917ae90912fa7cf40acd28>>. Citado na página 159.

RAY, P. et al. Demonetization and its aftermath: an analysis based on twitter sentiments. *SADHANA-ACADEMY PROCEEDINGS IN ENGINEERING SCIENCES*, v. 43, n. 11, 2018. ISSN ["0256-2499", "0973-7677"]. Citado 6 vezes nas páginas 59, 61, 65, 107, 121 e 147.

- RUDOLPH, K. E. et al. When Effects Cannot be Estimated: Redefining Estimands to Understand the Effects of Naloxone Access Laws. *Epidemiology*, v. 33, n. 5, p. 689–698, 2022. ISSN 10443983 (ISSN). Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85135730321&doi=10.1097%2fEDE.00000000001502&partnerID=40&md5=be21b34ab1393a0056c2962446adee22>>. Citado 6 vezes nas páginas 58, 62, 63, 107, 121 e 147.
- SARKER, I. H. Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*, v. 2, n. 3, p. 160, 2021. ISSN 2661-8907. Disponível em: <<https://doi.org/10.1007/s42979-021-00592-x>>. Citado na página 25.
- SASAKI, Y.; URA, T.; ZHANG, Y. *Unconditional Quantile Regression with High Dimensional Data*. 2022. Citado na página 164.
- SETODJI, C. M. et al. The Right Tool for the Job. *Epidemiology*, v. 28, n. 6, p. 802–811, 11 2017. ISSN 1044-3983. Citado na página 63.
- SHAH, V.; KREIF, N.; JONES, A. M. Machine learning for causal inference: Estimating heterogeneous treatment effects. In: *Handbook of Research Methods and Applications in Empirical Microeconomics*. University of York, United Kingdom: Edward Elgar Publishing Ltd., 2021. p. 438–487. ISBN 9781788976480 (ISBN); 9781788976473 (ISBN). Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85124551307&partnerID=40&md5=484ea63bb734d9da6c31517cdb1bf378>>. Citado 13 vezes nas páginas 55, 57, 58, 63, 69, 70, 71, 107, 121, 148, 161, 162 e 164.
- SHALIT, U.; JOHANSSON, F. D.; SONTAG, D. Estimating individual treatment effect: generalization bounds and algorithms. In: PRECUP, D.; TEH, Y. W. (Ed.). *Proceedings of the 34th International Conference on Machine Learning*. PMLR, 2017. (Proceedings of Machine Learning Research, v. 70), p. 3076–3085. Disponível em: <<https://proceedings.mlr.press/v70/shalit17a.html>>. Citado na página 164.
- SHARMA, A.; SHEKHAR, H. A predictive analytics framework for Sustainable Water Governance. *Sustainable Computing: Informatics and Systems*, v. 32, p. 100604, 2021. ISSN 2210-5379. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2210537921000925>>. Citado 6 vezes nas páginas 58, 61, 65, 107, 121 e 148.
- SHEPHERD-BANIGAN, M. et al. Individual and program Characteristics May Drive Variability in Outcomes After Caregivers Participate in a Tailored Support Intervention. *Journal of Applied Gerontology*, v. 41, n. 8, p. 1960–1970, 2022. ISSN 07334648 (ISSN). Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85130553289&doi=10.1177%2f07334648221091564&partnerID=40&md5=0efe19c52b8549164e40e891f65a8f11>>. Citado 6 vezes nas páginas 59, 65, 72, 102, 113 e 129.
- SHI, Z. et al. Abrupt but smaller than expected changes in surface air quality attributable to COVID-19 lockdowns. *Science Advances*, v. 7, n. 3, p. eabd6696, 2021. Disponível em: <<https://www.science.org/doi/abs/10.1126/sciadv.abd6696>>. Citado na página 164.
- SINGH, R.; XU, L.; GRETTON, A. Kernel Methods for Causal Functions: Dose, Heterogeneous, and Incremental Response Curves. *ArXiv*, 2020. Disponível em: <<http://arxiv.org/pdf/2010.04855v7>>. Citado 9 vezes nas páginas 54, 57, 59, 63, 71, 72, 108, 122 e 149.



SRIVASTAVA, A. Challenges for evaluation practices and innovative approaches: Lessons during COVID-19 pandemic. *Evaluation and Program Planning*, v. 92, 2022. ISSN 01497189 (ISSN). Citado na página 159.

STEINKRAUS, A.; others. Estimating Treatment Effects With Artificial Neural NetsA Comparison to Synthetic Control Method. *Economics Bulletin*, v. 39, n. 4, p. 2778–2791, 2019. Citado na página 164.

STETTER, C.; MENNIG, P.; SAUER, J. Using Machine Learning to Identify Heterogeneous Impacts of Agri-Environment Schemes in the EU: A Case Study. *European Review of Agricultural Economics*, v. 49, n. 4, p. 723–759, 2022. ISSN 01651587 (ISSN). Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85135020553&doi=10.1093%2ferae%2fjbab057&partnerID=40&md5=e7b479742aae426cbd49b59746a2ca01>>. Citado 9 vezes nas páginas 59, 63, 64, 70, 71, 108, 122, 149 e 163.

STOFFI, F. J. B.; GNECCO, G. Causal tree with instrumental variable: an extension of the causal tree framework to irregular assignment mechanisms. *International Journal of Data Science and Analytics*, v. 9, n. 3, p. 315–337, 2020. ISSN 2364415X (ISSN). Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85088162982&doi=10.1007%2fs41060-019-00187-z&partnerID=40&md5=8d2474e738fea3796c9347eb76c89101>>. Citado 9 vezes nas páginas 54, 59, 63, 71, 102, 113, 130, 155 e 164.

STOFFI, F. J. B. et al. Estimating heterogeneous causal effects in the presence of irregular assignment mechanisms. In: *5th IEEE International Conference on Data Science and Advanced Analytics, DSAA 2018*. Institute of Electrical and Electronics Engineers Inc., 2019. p. 1–10. ISBN 9781538650905 (ISBN). Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85062848423&doi=10.1109%2fD SAA.2018.00010&partnerID=40&md5=c39a9b8b31944b8fccd1851c003a5ecc>>. Citado na página 155.

STRITTMATTER, A. *What Is the Value Added by Using Causal Machine Learning Methods in a Welfare Experiment Evaluation?* 2021. Citado na página 164.

SU, X. et al. Facilitating Score and Causal Inference Trees for Large Observational Studies. *Journal of Machine Learning Research*, v. 13, n. 95, p. 2955–2994, 2012. Disponível em: <<http://jmlr.org/papers/v13/su12a.html>>. Citado na página 164.

SUN, Y. et al. Research on the Policy Analysis of Sustainable Energy Based on Policy Knowledge Graph TechnologyA Case Study in China. *Systems*, v. 11, n. 2, 2023. ISSN 20798954 (ISSN). Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85149212727&doi=10.3390%2fsystems11020102&partnerID=40&md5=e187b35dd91a0d7d389e4d0d3ec042c1>>. Citado na página 159.

TOTH, B.; LAAN, M. J. van der. *TMLE for marginal structural models based on an instrument*. bepress, 2016. Citado na página 164.

TUKEY, J. W. We Need Both Exploratory and Confirmatory. *The American Statistician*, Taylor & Francis, v. 34, n. 1, p. 23–25, 1980. Disponível em: <<https://www.tandfonline.com/doi/abs/10.1080/00031305.1980.10482706>>. Citado na página 63.

VALENTE, M. Policy evaluation of waste pricing programs using heterogeneous causal effect estimation. *Journal of Environmental Economics and Management*, v. 117, 2023. ISSN 00950696 (ISSN). Citado 11 vezes nas páginas 51, 58, 63, 69, 70, 71, 108, 122, 150, 162 e 163.

VALLE-CRUZ, D. et al. Assessing the public policy-cycle framework in the age of artificial intelligence: From agenda-setting to policy evaluation. *Government Information Quarterly*, Elsevier Ltd, v. 37, n. 4, 10 2020. ISSN 0740624X. Citado na página 26.

VARIAN, H. R. Big Data: New Tricks for Econometrics. *Journal of Economic Perspectives*, v. 28, n. 2, p. 3–28, 5 2014. Disponível em: <<https://www.aeaweb.org/articles?id=10.1257/jep.28.2.3>>. Citado na página 55.

VIVIANO, D. Policy Targeting under Network Interference. *ArXiv*, 2023. Disponível em: <<http://arxiv.org/pdf/1906.10258v13>>. Citado 8 vezes nas páginas 51, 54, 59, 63, 71, 108, 123 e 150.

VU, T. V. et al. Assessing the impact of clean air action on air quality trends in Beijing using a machine learning technique. *Atmospheric Chemistry and Physics*, v. 19, n. 17, p. 11303–11314, 2019. Disponível em: <<https://acp.copernicus.org/articles/19/11303/2019/>>. Citado na página 164.

WANG, J.; SUN, L. Reducing Bus Bunching with Asynchronous Multi-Agent Reinforcement Learning. *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*, p. 426–433, 8 2021. Disponível em: <<https://doi.org/10.24963/ijcai.2021/60>>. Citado na página 159.

WANG, L. et al. Satellite-based assessment of the long-term efficacy of PM2.5 pollution control policies across the Taiwan Strait. *Remote Sensing of Environment*, v. 251, 2020. ISSN 00344257 (ISSN). Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85090883215&doi=10.1016%2fj.rse.2020.112067&partnerID=40&md5=7329b32bfc4e67030f870ee1dd7cde08>>. Citado 7 vezes nas páginas 57, 58, 61, 64, 108, 123 e 151.

WANG, L. et al. Machine learning spatio-temporal epidemiological model to evaluate Germany-county-level COVID-19 risk. *Machine Learning: Science and Technology*, v. 2, n. 3, 2021. ISSN 26322153 (ISSN). Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85108533527&doi=10.1088%2f2632-2153%2fac0314&partnerID=40&md5=01cd3470823c2bf4ffe8957c5fcf54c3>>. Citado 5 vezes nas páginas 58, 65, 108, 123 e 151.

WOZNY, F. The Impact of COVID-19 on Airfares A Machine Learning Counterfactual Analysis. *Econometrics*, v. 10, n. 1, 2022. ISSN 22251146 (ISSN). Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85125054102&doi=10.3390%2feconometrics10010008&partnerID=40&md5=b6d5c0269a64dcff5212ef57ca9166dd>>. Citado 9 vezes nas páginas 58, 59, 64, 108, 123, 152, 161, 162 e 163.

YANG, C. et al. Data-driven prediction and evaluation on future impact of energy transition policies in smart regions. *Applied Energy*, v. 332, 2023. ISSN 03062619 (ISSN). Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85144476692&doi=10.1016%2fj.apenergy.2022.120523&partnerID=40&md5=71eb0643f2314d1d434da21b57cd732d>>. Citado 6 vezes nas páginas 51, 59, 65, 108, 124 e 152.

- YANG, Y.; GUO, L. Research on Diagnostic Test and Treatment for Higher Education System. In: *2021 IEEE 3rd International Conference on Frontiers Technology of Information and Computer (ICFTIC)*. [S.l.: s.n.], 2021. p. 291–300. ISBN VO -. Citado na página 160.
- YAO, L. et al. A Survey on Causal Inference. *ACM Trans. Knowl. Discov. Data*, v. 15, n. 5, 2021. ISSN 1556-4681. Disponível em: <<https://doi.org/10.1145/3444944>>. Citado na página 55.
- YIN, H. et al. Land use and land cover change in Inner Mongolia - understanding the effects of China's re-vegetation programs. *REMOTE SENSING OF ENVIRONMENT*, v. 204, p. 918–930, 2018. ISSN ["0034-4257", "1879-0704"]. Citado 6 vezes nas páginas 58, 61, 64, 109, 124 e 153.
- ZARE, S. et al. Use of Machine Learning to Determine the Information Value of a BMI Screening Program. *American Journal of Preventive Medicine*, v. 60, n. 3, p. 425–433, 2021. ISSN 0749-3797. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0749379720305134>>. Citado 5 vezes nas páginas 58, 65, 109, 124 e 153.
- ZHANG, D. et al. A machine learning model to estimate ambient PM2.5 concentrations in industrialized highveld region of South Africa. *Remote Sensing of Environment*, v. 266, 2021. ISSN 00344257 (ISSN). Citado 5 vezes nas páginas 58, 61, 109, 125 e 153.
- ZHANG, Y.; LI, H.; REN, G. Estimating heterogeneous treatment effects in road safety analysis using generalized random forests. *Accident Analysis and Prevention*, v. 165, 2022. ISSN 00014575 (ISSN). Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85120156264&doi=10.1016%2fj.aap.2021.106507&partnerID=40&md5=4fbceb4702a0ac7595d476a5a517b467>>. Citado 9 vezes nas páginas 59, 63, 71, 109, 125, 154, 161, 162 e 163.
- ZHANG, Y. et al. Significant Changes in Chemistry of Fine Particles in Wintertime Beijing from 2007 to 2017: Impact of Clean Air Actions. *Environmental Science & Technology*, American Chemical Society, v. 54, n. 3, p. 1344–1352, 2 2020. ISSN 0013-936X. Disponível em: <<https://doi.org/10.1021/acs.est.9b04678>>. Citado na página 164.
- ZHENG, Y.; ZHENG, H.; YE, X. Using machine learning in environmental tax reform assessment for sustainable development: A case study of Hubei Province, China. *Sustainability (Switzerland)*, v. 8, n. 11, 2016. ISSN 20711050 (ISSN). Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-85000715918&doi=10.3390%2fsu8111124&partnerID=40&md5=c7ba4f582d938e997e02930050259603>>. Citado na página 160.



# Apêndices



# APÊNDICE A – Classificação por características

Quadro 9 – Classificação por características

| N. | Referência                    | Palavras-chave  | País dos Autores | Revista  | Objetivo resumido  | Codificação do Objetivo |
|----|-------------------------------|---|------------------|--|--|-------------------------|
| 1  | Abrell, Kosch e Rausch (2022) | Carbon pricing; Carbon tax; Policy evaluation; Machine learning; Electricity; UK Carbon Price Support; Climate policy; Emissions abatement; Cost and Environmental Effectiveness        | Alemanha         | Journal of Environmental Economics and Management            | Analisar o impacto nas emissões e custo da política de imposto de carbono do Reino Unido.  | Aplicação               |
| 2  | Agasisti et al. (2022)        | Financial education; RCT; impact analysis; online learning  | Itália           | Applied Economics  | Investigar a efetividade de uma intervenção de educação financeira em uma escola de ensino médio italiana, conduzida durante o fechamento das escolas devido à Covid-19.       | Aplicação               |
| 3  | Ali e Shirazi (2022)          | e-waste; sustainability; extended producer responsibility; recycler qualification program; CO2 emission; data envelopment analysis; natural language processing; machine learning; BERT | Canadá           | Sustainability   | Análise da política de gestão de lixo eletrônico do Canadá por meio de uma avaliação comparativa com a Suíça.  | Aplicação               |
| 4  | Auerbach et al. (2020)        | Urban renewal; property tax; crowdsourcing; data science; income estimation   | Estados Unidos   | Environment and Planning B: Urban Analytics and City Science | Estimar o custo e impacto de um programa para prevenir o deslocamento de moradores na região Westside de Atlanta usando técnicas de ciência de dados e aprendizado de máquina. | Aplicação               |
| 5  | Ballestar et al. (2019b)      | Research evaluation; Machine learning; Longitudinal clustering; Incentive-based policies  | Espanha          | Technological Forecasting & Social Change                    | Avaliar os efeitos do programa de incentivo a pesquisadores implementado na região de Madri no período de 2005 a 2010  | Novo método             |

| N. | Referência                       | Palavras-chave  | País dos Autores | Revista   | Objetivo resumido   | Codificação do Objetivo |
|----|----------------------------------|---|------------------|---|---|-------------------------|
| 6  | Ballestar, Sainz e Sanz (2022)   | Public policy analysis; Machine Learning; educational efficiency; LOMLOE  | Espanha          | Revista Española de Pedagogia                                       | Demonstrar a necessidade de uma avaliação econômica na modificação da Lei Orgânica de Educação a fim de subsidiar a aplicação de novos recursos que serão investidos.   | Aplicação               |
| 7  | Shepherd-Banigan et al. (2022)   | caregiving; veterans; health outcomes; intervention; depression; machine learning; burden; problem solving  | Estados Unidos   | Journal of Applied Gerontology                                      | Explorar os impulsionadores de resultados relacionados aos cuidadores de família submetidos a uma intervenção.  | Aplicação               |
| 8  | Stoffi e Gnecco (2020)           | Machine learning; Causal inference; Causal trees; Instrumental variable; Application to social science; Policy evaluation   | Itália; Bélgica  | International Journal of Data Science and Analytics                 | Propor uma versão modificada do algoritmo Causal Tree para permitir inferência causal em intervenções com mecanismo irregular de designação para tratamento.  | Novo método             |
| 9  | Belloni et al. (2017)            | Machine learning; causality; Neyman orthogonality; heterogeneous treatment effects; endogeneity; local average and quantile treatment effects; instruments; local effects of treatment on the treated; propensity score; Lasso; inference after model selection; moment-condition models with a continuum of target parameters; Lasso and Post-Lasso with functional response data; randomized control trials | Estados Unidos   | Econometrica  | Fornecer estimadores eficientes e faixas de confiança honestas para uma variedade de efeitos de tratamento, incluindo efeitos médios de tratamento local (LATE) e efeitos de tratamento local por quantil (LQTE) em ambientes ricos em dados. | Novo método             |
| 10 | Brahma e Mukherjee (2018)        | LASSO; Machine Learning; Indias Mid-Day Meal scheme; Schooling  | Estados Unidos   | Applied Econometrics and International Development                  | Avaliar se o montante transferido e a quantidade de grãos fornecidos no programa de alimentação escolar são determinantes da cobertura de crianças alimentadas.   | Aplicação               |
| 11 | Brondeel, Kestens e Chaix (2017) | Active transport; Transport; Simulation study; Machine learning; Data integration; Health inequalities; RECORD cohort study; France   | França; Canadá   | International Journal of Behavioral Nutrition and Physical Activity | Avaliar o impacto do Plano de Mobilidade Urbana de Ilha de França (2010-2020) na atividade física, sob a hipótese de que as mudanças serão aplicadas, por meio de uma abordagem de simulação.   | Aplicação               |



| N. | Referência                         | Palavras-chave   | País dos Autores | Revista   | Objetivo resumido  | Codificação do Objeto |
|----|------------------------------------|--|------------------|---|--|-----------------------|
| 12 | Carter, Tjernstrom e Toledo (2019) | Agricultural productivity; Impact evaluation; Heterogeneous treatment effects  | Estados Unidos   | Journal of Development Economics  | Estudar o impacto de um programa de desenvolvimento rural que tem o objetivo de impulsionar a receita do setor de pequenas propriedades rurais na Nicarágua.   | Aplicação             |
| 13 | Chang (2020)                       | -  | Estados Unidos   | Carolina Digital Repository / Dissertation  | Avaliar os efeitos de prover métodos contraceptivos de curta e longa duração sem custo para mulheres da África Subsaariana.  | Aplicação             |
| 14 | Chen et al. (2021)                 | health insurance; childreus health; causal forest; heterogeneous treatment effects   | China            | International Journal of Environmental Research and Public Health   | Investigar o impacto do seguro médico básico (Urban and Rural Resident Basic Medical Insurance - UR-RBMI) em crianças em idade escolar e pré-escolar da zona rural da China.   | Aplicação             |
| 15 | Chen e Zhou (2022)                 | Transit fare; Transportation policy evaluation; Empirical data; Rail transit; Unsupervised learning; Wuhan                             | China            | Transportation Research Part A: Policy and Practice   | Explorar 24 semanas de dados de smartcard de Wuhan, China, com fins de quantificar empiricamente os impactos de uma política de tarifa de transporte nos padrões de viagem de diferentes usuários e grupos de usuários e a receita de caixa de tarifas.                    | Aplicação             |
| 16 | Chew et al. (2018)                 | Social network analysis; Machine learning; Social media; Public health; Tobacco control  | Estados Unidos   | International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction and Behavior Representation in Modeling and Simulation (SBP-BRIMS) | Identificar e entender as comunidades seguidoras de campanha no Twitter para informar a avaliação do programa  | Aplicação             |
| 17 | Cockx, Lechner e Bollens (2023)    | Policy evaluation; Active labour market policy; Causal machine learning; Modified causal forest; Conditional average treatment effects | Bélgica; Suíça   | Labour Economics  | Estimar os efeitos do fornecimento de políticas ativas de mercado de trabalho relacionadas a três programas de capacitação no mercado de trabalho, em vários níveis de agregação utilizando um estimador causal de aprendizado de máquina chamado Modified Causal Forest.  | Aplicação             |
| 18 | Colangelo (2021)                   | -  | Estados Unidos   | UC Irvine   | Capítulo 1: Estudar algoritmos de transferência de aprendizado e sua aplicação na inferência com amostras de tamanho insuficiente. Capítulo 3: avaliar a efetividade do estimador Double Machine Learning (DML) desenvolvido por Colangelo e Lee para tratamento contínuo. | Revisão               |

| N. | Referência                           | Palavras-chave   | País dos Autores              | Revista   | Objetivo resumido  | Codificação do Objeto |
|----|--------------------------------------|--|-------------------------------|---|--|-----------------------|
| 19 | Colangelo e Lee (2020)               | Average structural function; cross-fitting; dose-response function; doubly robust; high dimension                                | Estados Unidos                | ArXiv   | Propor um método de inferência não paramétrica para efeitos causais de variáveis de tratamento contínuo, sob inconfundibilidade e parâmetros incómodos não paramétricos ou de alta dimensionalidade.                               | Novo método           |
| 20 | Dai et al. (2022)                    | Air quality; The three-year action plan; Machine learning; Effectiveness; Assessment   | China; Estados Unidos         | Atmospheric Environment   | Avaliar a efetividade do Plano de Ação de 3 Anos para redução da poluição do ar na região de Fenwei Plain.   | Aplicação             |
| 21 | Duthil, Imoussaten e Montmair (2017) | -  | França                        | IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications | Avaliar a prioridade dada pelos municípios franceses à implementação da política de desenvolvimento sustentável da França.   | Aplicação             |
| 22 | Francke e Acosta (2021)              | children; Peru; nutrition programs and policies; anemia; child nutrition disorders   | Peru; Estados Unidos          | Apuntes   | Avaliar o impacto do programa nacional de alimentação escolar Qali Warma na anemia e desnutrição crônica de crianças de 3 a 5 anos do Peru.  | Aplicação             |
| 23 | Gao, Li e Wang (2021)                | Big data; Text mining; Machine learning; policy evaluation   | China                         | IEEE International Conference on Big Data Analytics   | Propor um sistema modular de avaliação de política pública baseado em mineração de texto e aprendizado de máquina.   | Aplicação             |
| 24 | Goller et al. (2020a)                | Programme evaluation; active labour market policy; causal machine learning; treatment effects; radius matching; propensity score | Alemanha; Suíça               | Labour Economics  | Investigar se algoritmos de aprendizado de máquina para estimar o score de propensão produzem estimativa mais confiável do efeito médio de tratamento sobre os tratados (ATE) usando uma estrutura de correspondência de raio.     | Revisão               |
| 25 | Gori et al. (2019)                   | Flood plain management; Flood risk; Land use/land cover change; Urbanization; Distributed hydrologic model                       | Estados Unidos                | Journal of Hydrology  | Aprimorar o entendimento sobre o impacto da urbanização em uma bacia hidrográfica de Houston e alterar os limites da planície de inundação de 100 anos   | Aplicação             |
| 26 | Grecov et al. (2022)                 | Causal effect; counterfactual analysis; globaltime series forecasting; neural networks (NNs); probabilistic forecasting          | Austrália                     | IEEE Transactions on Neural Networks and Learning Systems   | Apresentar um novo método para estimar os efeitos causais sobre múltiplas unidades tratadas, combinando técnicas de previsão probabilística com métodos de previsão global usando modelos de aprendizado profundo.                 | Novo método           |
| 27 | Guo et al. (2022)                    | Air quality; Clean air action; Weather normalization; Random forest model; Meteorological influence                              | China; Reino Unido; Austrália | Chemosphere   | Aplicar aprendizado de máquina para separar as influências meteorológicas das mudanças nas emissões de poluição e examinar as tendências e características de poluentes atmosféricos em 12 megacidades chinesas entre 2013 e 2020. | Aplicação             |

| N. | Referência                               | Palavras-chave  | País dos Autores                   | Revista  | Objetivo resumido  | Codificação do Objetivo |
|----|--|---|------------------------------------|--|--|-------------------------|
| 28 | <a href="#">Ha e Grubert (2023)</a>      | Public comments; Energy policy; Public attitudes; Convolutional neural networks; Supervised machine learning  | Estados Unidos                     | Energy Research & Social Science                           | Investigar se um método híbrido de codificação qualitativa com abordagem de aprendizado de máquina pode classificar grandes quantidades de dados textuais relacionados à política mantendo o esforço manual razoável.                            | Aplicação               |
| 29 | <a href="#">Hatamyar e Kreif (2023)</a>  | -   | Reino Unido                        | ArXiv  | Analisar diferentes métodos de aprendizado de máquina na tarefa de aprender regras de alocação que otimizem os efeitos de uma política.  | Revisão                 |
| 30 | <a href="#">Heiler e Knaus (2021)</a>    | causal inference; causal machine learning; double machine learning; heterogeneous treatment effects; overlap; treatment versions  | Dinamarca; Alemanha                | ArXiv  | Propor um método de decomposição para expor heterogeneidade mascarada, evitar descobertas espúrias e avaliar a qualidade da atribuição ao tratamento quando a variável binária de tratamento não corresponde com os vários tratamentos efetivos. | Novo método             |
| 31 | <a href="#">Hu e Huang (2018)</a>        | Big data analytics; automated machine learning(AutoML); causal impact analytics; counterfactuals; BayesianStructural Time Series (BSTS); high-dimensional machine learning; policy evaluation | China                              | International Conference on Telecommunication and Networks | Tratar os desafios da extensão de AutoML aplicada à análise de impacto causal de avaliação de política e desenvolver os passos necessários para automatização da geração de modelos de ML para inferência causal.                                | Revisão                 |
| 32 | <a href="#">Kallus e Zhou (2019)</a>     | -   | Estados Unidos                     | Conference on Neural Information Processing Systems        | Demonstrar a identificação de impacto discrepante de regras personalizadas no âmbito de uma intervenção.   | Revisão                 |
| 33 | <a href="#">Knaus (2022)</a>             | Causal machine learning; conditional average treatment effects; policy learning; individualized treatment rules; multiple treatments; DR-learner  | Suíça                              | The Econometrics Journal                                   | Revisar, aplicar e estender métodos baseados em Double Machine Learning (DML) com foco na avaliação de programa sob a suposição de inconfundibilidade.   | Novo método             |
| 34 | <a href="#">Kreif e DiazOrdaz (2019)</a> | Machine learning; causal inference; treatment effects; health economics; program evaluation; policy evaluation; doubly robust methods; matching   | Reino Unido                        | ArXiv  | Revisar algoritmos populares de aprendizado de máquina e ilustrar a aplicação na estimação de efeitos de tratamento binário sob a suposição de não existir confundidores não observados.   | Revisão                 |
| 35 | <a href="#">Kreif et al. (2022b)</a>     | Policy evaluation; Machine learning; Heterogenous treatment effects; Health insurance   | Reino Unido; Indonésia; Luxemburgo | Health Services and Outcomes Research Methodology          | Explorar a abordagem de aprendizado de máquina causal na avaliação de política de saúde com confundidores fortes observados, mais especificamente no contexto de seguro de saúde maternal na Indonésia.  | Aplicação               |

| N. | Referência                               | Palavras-chave   | País dos Autores      | Revista  | Objetivo resumido   | Codificação do Objetivo |
|----|--|--|-----------------------|--|---|-------------------------|
| 36 | <a href="#">Kumar e Sharma (2020)</a>    | Sustainable agriculture; Agricultural governance; Opinion mining; Machine learning; Twitter  | Índia                 | Sustainable Computing: Informatics and Systems                 | Propor uma abordagem para governança agrícola sustentável usando mineração de opinião em rede social usando aprendizado de máquina supervisionado.  | Aplicação               |
| 37 | <a href="#">Li et al. (2021)</a>         | food price; pig price; food policy control; text mining; PMC index model   | China; Estados Unidos | Agriculture  | Estudar as políticas de regulação de preço de suínos da China e seus efeitos utilizando métodos de mineração de textos, Latent Dirichlet Allocation e o modelo Policy Modeling Consistency Index para análise quantitativa. | Aplicação               |
| 38 | <a href="#">Linden e Yarnold (2016a)</a> | balance; causal inference; machine learning; matching; propensity score  | Estados Unidos        | Journal of Evaluation in Clinical Practice                     | Proposta de um framework analítico para avaliação de programa que utiliza aprendizado de máquina para análise do balanceamento das covariáveis após o pareamento e análise dos efeitos de tratamento.                       | Aplicação               |
| 39 | <a href="#">Linden e Yarnold (2018)</a>  | balance; bias; causal inference; confounding; interrupted time series analysis; machine learning; matching; optimal discriminant analysis                                | Estados Unidos        | Journal of Evaluation in Clinical Practice                     | Apresentação de uma nova abordagem de aprendizado de máquina com Optimal Discriminant Analysis para avaliação dos efeitos de tratamento em análises de séries temporais interrompidas de múltiplos grupos.                  | Aplicação               |
| 40 | <a href="#">Lu et al. (2022)</a>         | National mangrove nature reserve; Conservation effectiveness; Systematic assessment; Object-oriented deep learning classification; Landscape development intensity index | China; Estados Unidos | Ecological Indicators  | Avaliação sistemática da efetividade de conservação de reservas naturais nacionais de mangue da China.  | Aplicação               |
| 41 | <a href="#">Ma, Chen e Lei (2022)</a>    | -  | China                 | Mathematical Problems in Engineering                           | Proposta de um método baseado em rede neural para predição e avaliação de comportamento de conluio vertical de licitação de projeto de investimento governamental.  | Aplicação               |
| 42 | <a href="#">Ma e Zhang (2014)</a>        | hierarchical clustering; education quality; multi-agent systems; social network analysis   | Estados Unidos        | IEEE International Conference on Systems, Man, and Cybernetics | Avaliar o impacto de diferentes políticas de financiamento no desempenho de estudantes da educação básica a partir da estrutura hierárquica definida pelas interações sociais dos agentes do ambiente escolar.              | Aplicação               |
| 43 | <a href="#">Matenga (2022)</a>           | Energy health Sustainable; development goals; Clustering; Machine learning   | Estados Unidos        | Sustainable Energy Technologies and Assessments                | Desenvolver um modelo de avaliação holísticos de países sobre o progresso quanto ao Objetivo de Desenvolvimento Sustentável 7 que trata de energia limpa e acessível.   | Aplicação               |

| N. | Referência                                 | Palavras-chave  | País dos Autores                     | Revista   | Objetivo resumido  | Codificação do Objetivo |
|----|--|---|--------------------------------------|---|--|-------------------------|
| 44 | McFowland, So-<br>manchi e Neill<br>(2023) | -   | Estados Unidos                       | ArXiv   | Propor um método para descobrir qual subpopulação é mais afetada significativamente pelo tratamento em um experimento aleatório.   | Novo método             |
| 45 | Pan et al. (2017)                          | Multivariate analysis; Supervised learning; Random forest; HIV testing; Substance use; Sexual risk behaviors      | Estados Unidos                       | AIDS and Behavior   | Descobrir as variáveis mais relevantes para prever o recebimento de teste de HIV entre usuários de substâncias de programas de tratamento de transtornos por uso de substâncias.                                   | Aplicação               |
| 46 | Rana et al. (2022)                         | Natural climate solutions; Tree planting programs; Wasteful spending; Forests; Himalaya; Machine Learning         | Estados Unidos; Índia; Coreia do Sul | World Development   | Avaliar o impacto de programas de plantio de árvores em Himachal Pradesh.  | Aplicação               |
| 47 | Ratledge et al.<br>(2021a)                 | -   | Estados Unidos                       | ArXiv   | Mostrar como o avanço em imagens de satélite e aprendizado de máquina podem ser usados para desenvolver medidas de meio de sustento de vida locais e a aplicação na avaliação de impacto do acesso à eletricidade. | Aplicação               |
| 48 | Ray et al. (2018)                          | Demonetization; data analysis; data preprocessing; opinion extraction; sentiment classification; machine learning | Índia                                | Sdhan Academy Proceedings in Engineering Sciences                         | Propor uma abordagem para análise de sentimento usando dados do Twitter para o esforço de demonetização do Governo Indiano.  | Aplicação               |
| 49 | Rudolph et al.<br>(2022)                   | -   | Estados Unidos                       | Epidemiology  | Evidenciar situações práticas de violação da suposição de positividade que podem ocorrer na tentativa de estimar os efeitos de políticas de saúde.   | Aplicação               |
| 50 | Shah, Kreif e Jones<br>(2021)              | -   | Reino Unido                          | Handbook of Research Methods and Applications in Empirical Microeconomics | Prover conhecimento e gerar percepções sobre aprendizado de máquina causal e destacar alguns avanços recentes da literatura sobre estimação de efeitos heterogêneos.   | Revisão                 |
| 51 | Sharma e Shekhar<br>(2021)                 | Environmental sustainability; Natural resource conservation; Water governance; Opinion mining; Social web         | Índia                                | Sustainable Computing: Informatics and Systems                            | Propor um framework para entender a opinião pública sobre políticas governamentais para aprimorar a governança da água sustentável.  | Aplicação               |
| 52 | Martins et al.<br>(2021)                   | Machine Learning; Prescription Programs; Analgesics; Opioid   | Estados Unidos                       | Epidemiology  | Identificar cláusulas legais relacionadas à alta taxa de dispensação de opióides e de altas doses nas diferentes leis promulgadas pelos entes dos Estados Unidos.  | Aplicação               |

| N. | Referência                     | Palavras-chave   | País dos Autores                         | Revista   | Objetivo resumido   | Codificação do Objetivo |
|----|--------------------------------|--|--|---|---|-------------------------|
| 53 | Singh, Xu e Gretton (2020)     | -  | Estados Unidos; Reino Unido              | ArXiv   | Propor estimadores baseados no modelo de regressão de kernel ridge para funções causais não paramétricas como curvas de resposta de dose, heterogênea e incremental.            | Novo método             |
| 54 | Stetter, Mennig e Sauer (2022) | Agri-environment schemes; impact evaluation; heterogeneous treatment effects; causal machine learning; random forests (RFs); EU common agricultural policy (CAP) | Alemanha                                 | European Review of Agricultural Economics         | Identificar a efetividade ambiental dos esquemas agroambientais na mitigação dos efeitos ambientais adversos.   | Aplicação               |
| 55 | Valente (2023)                 | Waste pricing; Causal effect heterogeneity; Machine learning   | Áustria; Alemanha                        | Journal of Environmental Economics and Management | Estudar os efeitos heterogêneos da introdução de preços de resíduos domésticos não classificados no custo e demanda municipal.  | Aplicação               |
| 56 | Viviano (2023)                 | -  | Estados Unidos                           | ArXiv   | Estudar o problema e propor método de alocação ótima de tratamento de uma política na presença de spillover, usando dados de um quasi-experimento.                              | Novo método             |
| 57 | Wang et al. (2020)             | Air pollution control policy; Efficacy evaluation; PM2.5 exposure assessment; Random Forest; Taiwan Strait   | China; Estados Unidos                    | Remote Sensing of Environment                     | Avaliar os efeitos das políticas de controle da poluição do ar na concentração de material particulado final (PM2.5) na região do Estreito de Taiwan no período de 2005 a 2018. | Aplicação               |
| 58 | Wang et al. (2021)             | SUIR model; COVID-19 pandemic; deep learning; cellular automata  | China; Alemanha                          | Machine Learning: Science and Technology          | Desenvolver modelo preditivo para avaliar tempestivamente o risco da Covid-19.  | Aplicação               |
| 59 | Wozny (2022)                   | machine learning; policy evaluation; aviation  | Alemanha                                 | Econometrics                                      | Estudar o desempenho de previsões de modelos de aprendizado de máquina para análise contrafactual relacionada ao transporte aéreo.  | Aplicação               |
| 60 | Yang et al. (2023)             | Energy transition; Renewable energy; Policy prediction; Policy evaluation; Machine learning  | China; Singapura; Austrália; Reino Unido | Applied Energy                                    | Desenvolver uma plataforma para prever a avaliar políticas de transição energética aplicando rede neural artificial e modelo de difusão tecnológica.                            | Aplicação               |

| N. | Referência             | Palavras-chave   | País dos Autores              | Revista                                 | Objetivo resumido   | Codificação do Objetivo |
|----|------------------------|--|-------------------------------|---|---|-------------------------|
| 61 | Yin et al. (2018)      | Land use and land cover change; Deforestation; Afforestation; Cropland retirement; MODIS time series; Trajectory-based approach; Inner Mongolia; Machine learning; Random forest; Probability scores | Alemanha; China               | Remote Sensing of Environment           | Mapear o uso e cobertura da terra no interior da Mongólia à luz das políticas de proteção ambiental chinesas.   | Aplicação               |
| 62 | Zare et al. (2021)     | -  | Estados Unidos                | American Journal of Preventive Medicine | Avaliar o valor informacional do Programa de Rastreio de Índice de Massa Corporal de Arkansas por meio da verificação se a disponibilidade da informação de IMC na educação infantil melhora a predição de obesidade no quarto ano estudantil para além dos indicadores já disponíveis na ausência do programa. | Aplicação               |
| 63 | Zhang et al. (2021)    | PM2.5; MAIAC AOD; random forest; air quality standard; South Africa  | Estados Unidos; África do Sul | Remote Sensing of Environment           | Testar se modelos de aprendizado de máquina são adequados para regiões com observações terrestres esparsas, como na África do Sul, e quais preditores foram importantes na modelagem de PM2.5.  | Aplicação               |
| 64 | Zhang, Li e Ren (2022) | Heterogeneous treatment effects; Road safety evaluation; Speed enforcement cameras; Generalized random forest; Causal machine learning   | China                         | Accident Analysis and Prevention        | Introduzir o uso de Generalized Random Forests, uma técnica de aprendizado de máquina para inferência causal, para estimação de efeitos de tratamento heterogêneos na análise de segurança de rodovias.   | Aplicação               |





# APÊNDICE B – Classificação por objeto

Quadro 10 – Classificação por objeto

| N. | Referência                                    | Intervenção  | Função de Governo | Causal | Registros | Variáveis | Fonte   |
|----|---|--|-------------------|--------|-----------|-----------|---|
| 1  | <a href="#">Abrell, Kosch e Rausch (2022)</a> | Imposto de carbono sobre uso de combustível fóssil do Reino Unido chamado Carbon Price Support | Gestão Ambiental  | Sim    | -         | -         | Dados combinados de várias fontes públicas ELEXON: <a href="http://www.elexon.co.uk">www.elexon.co.uk</a> ; Thomson Reuters EIKON database; Carbon price floor (CPF) and the price support mechanism. House of Commons - Briefing paper (05927); Carbon price floor: Reform. <a href="http://www.gov.uk/government/publications/carbon-price-floor-reform">www.gov.uk/government/publications/carbon-price-floor-reform</a> ; European energy exchange: Daily spot prices EU emission allowances; Electricity then year statement 2011 - appendix F. <a href="http://www.nationalgrid.com/uk/publications">www.nationalgrid.com/uk/publications</a> ; Variable pitch. <a href="http://www.variablepitch.co.uk">www.variablepitch.co.uk</a> ; European Climate assessment & data. <a href="http://eca.knmi.nl/dailydata/predeterminedseries.php">http://eca.knmi.nl/dailydata/predeterminedseries.php</a> ; Revised 1996 IPCC Guidelines for National Greenhouse Gas Inventories. IPCC; DG Climate action: European union transaction log. <a href="http://ec.europa.eu/environment/ets">http://ec.europa.eu/environment/ets</a> . |
| 2  | <a href="#">Agasisti et al. (2022)</a>        | Ensino à distância de educação financeira para alunos do ensino médio                          | Educação          | Não    | 175       | 24        | Dados dos testes de proficiência e de registros administrativos dos alunos relacionados a comportamento financeiro, atitudes financeiras, socioeconômicos, demográficos e desempenho acadêmico.   |

| N. | Referência                                     | Intervenção   | Função de Governo | Causal | Registros | Variáveis | Fonte   |
|----|--|---|-------------------|--------|-----------|-----------|---|
| 3  | <a href="#">Ali e Shirazi (2022)</a>           | Política de gestão de lixo eletrônico do Canadá   | Gestão Ambiental  | Não    | 23028     | -         | 463 artigos de pesquisa obtidos por meio de pesquisa na base Scopus relacionadas a gestão de lixo eletrônico. Documentos públicos relacionados a gestão de lixo eletrônico no Canadá e Suíça, incluindo documentos governamentais, pesquisas, leis, entre outros.   |
| 4  | <a href="#">Auerbach et al. (2020)</a>         | Programa anti-deslocamento de moradores em uma região em reurbanização de Atlanta: Anti-Displacement Tax Fund | Urbanismo         | Não    | 2600      | 20        | Características individuais dos lares e histórico de avaliação de impostos de 2005 a 2016 de Fulton County Tax Assessor. Dados de título de dívida do proprietário de Georgia Superior Court Clerkso Cooperative Authority. Características de propriedades da região de US Bureau of Labour Statistics' Consumer Expenditure Survey. Dados de estimativa de aluguel raspados de Zillow em julho/2017. Estimativas populacionais de American Community Survey 2015. |
| 5  | <a href="#">Ballestar et al. (2019b)</a>       | Programa de incentivo a pesquisadores acadêmicos de Madrid  | Educação          | Não    | 5861      | 96        | Dados anonimizados de 25.000 acadêmicos participantes de 6 universidades públicas de Madrid que fizeram parte dos 6 anos do programa. O total que participou de todos os 6 anos foi 5.861. Um registro por pesquisador, sendo 12 variáveis de características do pesquisador (tipo de contrato, sexo e área de conhecimento) e 84 relativa ao seu desempenho longitudinal nos 6 anos de avaliação.  |
| 6  | <a href="#">Ballestar, Sainz e Sanz (2022)</a> | Programa para melhoria do êxito educacional da região de Castilla y León                                      | Educação          | Não    | 1739      | 5         | Dados de estudantes do quarto ano da Educação Secundária Obrigatória que participaram do Programa para la Mejora del Éxito Educativo no ano acadêmico de 2019-2020.   |

| N. | Referência                                       | Intervenção  | Função de Governo   | Causal | Registros | Variáveis | Fonte  |
|----|--|--|---------------------|--------|-----------|-----------|--|
| 7  | <a href="#">Shepherd-Banigan et al. (2022)</a>   | Programa de suporte a cuidadores de militares veteranos chamado Operation Family Caregiver   | Assistência social  | Não    | 992       | 30        | Dados de uma coorte de 2015 a 2019 relativo ao programa OFC que era administrado pelo Rosalynn Carter Institute for Caregivers.  |
| 8  | <a href="#">Stoffi e Gnecco (2020)</a>           | Programas de acesso facilitado a crédito para o desenvolvimento de pequenas empresas do ramo de artesanato na região da Toscana, Itália: Programs for the de Development of Crafts | Comércio e Serviços | Sim    | 987       | -         | Dados de Artigian Credito Toscano e da Chamber of Commerce sobre as empresas participantes e não participantes do programa, incluindo vendas, empregados, localização da empresa, ano de abertura, situação legal, canal de distribuição principal, se integrante do programa.   |
| 9  | <a href="#">Belloni et al. (2017)</a>            | Programa de previdência dos Estados Unidos patrocinado pelo empregador, chamado 401(k)   | Previdência Social  | Sim    | 9915      | 1756      | Dados no nível de moradia a partir da pesquisa Survey of Income and Program Participation (SIPP) de 1991, contém dados individuais, da família e de renda. Como são testados diferentes conjuntos de variáveis, as combinações variam de 32 a 1756.  |
| 10 | <a href="#">Brahma e Mukherjee (2018)</a>        | Programa de alimentação escolar da Índia   | Educação            | Não    | 120       | 35        | Relatórios de cobertura do programa do Ministry of Human Resource Development da Índia e níveis de renda por estado disponíveis em <a href="http://indiastat.com">Indiastat.com</a> ; Memorandos administrativos de grãos do sítio oficial e projeção da população do Ministry of Human Resource Development da Índia. |
| 11 | <a href="#">Brondeel, Kestens e Chaix (2017)</a> | Plano de Mobilidade Urbana de Ilha de França   | Urbanismo           | Não    | 82084     | 45        | Pesquisa "Enquête Globale Transport" realizada na Ilha de França a cada 10 anos, foram utilizados dados de pessoas de 35 a 83 anos, total de 21.332 pessoas e 82.084 viagens. Estudo RECORD GPS que coletou dados de transporte e acelerômetro de 236 participantes por 7 dias, totalizando 7.138 viagens.             |

| N. | Referência   | Intervenção   | Função de Governo | Causal | Registros | Variáveis | Fonte   |
|----|--|---|-------------------|--------|-----------|-----------|---|
| 12 | <a href="#">Carter, Tjernstrom e Toledo (2019)</a> | Programa de desenvolvimento dirigido a pequenas propriedades rurais na Nicarágua chamado Rural Business Development | Agricultura       | Sim    | 1396      | 14        | Dados do programa relacionados aos fazendeiros, propriedade, moradores da propriedade, despesas, crédito, eficiência técnica, atividade alvo. Os pesquisadores atuaram junto com os gestores para conduzir o experimento desde a seleção dos participantes.   |
| 13 | <a href="#">Chang (2020)</a>                       | Fornecimento de contraceptivos de curta e longa duração para mulheres da África Subsaariana                         | Saúde             | Sim    | 29833     | 38        | Dados da pesquisa Performance Monitoring and Accountability 2020, coletados de 2014 a 2017 de moradores, mulheres e prestadores de serviço de saúde em 11 países que assumiram compromisso com o esforço Family Planning 2020. Os dados utilizados são de 2016/2017 referentes a 8 países da África Subsaariana: Burkina Faso, Cote d'Ivoire, Ethiopia, Ghana, Kenya, Niger, Nigeria, and Uganda. |
| 14 | <a href="#">Chen et al. (2021)</a>                 | Seguro médico básico para residentes da China chamado Urban and Rural Resident Basic Medical Insurance              | Saúde             | Sim    | 5552      | 11        | China Family Panel Studies (CFPS) 2018 do Institute of SocialScience Survey (ISSS) de Peking University. Chindrens Database in 2018.  |
| 15 | <a href="#">Chen e Zhou (2022)</a>                 | Política de ajuste tarifário do metrô de Wuhan, China   | Transporte        | Não    | 7632369   | -         | Registros de viagens dos smartcards de metrô de Wuhan Transportation Development Strategy Institute do período de jan/2018 a dez/2019, uma semana por mês. As transações são identificadas pelo número único do smartcard, sendo que os passageiros que viajaram ao menos uma vez nessas semanas foram considerados passageiros frequentes.   |

| N. | Referência                                      | Intervenção  | Função de Governo | Causal | Registros | Variáveis | Fonte  |
|----|---|--|-------------------|--------|-----------|-----------|--|
| 16 | <a href="#">Chew et al. (2018)</a>              | Campanha pública no Twitter para educar jovens sobre os efeitos nocivos de uso de tabaco   | Saúde             | Não    | 3184      | 38200     | Lista de seguidores da conta @knowtherealcast, que era a conta da campanha, e seus respectivos seguidores, no ano de 2014. Os últimos tweets de cada conta, até o máximo de 200 de cada uma. Dados obtidos por meio da API REST do Twitter.  |
| 17 | <a href="#">Cockx, Lechner e Bollens (2023)</a> | Três ações de capacitação de desempregados parte do fornecimento de políticas ativas de mercado de trabalho da região de Flandres na Bélgica | Trabalho          | Sim    | 73582     | 54        | Dados de registros administrativos contendo informações sociodemográficas, de emprego, histórico de trabalho, participação no programa, entre outros, de desempregados que solicitação seguro desemprego após demissão involuntária no período de dezembro de 2014 a junho de 2016.  |
| 18 | <a href="#">Colangelo (2021)</a>                | Programa de capacitação para o trabalho direcionado a jovens desfavorecidos nos Estados Unidos chamado Job Corps                             | Trabalho          | Sim    | 11250     | 40        | Capítulo 1: originário de estudo anterior sobre o programa: Schochet, Peter Z, John Burghardt, and Sheena McConnell (2008). Does job corps work? Impact findings from the national job corps study. In: American economic review 98.5, pp. 186486. 11.250 registros. Capítulo 3: originário de estudo anterior sobre o programa: Hsu, Yu-Chin et al. (2018). Direct and indirect effects of continuous treatments based on generalized propensity score weighting. SES Working Paper 495, University of Frimbourg. 4024 registros e quantidade de variáveis não informada. |

| N. | Referência   | Intervenção  | Função de Governo    | Causal | Registros | Variáveis | Fonte  |
|----|--|--|----------------------|--------|-----------|-----------|--|
| 19 | <a href="#">Colangelo e Lee (2020)</a>               | Programa de capacitação para o trabalho direcionado a jovens desfavorecidos nos Estados Unidos chamado Job Corps   | Trabalho             | Sim    | 4024      | 40        | Originário de estudo anterior sobre o programa: Hsu, Y.-C., M. Huber, Y.-Y. Lee, and L. Lettry (2020). Direct and indirect effects of continuous treatments based on generalized propensity score weighting. <i>Journal of Applied Econometrics</i> 35(7), 8148-40.  |
| 20 | <a href="#">Dai et al. (2022)</a>                    | Plano de Ação de 3 Anos para redução da poluição do ar na região de Fenwei Plain   | Gestão Ambiental     | Não    | -         | 15        | Concentração dos poluentes por hora de 2017 a 2020 em Fenwei Plain de China National Environmental Monitoring Network (CEM). Dados meteorológicos por hora de Integrated Surface Database da National Oceanic and Atmospheric Administration (pacote R worldMet e dataset ERA5). Emissões mensais dos poluentes em Fenwei Plain de Multi-resolution Inventory for China. |
| 21 | <a href="#">Duthil, Imoussaten e Montmain (2017)</a> | Financiamento de municípios para ações de desenvolvimento sustentável  | Administração        | Não    | 13794     | -         | Relatórios publicados pelas administrações municipais ou grandes corporações de 265 municípios franceses, em torno de 50 documentos por município.   |
| 22 | <a href="#">Francke e Acosta (2021)</a>              | Programa nacional de alimentação escolar do Peru chamado Qali Warma  | Educação             | Sim    | 15334     | 19        | Características das crianças, incluindo saúde, e suas moradias da Pesquisa Demográfica e de Saúde Familiar do Peru (ENDES). Dados administrativos das escolas sobre as características de fornecimento de alimentação de Ministério de Desarrollo e Inclusión Social.  |
| 23 | <a href="#">Gao, Li e Wang (2021)</a>                | Plano de Trabalho da Comissão Municipal de Pequim de Ciência e Tecnologia para a Construção da Comunidade de Inovação Colaborativa Beijing-Tiangjin-Heibei | Ciência e Tecnologia | Não    | 74        | -         | Textos de políticas baixados de Ministérios e portais oficiais específicos do Governo da China a partir de buscas pelas palavras-chave "colaboração", "ciência e tecnologia" e "inovação".   |

| N. | Referência                            | Intervenção  | Função de Governo | Causal | Registros | Variáveis | Fonte   |
|----|---------------------------------------|--|-------------------|--------|-----------|-----------|---|
| 24 | <a href="#">Goller et al. (2020a)</a> | Programa de treinamento para desempregados da Alemanha chamado Determining, Reducing and Removing Employment Impediments   | Trabalho          | Sim    | 276637    | 27        | Integrated Employment Biographies e outros registros administrativos disponíveis no Institute for Employment Research, dados locais de Statistics Department of the German Federal Employment Agency.   |
| 25 | <a href="#">Gori et al. (2019)</a>    | Requisitos para mecanismos de drenagem e controle de inundação do Condado de Harris em Houston, Texas: Harris County Flood Control Districts Policy; Criteria and Procedure Manual | Gestão Ambiental  | Não    | -         | 12        | Mapas de uso e cobertura de terra de National Land Cover Dataset de 2001, 2006 e 2011. Dados de impulsoadores de desenvolvimento urbano (biofísicos, de infraestrutura e socioeconômicos) de Census Bureau, National Hydrography Dataset, Houston-Galveston Area Council e Texas Education Agency.  |
| 26 | <a href="#">Grecov et al. (2022)</a>  | Emissão de licença para comercialização de bebidas alcoólicas na Austrália   | Administração     | Sim    | -         | -         | Registros de atendimento de ambulância no período de janeiro de 2015 a maio de 2019 de 79 áreas de governo local de um estado da Austrália, obtidos de National Ambulance Surveillance System (NASS). O período até maio de 2018 é pré-intervenção e a partir de junho é intervenção quando foi identificado o aumento anormal na concessão de licenças.                                      |
| 27 | <a href="#">Guo et al. (2022)</a>     | Ações de combate à poluição do ar na China chamadas Air Pollution Prevention and Control Action Plan (20132017) e Blue Sky Protection Campaign (20182020)                          | Gestão Ambiental  | Não    | -         | -         | Dados de qualidade do ar com 1 hora de resolução de 13/05/2014 a 31/12/2020 de <a href="http://beijingair.sinaapp.com">http://beijingair.sinaapp.com</a> , e de 18/01/2013 a 13/05/2014 de <a href="http://data.epmap.org/air/nations">http://data.epmap.org/air/nations</a> . Dados meteorológicos de 18/01/2013 a 31/12/2020 de <a href="http://hz.zc12369.com">http://hz.zc12369.com</a> . |
| 28 | <a href="#">Ha e Grubert (2023)</a>   | Plano de Energia Limpa dos Estados Unidos  | Energia           | Não    | 1900      | 584       | Dados de comentários públicos transcritos de audiências coletadas de <a href="https://www.regulations.gov">regulations.gov</a> .  |

| N. | Referência                               | Intervenção   | Função de Governamento | Causal | Registros | Variáveis | Fonte   |
|----|--|---|------------------------|--------|-----------|-----------|---|
| 29 | <a href="#">Hatamyar e Kreif (2023)</a>  | Programa nacional de seguro de saúde da Indonésia   | Saúde                  | Sim    | 10622     | 64        | Dados de nascimento entre 2002 e 2014, incluindo características da mãe e moradia, da Indonesian Family Life Survey (IFLS).   |
| 30 | <a href="#">Heiler e Knaus (2021)</a>    | Programa de capacitação para o trabalho direcionado a jovens desfavorecidos nos Estados Unidos chamado Job Corps  | Trabalho               | Sim    | 9708      | 55        | Dados de estudo anterior. Schochet, P. Z., Burghardt, J., & McConnell, S. (2019). Replication data for: Does job corps work? Impact findings from the national job corps study. Inter-university Consortium for Political and Social Research (ICPSR).        |
| 31 | <a href="#">Hu e Huang (2018)</a>        | Políticas de estímulo econômico americanas  | Administração          | Sim    | -         | 35582     | Dados de notícias GDELT ( <a href="https://www.gdeltproject.org/">https://www.gdeltproject.org/</a> ).  |
| 32 | <a href="#">Kallus e Zhou (2019)</a>     | Programa de aconselhamento para obtenção de emprego na França   | Trabalho               | Sim    | 7950      | 16        | Dados de estudo anterior que avaliou o mesmo programa.  |
| 33 | <a href="#">Knaus (2022)</a>             | Quatro programas de uma política de trabalho na Suíça chamada Swiss Active Labour Market Policy relacionados a procura de trabalho, treinamento vocacional, cursos de computação e cursos de línguas. | Trabalho               | Sim    | 62497     | 45        | Dados de estudo anterior que avaliou o mesmo programa: Huber, M., Lechner, M., & Mellace, G. (2017). Why do tougher caseworkers increase employment? The role of program assignment as a causal mechanism. Review of Economics and Statistics, 99(1), 180183. |
| 34 | <a href="#">Kreif e DiazOrdaz (2019)</a> | Programa nacional de seguro de saúde da Indonésia   | Saúde                  | Sim    | 10985     | 34        | Dados de nascimento entre 2002 e 2014, incluindo características da mãe e moradia, da Indonesian Family Life Survey (IFLS).   |
| 35 | <a href="#">Kreif et al. (2022b)</a>     | Programa nacional de seguro de saúde da Indonésia   | Saúde                  | Sim    | 12076     | 25        | Pesquisa Indonesian Family Life Survey (IFLS) do Central Bureau of Statistics longitudinal com 7224 lares revisitados nos anos de 2000, 2007/2008 e 2014/2015.  |
| 36 | <a href="#">Kumar e Sharma (2020)</a>    | Política agrícola da Índia  | Agricultura            | Não    | 1008      | -         | Postagens no Twitter obtidos por meio da API em 2016  |



| N. | Referência                               | Intervenção  | Função de Governo | Causal | Registros | Variáveis | Fonte  |
|----|--|--|-------------------|--------|-----------|-----------|--|
| 37 | <a href="#">Li et al. (2021)</a>         | Políticas de regulação de preço de suínos da China   | Agricultura       | Não    | 74        | -         | Textos de documentos normativos relacionados à regulação de preço de suínos obtidos por meio de buscas de palavras-chave em sites oficiais, incluindo National Development and Reform Commission, Peking University e base central de interpretação legal.   |
| 38 | <a href="#">Linden e Yarnold (2016a)</a> | Programa de cuidados médicos em casa   | Saúde             | Sim    | 2002      | 11        | Dados de avaliação prévia de um programa piloto de assistência médica primária residencial   |
| 39 | <a href="#">Linden e Yarnold (2018)</a>  | Lei para redução de vendas de cigarro  | Saúde             | Sim    | 1178      | 2         | Obtidos de outro estudo publicado.   |
| 40 | <a href="#">Lu et al. (2022)</a>         | Reservas naturais nacionais para conservação de mangue                                     | Gestão Ambiental  | Não    | -         | 22        | Imagens de satélite das regiões das reservas de 1987, 1993, 2000, 2005, 2010 e 2019, obtidas da plataforma Google Earth. Amostras de pesquisas de 2016 a 2019 de pesquisas sobre cobertura da terra. Dados anuais de precipitação e média de temperatura de 1951 a 2019 de estações meteorológicas localizadas nas reservas ou ao redor. |
| 41 | <a href="#">Ma, Chen e Lei (2022)</a>    | Regulamento de Medidas em Licitações e Concursos para Projetos de Engenharia de Construção | Administração     | Não    | 166       | 13        | Pesquisas Chinese General Social Survey 2015 e 2017 e dados do National Bureau of Statistics da China  |
| 42 | <a href="#">Ma e Zhang (2014)</a>        | Financiamento da educação básica no Condado de Bexar do Texas                              | Educação          | Não    | 345682    | -         | Dados da Agência de Educação do Texas.   |
| 43 | <a href="#">Matenga (2022)</a>           | Objetivo de Desenvolvimento Sustentável 7 - energia limpa e acessível                      | Energia           | Não    | 6510      | 17        | Indicadores do Banco Mundial de 1990 a 2019.   |

| N. | Referência   | Intervenção   | Função de Governo | Causal | Registros | Variáveis | Fonte  |
|----|--|---|-------------------|--------|-----------|-----------|--|
| 44 | <a href="#">McFowland, So-<br/>manchi e Neill<br/>(2023)</a> | Disponibilização de professores auxiliares treinados em salas de aula de escolas do Tennessee | Educação          | Sim    | 11600     | 11        | Dados do experimento Tennessee STAR: E. R. Word, J. Johnston, H. P. Bain, and Others. The state of Tennessee Student/Teacher Achievement Ratio (STAR) project: Technical report 19851990. Nashville: Tennessee State Department of Education, 1990   |
| 45 | <a href="#">Pan et al. (2017)</a>                            | Programas de tratamento de transtornos por uso de substâncias dos Estados Unidos              | Saúde             | Não    | 1281      | 119       | Estudo realizado pelo National Institute on Drug Abuse Clinical Trials Network sobre teste de HIV e aconselhamento com participantes de 12 programas de tratamento de transtorno por uso de substância baseados em comunidade. Dados incluem fatores sociodemográficos, uso de substância, comportamento sexual de risco, uso de preservativos e de depressão. |
| 46 | <a href="#">Rana et al. (2022)</a>                           | Programas Plantio de árvores no Estado Indiano de Himachal Pradesh entre 2016 e 2019          | Gestão Ambiental  | Não    | 16674     | 32        | Dados de cobertura de árvores de Forest Survey of India. Dados dos polígonos das florestas de Himachal Pradesh Forest Department GIS Lab. Dados de orçamento de arborização reportados pelo Forest Department. Dados populacionais do Censo da Índia. Outros dados climáticos, áreas e solo de outros estudos e entidades.                                     |
| 47 | <a href="#">Ratledge et al. (2021a)</a>                      | Expansão da malha de energia elétrica rural em Uganda   | Energia           | Sim    | 641621    | 13        | 14 anos de pesquisa domiciliar de Uganda (Demographic and Health Surveys) de 641.621 lares de 27.174 áreas de 25 países da África Subsariana. Também são usados dados públicos de expansão da malha elétrica construída a partir de fontes do Governo de Uganda, Banco Mundial e parceiros, e artigos acadêmicos.  |

| N. | Referência                                 | Intervenção  | Função de Governo | Causal | Registros | Variáveis | Fonte   |
|----|--|--|-------------------|--------|-----------|-----------|---|
| 48 | <a href="#">Ray et al. (2018)</a>          | Demonetização do Governo da Índia para retirar determinadas notas de circulação        | Administração     | Não    | 57536     | -         | Postagens no Twitter obtidos por meio da API entre 08/11/2016 e 14/12/2016 utilizando palavras-chave relacionadas ao projeto.   |
| 49 | <a href="#">Rudolph et al. (2022)</a>      | Leis de acesso a naloxona nos Estados Unidos   | Saúde             | Sim    | -         | 20        | Dados de vigência das leis de acesso de Prescription Drug Abuse Policy System. Dados restritos de overdose por opióide de National Center for Health Statistics. Dados de taxa de mortalidade de National Center for Health Statistics. Dados de condados, incluindo características de moradores, de Geolytics. Outros dados relacionados à prescrição de naloxona, overdose e leis correlatas no nível estadual de diversas fontes. |
| 50 | <a href="#">Shah, Kreif e Jones (2021)</a> | Programa nacional de seguro de saúde da Indonésia                                      | Saúde             | Sim    | 912812    | 97        | Dados de características individuais e moradias incluindo idade, gênero, estado civil, educação, local de residência, fatores socioeconômicos, diferenças geográficas de regiões, entre outras, da Pesquisa nacional socioeconômica de moradores da Indonésia de 2017 (SUSNAS).   |
| 51 | <a href="#">Sharma e Shekhar (2021)</a>    | Programa Namami Gange criado para conservação e rejuvenescimento do rio Ganga na Índia | Gestão Ambiental  | Não    | 1497      | -         | Publicações no Twitter por meio da API com filtro #NamamiGange por 3 meses.   |
| 52 | <a href="#">Martins et al. (2021)</a>      | Regulação de opióides dos entes dos Estados Unidos                                     | Saúde             | Não    | 162       | -         | Banco de dados de prescrição longitudinal IQVIA (2006-2016) que contém vendas de varejo de farmácias dos Estados Unidos. A população estimada dos condados do U.S. Census Bureau. Dados de decretação das cláusulas do Prescription Drug Abuse Policy System (PDAPS). Limites de prescrição de outro estudo referenciado no texto.  |

| N. | Referência                                     | Intervenção  | Função de Governo | Causal | Registros | Variáveis | Fonte  |
|----|--|--|-------------------|--------|-----------|-----------|--|
| 53 | <a href="#">Singh, Xu e Gretton (2020)</a>     | Programa de capacitação para o trabalho direcionado a jovens desfavorecidos nos Estados Unidos chamado Job Corps         | Trabalho          | Sim    | 3906      | 40        | Sugere os mesmos dados de outro estudo: Kyle Colangelo and Ying-Ying Lee. Double debiased machine learning nonparametric inference with continuous treatments.arXiv:2004.03036, 2020.  |
| 54 | <a href="#">Stetter, Mennig e Sauer (2022)</a> | Esquemas agroambientais da União Europeia como parte do Programa de Desenvolvimento Rural na região da Bavária, Alemanha | Agricultura       | Sim    | 2758      | 130       | Dados de contabilidade das fazendas, incluindo propriedade, cultivo, famílias moradoras e finanças, de European Farm Accountancy Data Network. Dados das fazendas oficiais contendo participação no programa e ambiente socioeconômico, espacial e estrutural onde se encontram.   |
| 55 | <a href="#">Valente (2023)</a>                 | Programas de pagamento pelo volume de lixo doméstico descartado não classificado em municípios da Itália                 | Gestão Ambiental  | Sim    | 19948     | 90        | Dados anuais sobre preços de resíduos, quantidades de resíduos domésticos e custos de gestão, determinantes socioeconômicos, geográficos e políticos da geração de resíduos e adoção de preços, de 2010 a 2015, obtidos por meio de raspagem e de registros administrativos de Northern and Central Italian municipalities, National Environmental Protection Agency (ISPR), National Institute of Statistics (ISTAT), comunitaliani.it, Ministry of the Interior, European Pollution Release and Transfer Register (E-PRTR), ational Environmental Protection Agency (ISFRA), the National Association of Italian Municipalities (ANCI) e empresas de gestão de resíduos. |

| N. | Referência                         | Intervenção   | Função de Governo | Causal | Registros | Variáveis | Fonte  |
|----|------------------------------------|---|-------------------|--------|-----------|-----------|--|
| 56 | <a href="#">Viviano (2023)</a>     | Experimento de sessões informativas sobre os benefícios de seguro para aumentar a adesão dos seguros contra efeitos climáticos por fazendeiros em um região suscetiva a desastre ambiental na China | Administração     | Sim    | 4511      | -         | Dados coletados no experimento, incluindo relacionamento entre os beneficiários, de Cai, J., A. De Janvry, and E. Sadoulet (2015). Social networks and the decision to insure. <i>American Economic Journal: Applied Economics</i> 7(2), 81108   |
| 57 | <a href="#">Wang et al. (2020)</a> | Políticas de controle de poluição do ar implementadas entre 2005 e 2018 na região do Estreito de Taiwan   | Gestão Ambiental  | Não    | 16540     | 6         | Concentração de PM2.5 de China National Environmental Monitoring Center, e Environmental Monitoring Centers of Xiamen, Quanzhou City and Zhangzhou in Fujian e Taiwan Environmental Protection Administration. Múltiplos produtos de dados dos satélites Terra e Aqua. Dados de clima e tempo do ERA5. Dados de uso da terra de Chinese land use/land cover (CNLUCC). Dados de rodovia de Google road map e Google Earth. Dados de população de Gridded Population of the World (GPW). |
| 58 | <a href="#">Wang et al. (2021)</a> | Política de lockdown em resposta à pandemia do Covid-19 na Alemanha   | Saúde             | Não    | 10000     | -         | Dados sobre a pandemia na Alemanha publicados pelo Robert Koch-Institut.   |
| 59 | <a href="#">Wozny (2022)</a>       | Políticas mundiais de tráfego aéreo de passageiros em resposta à pandemia do Covid-19   | Transporte        | Não    | 23508     | 100       | Dados mensais de tarifas aéreas no nível de empresas aéreas de Sabre Market Intelligence. Dados anuais das empresas aéreas de ICAO Data Plus coletados pela autoridades de aviação nacional dos estados membros.   |

| N. | Referência                         | Intervenção  | Função de Governo | Causal | Registros | Variáveis | Fonte  |
|----|------------------------------------|--|-------------------|--------|-----------|-----------|--|
| 60 | <a href="#">Yang et al. (2023)</a> | Programas de transição energética de Londres, Singapura e Califórnia   | Energia           | Não    | -         | 14        | World Bank Open Data. Statistics Singapore. London Datastore. US Bureau of Economic Analysis. UCLA Anderson Forecast. Bank of England. Federal Reserve Board of USA. United Nations: World Population Prospects 2019. Department of Finance, State of California. Public Policy Institute of California. Global Data Lab. California Air Resources Board. Singapore Energy Market Authority. U.S. Energy Information Administration. UK Office for National Statistics. Singapore National Environment Agency. UK Met Office California. UCANR. Energy Trend: UK Weather. Lazard, IRENA. |
| 61 | <a href="#">Yin et al. (2018)</a>  | Programas de proteção ambiental na China chamados Natural Forest Conservation Program (NFCP), Grain to Green Program (GGP) e Beijing and Tianjin Sandstorm Source Treatment Project (BTSS) | Gestão Ambiental  | Não    | -         | -         | Dados de satélite Terra MOD13Q1 e Aqua MYD13Q1. Dados de classificação de terra de Chinese Academy of Science.   |
| 62 | <a href="#">Zare et al. (2021)</a> | Programa de Rastreamento do Índice de Massa Corporal de Estudantes da Educação Infantil de Arkansas nos Estados Unidos   | Saúde             | Não    | 244053    | 31        | Dados do programa da população estudantil de escolas públicas de Arkansas. Dados censitários de Census Population e American Community Survey. Dados histórico de comércio de ReferenceUSA Database.   |

| N. | Referência             | Intervenção   | Função de Governo | Causal | Registros | Variáveis | Fonte   |
|----|------------------------|---|-------------------|--------|-----------|-----------|---|
| 63 | Zhang et al. (2021)    | Mudança no padrão nacional de concentração de PM2.5 ocorrida na África do Sul em 2016 | Gestão Ambiental  | Não    | 9853      | 40        | Dados de concentração de PM2.5 medidos por hora em 20 estações de monitoramento da South African Weather Service. Dados de profundidade ótica de aerossóis (AOD) dos satélites Terra e Aqua da rede AERONET. Dados meteorológicos diários de Goddard Earth Observing System Data Assimilation System GEOS- 5 Forward Processing. Dados de uso da terra de Sentinel 2. Dados populacionais de Gridded Population of the World. Dados de elevação de Advanced Spaceborne Thermal Emission and Reflection Radiometer Global Digital Elevation Model Version 3. |
| 64 | Zhang, Li e Ren (2022) | Programa de radar de velocidade do Reino Unido  | Transporte        | Sim    | -         | 14        | Dados de 771 locais com câmera (tratamento) e 4743 locais de comparação (controle) de 8 distritos administrativos: Cheshire, Dorset, Greater Manchester, Lancashire, Leicestershire, Merseyside, Sussex e West Midlands. Dados de colisões de UK Department for Transport. Dados de índice de deprivação de National Archives. Dados de população e emprego obtidos de Office for National Statistics.  |





# APÊNDICE C – Classificação por método

Quadro 11 – Classificação por método

| N. | Referência                    | Aplicação  | Aplicação codificada   | Métodos   | Tipo           | Lingua-gem | Pacotes                |
|----|-------------------------------|--|------------------------|---|----------------|------------|------------------------|
| 1  | Abrell, Kosch e Rausch (2022) | A estratégia da avaliação foi treinar um modelo, baseado em um modelo de equilíbrio econômico, que prevê resultados sobre o tratamento observado e o contrafactual. O modelo empregado para gerar o resultado das plantas supondo que não houvesse a aplicação da política foi o LASSO, que obteve melhor desempenho que o Random Forest para esse caso. Adicionalmente, o modelo foi utilizado para realizar simulações variando-se os níveis de imposto de carbono com o objetivo de investigar o que direciona a heterogeneidade no abatimento da quantidade e impacto nos custos causados pelo imposto de carbono.                                 | Previsão               | Least Absolute Shrinkage and Selection Operator (LASSO); Random Forest (RF) | Supervisionado | -          | -                      |
| 2  | Agasisti et al. (2022)        | As salas de aula foram separadas em grupos de tratamento (receberam o curso) e controle (não receberam o curso). Para avaliação foram realizados testes de proficiência com todos antes e após a intervenção para verificar o impacto no conhecimento dos alunos. Para inferência causal foram utilizados modelos OLS de regressão linear. Modelos de aprendizado de máquina foram utilizados com o objetivo de explorar o relacionamento entre as variáveis. Foram aplicados Decision Tree e Random Forest. O método SHapley Additive exPlanations (SHAP) foi empregado para auxiliar na interpretação da importância das variáveis para o resultado. | Análise dos preditores | Decision Tree (DT); Random Forest (RF)                                      | Supervisionado | R          | Pacote R: randomForest |

| N. | Referência             | Aplicação  | Aplicação codificada            | Métodos  | Tipo               | Lingua-<br>gem | Pacotes  |
|----|------------------------|--|---------------------------------|--|--------------------|----------------|--|
| 3  | Ali e Shirazi (2022)   | O aprendizado de máquina foi utilizado em um dos passos para subsidiar a avaliação por análise envoltória de dados (DEA). Como entrada para o modelo, documentos públicos foram processados no software NVivo e foram extraídos vários critérios de análise, incluindo indicadores para os dois países, que foram agrupados em quatro categorias: clareza; responsividade e inclusão; dados, comunicação, avaliação e conformidade; e emissões. O segundo passo foi classificar manualmente os artigos de pesquisa nas quatro categorias de análise. Na sequência, o modelo BERT, utilizado por meio do KeyBERT, foi aplicado para identificar nos artigos os termos mais relevantes relacionados às categorias, no total de 25. Esses dados foram utilizados como saída, ou eficiência, da DEA. A entrada para o DEA foi a quantificação dos critérios de análise presentes em cada um dos dois países. O resultado da DEA foi a análise de eficiência. | Descoberta de termos relevantes | Bidirectional Encoder Representations from Transformers (BERT)   | Não supervisionado | Python; Stata  | Pacote keybert; python Sklearn; Pacote python NLTK; Pacote python PyPDF2; Software NVivo |
| 4  | Auerbach et al. (2020) | Inicialmente, foi utilizada regressão (modelo não detalhado) para estimativa da elegibilidade por renda por moradia para o programa utilizando dados de características dos individuais dos imóveis, estimativas de valor de aluguel e valores de renda relacionados com características de imóveis. Previsão da taxa de imposto futura dos imóveis usando dados de séries temporais de imposto de uma vizinhança similar. Para isso foram testados modelos de MLR, RR, LR, DT e RF, sendo Random Forest o que apresentou melhor desempenho. A quantidade de residências elegíveis e a previsão de taxa de imposto para 7 anos foram utilizadas para demonstrar que a avaliação inicial tinha inconsistências decorrentes de dados menos granulares e também para dar transparência à comunidade.  | Previsão                        | Multiple Linear Regression (MLR); Ridge Regression (RR); Least Absolute Shrinkage and Selection Operator (LASSO); Decision Tree (DT); Random Forest (RF) | Supervisionado     | -              | -  |

| N. | Referência                     | Aplicação   | Aplicação codificada   | Métodos  | Tipo  | Lingua-gem | Pacotes        |
|----|--------------------------------|---|------------------------|--|---|------------|----------------|
| 5  | Ballestar et al. (2019b)       | Clusterização como segundo passo do modelo multinível proposto (Automated Nested Longitudinal Clustering) após a estratificação dos pesquisadores por tipo de contratação. A seguir uma rede neural para classificação nos mesmos grupos da clusterização, de forma a avaliar a robustez do primeiro método. Realizou-se análise descritiva dos grupos para avaliação.  | Classificação          | Automated Nested Longitudinal Clustering (ANLC); Multilayer Perceptron (MLP);  | Não supervisionado (ANLC); Supervisionado (MLP) | -          | -              |
| 6  | Ballestar, Sainz e Sanz (2022) | Treinamento de modelo baseado em árvore de decisão CHAID para predição se o aluno do programa irá concluir com êxito o ano escolar. As variáveis mais relevantes para predição do êxito foram anos de repetição prévia e se o estudante já repetiu de ano alguma vez. Um novo modelo foi treinado usando ANN-MLP usando os mesmos dados a fim de verificar se os resultados são equivalentes ao do primeiro modelo. As métricas foram consistentes para ambos, sem diferenças significantes. Apenas quanto à importância dos preditores houve diferença, sendo a mais relevante se o aluno já repetiu, seguido da categoria do programa, que era de três tipos. A avaliação se deu pela descrição da importância das variáveis para predição. | Análise dos preditores | Chi-square Automatic Interaction Detector (CHAID); Multilayer Perceptron (MLP) | Supervisionado                                  | -          | -              |
| 7  | Shepherd-Banigan et al. (2022) | O objetivo era verificar impulsionadores para as variáveis de resultado relacionadas a sintomas depressivos, sobrecarga subjetiva e resolução de problema. Calculou-se a mudança para cada cuidador pré e pós-intervenção. Com isso, foram treinados modelos baseados em Elastic Net Regression para seleção dos preditores importantes para cada uma das três variáveis de resultado de interesse.   | Análise dos preditores | Elastic Net Regression   | Supervisionado                                  | R; SAS     | Pacote R caret |

| N. | Referência             | Aplicação   | Aplicação codificada | Métodos  | Tipo           | Lingua-gem | Pacotes      |
|----|------------------------|---|----------------------|--|----------------|------------|--------------|
| 8  | Stoffi e Gnecco (2020) | Foram treinados modelos utilizando os novos algoritmos propostos CT-IV e HCT-IV, a variável de resultado é uma variável categórica que indica de houve aumento no número de empregados no ano subsequente ao tratamento aumentou, diminuiu ou foi mantido (1, -1 e 0, respectivamente). Como variável instrumental, vez que o tratamento não é aleatório, foi utilizada o dado se a empresa se inscreveu para o financiamento ou não, tratado por meio de escore de propensão, vez que também não é aleatório. É feita análise do Efeito Médio Causal nos Compliers (CACE).   | Inferência causal    | Causal Tree with Instrumental Variable (CT-IV); Honest Causal Tree with Instrumental Variable (HCT-IV) | Supervisionado | -          | -            |
| 9  | Belloni et al. (2017)  | São utilizados diferentes conjuntos de controle para comparação, variando de 32 a 1756 variáveis com combinações polinomiais quadráticas. A variável de resultado são os ativos financeiros do indivíduo, a de tratamento é um balanço positivo no plano de previdência, a instrumental é ser elegível ao plano. Os autores propõem um método baseado em condições de momento ortogonais e duplamente robustas. O método permite a utilização de diferentes modelos de aprendizado de máquina para aprender os componentes não-paramétricos do modelo. Na aplicação é utilizado post-LASSO. São reportados intenção média de tratamento (ATE), intenção média de tratamento nos tratados (ATE-T), intenção quantil de tratar (QTE) e intenção quantil de tratar nos tratados (QTE-T), efeitos da participação no programa como efeitos médios de tratamento local (LATE), efeitos médios de tratamento local nos tratados (LATE-T), efeitos quantis locais de tratamento (LQTE) e efeitos quantis locais de tratamento nos tratados (LQTE-T). | Inferência causal    | Least Absolute Shrinkage and Selection Operator (LASSO)  | Supervisionado | R; Matlab  | Pacote R hdm |

| N. | Referência                         | Aplicação   | Aplicação codificada   | Métodos  | Tipo           | Lingua-gem | Pacotes      |
|----|------------------------------------|---|------------------------|--|----------------|------------|--------------|
| 10 | Brahma e Mukherjee (2018)          | Seleção das variáveis importantes para predição do resultado esperado que, no caso, é a cobertura de crianças alimentadas.  | Análise dos preditores | Least Absolute Selection and Shrinkage Operator (LASSO); LASSO hierárquico | Supervisionado | -          | -            |
| 11 | Brondeel, Kestens e Chaix (2017)   | Construção de 3 modelos para predição de variáveis que compuseram os cenários elaborados. Variáveis: Duração do percurso, Modal de Transporte, Atividade Física Relacionada a Transporte de Moderada a Vigorosa. Os cenários foram analisados por estatística descritiva.   | Previsão               | Random Forest  | Supervisionado | R          | -            |
| 12 | Carter, Tjernstrom e Toledo (2019) | Os pesquisadores atuaram em conjunto dos gestores desde o início do programa. A avaliação se deu sobre dois grupos de fazendeiros separados aleatoriamente que receberam o tratamento. Um deles recebeu o tratamento após 20 meses do primeiro, assim funcionou como uma espécie de grupo de controle. Ambos foram avaliados com 18 meses do início do programa para o primeiro grupo e ao final de quatro anos da mesma data. Métodos econômicos tradicionais são utilizados para estimar intenção de tratar (ITT), efeitos médios locais de tratamento (LATE), efeito contínuo de tratamento e estimação por quantis. O algoritmo Generalized Random Forest é usado para treinar um modelo com o objetivo de identificar características mais associadas a grandes impactos no programa (CATE). | Inferência causal      | Generalized Random Forest  | Supervisionado | R          | Pacote R grf |

| N. | Referência                         | Aplicação  | Aplicação codificada                            | Métodos   | Tipo           | Lingua-gem | Pacotes                        |
|----|------------------------------------|--|---|---|----------------|------------|--------------------------------|
| 13 | <a href="#">Chang (2020)</a>       | A avaliação adotou uma abordagem com escore de propensão, vez que características dos indivíduos e acesso a serviços de planejamento familiar podem afetar o uso de contraceptivos. Para estimar o escore foram utilizados modelos de aprendizado de máquina baseado em GBM a partir de 38 covariáveis individuais. O modelo estima a probabilidade de a mulher ter acesso gratuito aos métodos contraceptivos. Os pesos gerados são aplicados em um modelo de regressão logística cujo coeficiente de interesse representava o efeito estimado do acesso gratuito a contraceptivos de longo e curto prazo. Os pesos dos modelos GBM foram comparados com modelo de regressão logística e produziram efeitos similares, com exceção dos efeitos do acesso gratuito no uso de contraceptivo de longa duração. Adicionalmente, foi utilizado um modelo de aprendizado de máquina baseado em Recursive Partitioning para identificar subgrupos que melhor predizem o uso de contraceptivo moderno (CATE). | Balanceamento de covariáveis; Inferência causal | Generalized Boosted Model (GBM); Recursive Partitioning (RPART) | Supervisionado | R; Stata   | Pacote R rpart; Pacote R twang |
| 14 | <a href="#">Chen et al. (2021)</a> | As crianças são separadas em 4760 no grupo de tratamento, que estavam no seguro médico, e 792 no grupo de controle, não tinha nenhum seguro médico. A variável dependente para crianças pré-escolares era o status de saúde e para crianças a partir de 6 anos era índice de massa corporal. Foi aplicado Propensity Score Matching para inferência causal e um modelo de aprendizado de máquina baseado em Causal Forest para identificar subgrupos com efeitos heterogêneos (Conditional Treatment Average Effects).   | Inferência causal                               | Causal Forest (CF)  | Supervisionado | -          | -                              |

| N. | Referência                         | Aplicação   | Aplicação codificada | Métodos                           | Tipo               | Lingua-gem | Pacotes |
|----|------------------------------------|---|----------------------|-----------------------------------|--------------------|------------|---------|
| 15 | <a href="#">Chen e Zhou (2022)</a> | <p>Com o objetivo de identificar os diferentes grupos de usuários do metrô, os autores treinaram dois modelos de aprendizado de máquina com o algoritmo Latent Dirichlet Allocation. Em ambos, os passageiros foram considerados como documentos para o modelo. As características dos indivíduos foram consideradas palavras, em um caso as características relacionadas ao tempo de viagem e em outro as geográficas. A combinação do tópico mais relevante de cada modelo foi utilizada para estabelecer o perfil espaço-temporal de cada usuário, gerando 16 grupos. Os grupos foram utilizados para as análises de objetivo, padrões e receitas por meio de análises estatísticas.</p> | Classificação        | Latent Dirichlet Allocation (LDA) | Não supervisionado | Matlab     | -       |
| 16 | <a href="#">Chew et al. (2018)</a> | <p>Criação de um modelo preditivo para classificar os seguidores de @knowtherealcast em jovem (13-17 anos), jovens adultos (18 a 24 anos) e adultos (25 anos e acima) utilizando metadados dos usuários e comportamento de postagens. A base rotulada foi montada a partir das últimas 200 postagens dos usuários do Twitter para os quais foi possível identificar a idade em postagens relacionadas ao aniversário.</p>   | Classificação        | Gradient Boosting                 | Supervisionado     | -          | -       |

| N. | Referência                                    | Aplicação   | Aplicação codificada | Métodos  | Tipo  | Lingua-gem | Pacotes           |
|----|---|---|----------------------|--|---|------------|-------------------|
| 17 | <a href="#">Cox, Lechner e Bollens (2023)</a> | Considerando que a avaliação buscar verificar os efeitos do programas entre aqueles que ainda não receberam tratamento, uma das três capacitações, o primeiro passo foi usar um estimador baseado em post-LASSO para identificar um pseudo início do programa para os não-participantes e eliminar os que nunca serão tratados ou tratados mais de uma vez. Modelos baseados em Modified Causal Forests são treinados para identificar os efeitos da participação nos programas de capacitação contra não participar em nenhum. São reportados efeitos médio de tratamento (ATE), efeitos médios de tratamento nos tratados (ATE <sub>T</sub> ), efeitos médios de tratamento individualizados (IATE). A fim contribuir na verificação de padrões de heterogeneidade, o IATE dos 3 programas são clusterizados utilizando K-means. Finalmente, são avaliadas estratégias diferentes de atribuição aos programas: aleatória, caixa preta com base nos IATEs calculados e com base em uma árvore de decisão respeitando a capacidade dos programas. | Inferência causal    | Least Absolute Shrinkage and Selection Operator (LASSO); Modified Causal Forest (MCF); K-means | Supervisionado (LASSO, MCF); Não supervisionado (K-Means) | Python     | Pacote Python mcf |



| N. | Referência             | Aplicação   | Aplicação codificada | Métodos   | Tipo           | Lingua-gem | Pacotes   |
|----|------------------------|---|----------------------|---|----------------|------------|---|
| 18 | Colangelo (2021)       | Capítulo 1: O objetivo é avaliar a transferência de conhecimento na estimativa do efeito médio de tratamento do programa. A estratégia foi selecionar duas variáveis de resultado, remuneração no terceiro ano após início do programa e percentual de semanas empregado no mesmo ano, criar artificialmente a ausência de valores para a de resultado alvo e usar a de resultado fonte para transferência de aprendizado. Os algoritmos de aprendizado usados no método Double Debiased Machine Learning (DML) são LASSO e Deep Neural Network. Capítulo 3: O objetivo do estimador avaliado, chamado Double Debiased Machine Learning (DML), é a função média de reposta à dose de tratamento, sendo a resposta a proporção de semanas empregado no segundo ano após o programa. A variável contínua de tratamento é o total de horas de treinamento acadêmico e vocacional no primeiro ano. O DML é implementado de cinco formas, cada um com um algoritmo de aprendizado de máquina para estimar a função de expectativa condicional e densidade condicional: LASSO, Random Forest, Generalized Random Forest, Deep Neural Network e K Deep Neural Network. Os 5 estimadores apresentaram padrões semelhantes de resultado. | Inferência causal    | Least Absolute Shrinkage and Selection Operator (LASSO); Random Forest (RF); Generalized Random Forest (GRF); Deep Neural Network (DNN); K Deep Neural Network (KNN); Double Machine Learning (DML) | Supervisionado | Python; R  | Pacote sklearn; python pytorch; Pacote python numpy; python pandas; Pacote python rpy2; Pacote python scipy; Pacote R grf |
| 19 | Colangelo e Lee (2020) | O objetivo do estimador proposto, chamado Double Debiased Machine Learning (DML), é a função média de reposta à dose de tratamento. Para isso, aplica modelos de aprendizado máquina para estimar a função de expectativa condicional e densidade condicional. O estimador é aplicado nos dados do programa com 3 algoritmos de aprendizado de máquina diferentes, Lasso, GRF e DNN, a fim de verificar o desempenho dos mesmos. Os 3 estimadores apresentaram padrões semelhantes de resultado.  | Inferência causal    | Least Absolute Shrinkage and Selection Operator (LASSO); Generalized Random Forest (GRF); Deep Neural Network (DNN); Double Machine Learning (DML)  | Supervisionado | Python; R  | Pacote sklearn; python pytorch; Pacote python numpy; python pandas; Pacote python rpy2; Pacote python scipy; Pacote R grf |

| N. | Referência                           | Aplicação  | Aplicação codificada            | Métodos   | Tipo               | Lingua-gem | Pacotes                                     |
|----|--------------------------------------|--|---------------------------------|---|--------------------|------------|---|
| 20 | Dai et al. (2022)                    | Os sete modelos foram treinados e o desempenho comparado com o objetivo de verificar qual o melhor para predição da concentração do poluente por hora. As variáveis de entrada incluíam uma variável de tempo como tendência, nove variáveis relacionadas a meteorologia e quatro relacionadas ao instante de tempo. O modelo que obteve melhor desempenho foi Random Forest. Foi criado um modelo por poluente e para cada uma das 57 localidades de Fenwei Plain. Os autores então aplicaram um método de normalização utilizando predição e variáveis do modelo Random Forest para possibilitar a remoção dos efeitos das mudanças meteorológicas nas concentrações dos poluentes, a fim de verificar as mudanças que são de fato decorrentes das emissões. | Previsão                        | k-Nearest Neighbors (KNN); Random Forest (RF); Extreme Gradient Boosting (XGB); Gradient Boosted Decision Trees (GBDT); AdaBoost Regression (ABR); Support Vector Regression (SVR); Multilayer Perceptron (MLP) | Supervisionado     | Python; R  | Pacote R worldMet;<br>Pacote Python sklearn |
| 21 | Duthil, Imoussaten e Montmain (2017) | Intuitivamente, o objetivo é verificar a importância que cada município financiado pelo governo francês para desenvolver ações relacionadas a desenvolvimento sustentável dá a cada critério. Para isso, os relatórios coletados de cada município passar um modelo baseado no método Synopsi para extração de segmentos relacionados a cada um dos critérios de política. Na sequência, um método fusão de dados imprecisos é utilizado para definir um escore de importância de cada critério por município.   | Descoberta de termos relevantes | Synopsis  | Não supervisionado | -          | -   |
| 22 | Francke e Acosta (2021)              | O aprendizado de máquina foi utilizado para gerar um grupo de contrafactual, vez que todas as crianças nessa idade devem ser beneficiárias do programa. Para isso, foi utilizado o modelo baseado em Gradient Boosting Modelo combinado com balanceamento por entropia para geração da amostra do grupo de controle. Para estimação dos efeitos foram aplicados métodos econométricos de estimação <i>probit</i> e <i>OLS</i> .  | Previsão                        | Gradient Boosting (GBM)   | Supervisionado     | R          | Pacote R twang                              |

| N. | Referência            | Aplicação  | Aplicação codificada                           | Métodos   | Tipo   | Lingua-gem | Pacotes |
|----|-----------------------|--|--|---|--|------------|---------|
| 23 | Gao, Li e Wang (2021) | Um modelo LDA2Vec é treinado sobre os textos relacionados a ciência e tecnologia, colaboração e inovação, previamente processados. Esse modelo é aplicado sobre a política pública avaliada e gera os termos relevantes de texto que serão atribuídos para o índice de avaliação. Uma Rede Neural Convolutional é treinada para classificar os indicadores (0 ou 1) da política com base nos textos. Finalmente os indicadores são agregados e uma nota final é atribuída que pode ser traduzida em faixas de avaliação da política que vai de insuficiente a uma política perfeita.   | Descoberta de termos relevantes; Classificação | LDA2Vec; Convolutional Neural Network                                       | Não supervisionado (LDA2Vec); Supervisionado (CNN) | Python     | -       |
| 24 | Goller et al. (2020a) | A fim de avaliar o efeito médio de tratamento sobre os tratados (ATE) é utilizado o estimador Radius matching on the propensity score with bias adjustment (RMBA). A variável de resultado é a quantidade de dias empregados nos 3 anos após o tratamento. Os dados são relativos aos grupos de tratamento e controle com 14.817 e 261,820 indivíduos, respectivamente, contendo 37 covariáveis de características dos indivíduos, histórico de trabalho, situação de moradia e estrutura dos centros de trabalho responsáveis por eles. Os modelos de aprendizado de máquina são utilizados para estimar o escore de propensão a ser aplicado no estimador. São comparados modelos baseados em Random Forest e LASSO. Adicionalmente os resultados são comparados com um modelo estatístico Probit. | Balanciamento de covariáveis                   | Least Absolute Shrinkage and Selection Operator (LASSO); Random Forest (RF) | Supervisionado                                     | -          | -       |

| N. | Referência                           | Aplicação  | Aplicação codificada | Métodos  | Tipo           | Lingua-gem | Pacotes   |
|----|--------------------------------------|--|----------------------|--|----------------|------------|---|
| 25 | <a href="#">Gori et al. (2019)</a>   | A utilização de aprendizado se dá na primeira etapa da abordagem para subsidiar uma modelagem hidrológica e hidráulica com softwares especializados. Nessa primeira etapa foi treinado um modelo com base em ANN-MLP com o objetivo de prever a transição de cada setor da bacia hidrográfica de Cypress Creek da situação de não desenvolvido para desenvolvido. A seleção dos preditores se deu por testes de poder explanatório potencial pelo método de Cramer's V e, após o treinamento inicial do modelo, com a remoção de forma iterativa daqueles com baixo poder preditivo. As etapas seguintes envolveram a criação de um modelo baseado em cadeia de Markov para construção de cenários de desenvolvimento para 2050 e aplicação dos softwares de modelagem hidrológica e hidráulica para análise da elevação da superfície de água e dos limites da planície de inundação de 100 anos. | Previsão             | Multilayer Perceptron (MLP)                            | Supervisionado | -          | -   |
| 26 | <a href="#">Grecov et al. (2022)</a> | A aplicação objetiva avaliar os efeitos do aumento na emissão de licença de comercialização de bebidas alcoólicas no atendimento de serviços médicos relacionados a intoxicação por álcool reportados em um estado da Austrália. Os autores desenvolveram um método chamado Deep Probabilistic Counterfactual Prediction Net (DeepProbCP) que utiliza o algoritmo de aprendizado de máquina Long Short-Term Memory Recurrent Neural Network com funções de quantis condicionais para aprender um modelo global a partir da série temporal de dados e estimar os efeitos causais produzidos em um grupo de múltiplas unidades tratadas. O resultado do método é uma função probabilística de previsão por quantis baseada em curva spline que permite estimar o efeito médio de tratamento (ATE) por quantil.   | Inferência causal    | Long Short-Term Memory Recurrent Neural Network (LSTM) | Supervisionado | Python; R  | Pacote Python tensorflow; Pacote Python scipy; Pacote R CausalImpact; Pacote R ArCo |

| N. | Referência                              | Aplicação  | Aplicação codificada             | Métodos  | Tipo           | Linguagem | Pacotes                 |
|----|---|--|----------------------------------|--|----------------|-----------|-------------------------|
| 27 | <a href="#">Guo et al. (2022)</a>       | Um modelo foi treinado com Random Forest para predição da concentração de 6 poluentes com variáveis como tempo, dia da semana, velocidade do vento, direção, pressão, temperatura, humidade. Com uma técnica de reamostragem das variáveis de tempo e meteorológicas por cidade, os dados foram utilizados no modelo para calcular a concentração de poluentes removendo a influência das variações meteorológicas. Dessa forma, foi possível comparar o observado com o estimado.   | Previsão                         | Random Forest  | Supervisionado | R         | Pacote R normalweatherr |
| 28 | <a href="#">Ha e Grubert (2023)</a>     | Uma amostra de comentários foram classificados manualmente em favorável, não favorável e outros. Adicionalmente, foram classificados em tópicos de discussão (trabalho, clima, justiça, etc). Foram aplicadas técnicas de aumento da base de dados usando substituição de sinônimos. Foi treinado um modelo CNN para classificar quanto ao suporte ou não do Plano, rótulo único. É um modelo para classificar quanto aos assuntos abordados no comentários, portanto, múltiplos rótulos. Isso permitiu análises descritivas relacionadas ao suporte e relação com tópico e inter-tópicos.   | Análise de sentimento            | Convolutional Neural Network (CNN)   | Supervisionado | -         | -                       |
| 29 | <a href="#">Hatamyar e Kreif (2023)</a> | Objetivo é aprender regras que maximizem o resultado da política de seguro de saúde subsidiado para mães considerando a redução na mortalidade de bebês no ano seguinte ao nascimento. As variáveis incluem dados das mães e da família. São testadas duas classes de políticas, uma que pesquisa pelas árvores de tratamento buscando maximizar o resultado, chamada exhaustive tree search, e outra que indica o subgrupo com menor efeito médio de tratamento condicional (CATE), chamada de plug-in policies. Os algoritmos testados em cada uma são Bayesian Additive Regression Trees, Double-Robust Machine Learning usando Random Forest e variações, Causal Forest e variações. | Otimização de regras de alocação | Bayesian Additive Regression Trees (BART); Double Machine Learning (DML); Random Forest (RF); Causal Forest (CF) | Supervisionado | R         | Pacote R grf            |

| N. | Referência            | Aplicação   | Aplicação codificada | Métodos   | Tipo           | Lingua-<br>gem | Pacotes  |
|----|-----------------------|---|----------------------|---|----------------|----------------|--|
| 30 | Heiler e Knaus (2021) | A abordagem apresentada permite a utilização de modelos de aprendizado de máquina para estimativa dos parâmetros de incômodo, resultados e escores de propensão, como um dos passos. No entanto, foi utilizado um Ensemble de Random Forest, LASSO e Ridge Regression. Como resultado do novo método são produzidas duas métricas, natural conditional average treatment effect (nATE) e random conditional average treatment effect (rATE), que, analisadas em conjunto permitem aferir o resultado decorrente da heterogeneidade dos tratamentos. | Inferência causal    | Least Absolute Shrinkage and Selection Operator (LASSO); Random Forest (RF); Ridge Regression | Supervisionado | R              | Pacote R causalDML; Pacote R sas7bdat  |
| 31 | Hu e Huang (2018)     | O impacto causal é estimado pela diferença no índice de mercado (stock index) entre o valor real observado e o valor predito do contrafactual derivado da aplicação do modelo de aprendizado de máquina Random Forest. Para inferência é utilizado o modelo estatístico Bayesian Structural Time Series Model (BSTS).   | Previsão             | Random Forest   | Supervisionado | Python         | Pacote Python sklearn  |
| 32 | Kalhus e Zhou (2019)  | Um modelo foi treinado com Generalized Random Forests para aprender o Efeito de Tratamento Médio Condicional (CATE) e inferência dos limites de Taxa de Positivo-Verdadeiro, Taxa de Falso-Negativo e curvas ROC e xROC para análise de efeitos heterogêneos nos grupos por idade, sexo e nacionalidade.  | Inferência causal    | Generalized Random Forest   | Supervisionado | Python; R      | Pacote Python sklearn; Pacote Python scipy; Pacote Python matplotlib; Pacote Python gurobipy; Pacote R grf |

| N. | Referência                   | Aplicação   | Aplicação codificada                                | Métodos   | Tipo           | Linguagem | Pacotes  |
|----|------------------------------|---|---|---|----------------|-----------|--|
| 33 | <a href="#">Knaus (2022)</a> | <p>A avaliação dos programas foi realizada considerando a quantidade de meses empregado acumulados nos 31 meses subsequentes ao início do programa. Os dados contêm 45 variáveis relacionadas a características dos indivíduos desempregados, do assistente social, do processo de aplicação, do emprego anterior e indicadores econômicos regionais. Na abordagem de Double Machine Learning foi utilizada o algoritmo de Random Forest para estimação dos parâmetros de incômodo (resultado e score de propensão) com todas as variáveis para cada um dos grupos de tratamento (todos os programas e os sem programa). São estimados resultado médio do programa (APO), efeito médio de tratamento (ATE), efeito médio de tratamento sobre os tratados (ATET), efeito médio de tratamento por grupo (GATE), efeito médio de tratamento individualizado (IATE). O Random Forest também é utilizado no cálculo do IATE por como componente do métodos DR-Learner e NDR-Learner, esse último proposto neste estudo. Finalmente são estimadas árvores de decisão a fim de inferir regras ótimas de tratamento considerando os diferentes tipos de tratamento.</p> | Inferência causal; Otimização de regras de alocação | Random Forest; Decision Tree; Double Machine Learning (DML) | Supervisionado | R         | Pacote R causalDML; Pacote R grf; Pacote R mp; Pacote R crs; Pacote R polycytree |

| N. | Referência                               | Aplicação   | Aplicação codificada  | Métodos   | Tipo           | Lingua-<br>gem | Pacotes  |
|----|--|---|-----------------------|---|----------------|----------------|--|
| 34 | <a href="#">Kreif e DiazOrdaz (2019)</a> | O objetivo é estimar efeito médio de tratamento (ATE) sob a suposição de nenhum confundidor não observado. A variável de tratamento se refere a ser segurado ou não pelo plano de saúde, enquanto que a de resultado é a probabilidade de o nascimento ser realizado por um profissional de saúde. As covariáveis incluem características da mãe e de moradia. São aplicados eXtreme Gradient Boosting; Super Learner tendo como base Random Forest, BART e Generalized Boosting Model. SuperLearner também é utilizado como base da abordagem Double Machine Learning. LASSO é utilizado nas abordagens Double Lasso e Collaborative Targeted Minimum Loss based Estimation. | Inferência causal     | Extreme Gradient Boosting; Super Learner; Random Forest; Bayesian Additive Regression Trees (BART); Generalized Boosted Model; Least Absolute Shrinkage and Selection Operator (LASSO); Double Machine Learning (DML) | Supervisionado | R              | Pacote R glmnet; Pacote R caret; Pacote R ranger; Pacote R gbm; Pacote R bartMachine; Pacote R dbarts; Pacote R SuperLearner; Pacote R h2oEnsemble; Pacote R subsemble; Pacote R twang; Pacote R Matching; Pacote R xgboost; Pacote R tmle |
| 35 | <a href="#">Kreif et al. (2022b)</a>     | Avaliação do impacto dos programas (Average Treatment Effect, Average Treatment Effect on the Controls), em comparação com o grupo de controle, com a utilização do algoritmo Generalized Random Forest. Análise da associação da importância das covariáveis com os efeitos heterogêneos observados, visando identificar subgrupos com efeitos distintos a partir do efeitos médios de tratamento condicional (CATE).  | Inferência causal     | Generalized Random Forest (GRF)   | Supervisionado | R              | Pacote R grf   |
| 36 | <a href="#">Kumar e Sharma (2020)</a>    | A partir da coleta de tweets relacionados à divulgação da lei e no mês seguinte à sua publicação foram treinados cinco modelos utilizando diferentes algoritmos para classificação da postagem em positivo, neutro e negativo. Ao final, o modelo SVM obteve melhor desempenho e foi utilizado. A avaliação da política com base na opinião foi realizada pela verificação dos percentuais de postagens em cada tipo de sentimento.   | Análise de sentimento | Naive Bayes (NB); Support Vector Machine (SVM); Multilayer Perceptron (MLP); k-Nearest Neighbors (kNN); Decision Tree (DT)  | Supervisionado | Python         | -  |



| N. | Referência               | Aplicação  | Aplicação codificada                            | Métodos                             | Tipo               | Lingua-gem | Pacotes                                  |
|----|--------------------------|--|---|-------------------------------------|--------------------|------------|--|
| 37 | Li et al. (2021)         | Com o objetivo de subsidiar a análise quantitativa dos documentos foi criado um modelo de análise baseado em Policy Modeling Consistency (PMC-Index) que consiste em indicadores de primeiro e segundo nível relacionados a critérios desejados nas normas. Os indicadores são avaliados em ruim, aceitável, bom e excelente. A nota da política é a consolidação de todos conforme equação definida para avaliação. Foi utilizado um modelo de aprendizado de máquina Latent Dirichlet Allocation para descoberta de grupos e tópicos dos documentos para subsidiar a criação dos indicadores. Ao final 5 políticas foram selecionadas para serem avaliadas pelo modelo PMC-Index contendo 9 indicadores de primeiro nível e 36 de segundo nível. | Descoberta de termos relevantes                 | Latent Dirichlet Allocation (LDA)   | Não supervisionado | Python     | Software ROSTCM;<br>Software Ucinet      |
| 38 | Linden e Yarnold (2016a) | Análise do balanceamento das covariáveis dos grupos de tratamento e controle após aplicação de uma técnica de pareamento e análise dos efeitos na variável de resultado.   | Balanceamento de covariáveis; Inferência causal | Optimal Discriminant Analysis (ODA) | Supervisionado     | Stata      | Software UniODA                          |
| 39 | Linden e Yarnold (2018)  | Em dois momentos. Primeiro para avaliar o balanceamento entre a unidade de tratamento e o grupo de controle na série temporal de resultado pré-intervenção. Em seguida, na estimação do efeito de tratamento, ou seja, a diferença entre a unidade de tratamento e o grupo de controle na série temporal de resultado pós-intervenção. A variável de resultado é a quantidade de pacotes de cigarros per capita de 1970 a 2000, sendo 1989 o primeiro ano da intervenção.  | Balanceamento de covariáveis; Inferência causal | Optimal Discriminant Analysis (ODA) | Supervisionado     | Stata      | Pacote Stata ITSA-MATCH; Software UniODA |

| N. | Referência            | Aplicação  | Aplicação codificada    | Métodos  | Tipo   | Lingua-gem | Pacotes  |
|----|-----------------------|--|-------------------------|--|--|------------|--|
| 40 | Lu et al. (2022)      | O aprendizado de máquina foi utilizado para classificar a cobertura da terra das sete reservas naturais objeto da avaliação. Foi utilizado um modelo de aprendizado profundo baseado em U-Net, treinado a partir de características extraídas de imagens de satélite do Google Earth e histórico de pesquisas de campos de anos anteriores. Os tipos de cobertura da terra foram analisados usando GIS para caracterizar a dinâmica da paisagem de mangue e sua conversão em outras coberturas, e a eficácia de conservação das reservas foi finalmente avaliada quantitativamente com base em métricas de paisagem e no modelo de peso de entropia. | Classificação           | U-Net Convolutional Neural Network   | Supervisionado   | -          | Software ArcGIS; Software Fragstats                                  |
| 41 | Ma, Chen e Lei (2022) | Predição do percentual de decréscimo na taxa de crescimento de casos de conluio vertical e do percentual de decréscimo na variância a fim de avaliar as mudanças na quantidade de casos de conluio vertical após adoção de regulamento no período de 2005 a 2020   | Previsão                | BP Neural Network Optimized by Annealing Algorithm                             | Supervisionado   | -          | -  |
| 42 | Ma e Zhang (2014)     | O algoritmo AHC é utilizado para descobrir as estruturas entre os agentes (estudantes, professores, administrativos) por meio de uma métrica definida como uma função do tempo de interação entre eles. A segunda parte avalia o impacto do montante do financiamento por meio do modelo hierárquico de aprendizado por reforço definido com uma política de recompensa que é uma função do desempenho dos alunos dado pelo Texas Assessment of Knowledge and Skills - TAKS.   | Classificação; Previsão | Agglomerative Hierarchical Clustering (AHC); MAXQ Value Function Decomposition | Não supervisionado (AHC); Aprendizado por reforço (MAXQ) | Python     | Pacote Python fastcluster; Pacote Python matplotlib; Software NodeXL |

| N. | Referência  | Aplicação  | Aplicação codificada | Métodos  | Tipo               | Lingua-gem | Pacotes |
|----|---|--|----------------------|--|--------------------|------------|---------|
| 43 | <a href="#">Matenga (2022)</a>                      | <p>Os indicadores relacionados ao ODS 7 obtidos do Banco Mundial foram utilizados para clusterização dos países utilizando o algoritmo k-Means. Foram gerados 8 grupos. A fim de permitir analisar sobre perspectivas de acesso e qualidade. Além do agrupamento inicial, foram gerados outras duas versões usando pesos para os indicadores relacionados a cada uma dessas perspectivas. Para cada versão, os grupos foram ordenados de acordo com a magnitude dos centroides, sendo de 0,1 e 2 considerados nível baixo; 3, 4 e 5, nível médio; e 6 e 7 nível alto.</p>  | Classificação        | K-means  | Não supervisionado | -          | -       |
| 44 | <a href="#">McFowland, So-manchi e Neill (2023)</a> | <p>O método proposto, Treatment Effect Subset Scan (TESS), visa identificar subpopulações que tenham efeitos de tratamento heterogêneos significantes em experimento aleatório, maximizando o estimador proposto efeitos médios de tratamento não-paramétrico (NATE). Intuitivamente, o objetivo é identificar subgrupos com efeitos anômalos. Um dos passos envolve a envolve o cômputo de uma distribuição de resultado condicional para cada perfil de covariáveis. Nesse passo é utilizado o algoritmo de Random Forest com adaptação dos seus pesos. A avaliação considerou como variável de resultado a soma das notas de matemática e leitura. Os grupos de tratamento e controle foram considerados, respectivamente, alunos em sala com professor auxiliar e sem professor auxiliar. Como resultado, observou-se que a subpopulação de alunos do segundo e terceiro ano, de escolas urbanas ou próximas da cidade, com professores com 10 ou mais anos de experiência tiveram efeito significante maior que a média. A literatura sempre apontou que não havia melhora com o uso de professores auxiliares, a aplicação aponta que para um grupo pode ter sido diferencial.</p> | Inferência causal    | Random Forest; Treatment Effect Subset Scan (TESS) | Supervisionado     | -          | -       |

| N. | Referência              | Aplicação   | Aplicação codificada                              | Métodos   | Tipo           | Lingua-gem | Pacotes                   |
|----|-------------------------|---|---|---|----------------|------------|---------------------------|
| 45 | Pan et al. (2017)       | Um modelo baseado em Random Forest foi treinado para predição do auto-reporte de recebimento de resultado de teste de HIV. Em uma primeira etapa, o modelo foi treinado com a exclusão de registros com dados faltante e algumas variáveis para evitar a multicolinearidade, tendo como objetivo comparar o desempenho com um modelo de regressão logística. O Random Forest teve melhor acurácia em todos os cenários testados com variação do tamanho da amostra. Na segunda etapa, o modelo foi treinado com todas as variáveis e registros usando imputação dos valores faltantes que, no caso, também foram imputados por um modelo de Random Forest. Finalmente, a importância das variáveis, incluindo interação entre elas, foram analisadas. | Análise dos preditores                            | Random Forest (RF)  | Supervisionado | R          | Pacote R random-ForestSRC |
| 46 | Rana et al. (2022)      | Treinamento de um modelo ensemble combinando algoritmos de Random Forest, Naive Bayes e Extreme Gradient Boosting para predição de perda de cobertura de árvores em polígonos de plantação. Com isso, as probabilidades de perda de cobertura de árvores preditas para as áreas em avaliação (0 a 100%) foram utilizadas para projetar o desperdício no estado de Himachal Pradesh com base produzido com os recursos aplicados para plantio nessas mesmas áreas e com o orçamento previsto até 2030.   | Previsão  | Extreme Gradient Boosting; Random Forest; Naive Bayes   | Supervisionado | -          | -                         |
| 47 | Ratledge et al. (2021a) | Predição dos registros faltantes da variável dependente (índice de meios de sustento) por meio do treinamento da CNN com imagens de satélite. Utilização de MC e SC-EM para estimar o efeito causal da expansão da malha de energia elétrica no índice de meios de sustento.  | Imputação de valores faltantes; Inferência causal | Convolutional Neural Network (CNN); Matrix Completion (MC); Synthetic Controls with Elastic Net (SC-EN) | Supervisionado | R          | -                         |

| N. | Referência            | Aplicação   | Aplicação codificada                                   | Métodos  | Tipo                  | Linguagem | Pacotes   |
|----|-----------------------|---|--|--|-----------------------|-----------|---|
| 48 | Ray et al. (2018)     | <p>CoreNLP é utilizado para definição da classe gramatical dos termos dos tweets. Então, os termos são passados pelo SentiWordNet para detecção da polaridade positiva, neutra ou negativa dos termos. Uma fórmula é aplicada para definir a polaridade final de cada tweet. As frequências então são analisadas por cidade, data, grupos de posição e oposição. A diferença entre grupos é testada via ANOVA. Os modelos de aprendizado de máquina baseados em Naive Bayes e SVM foram treinados em um dicionário pré-classificado para classificar em sentimentos como raiva, antecipação, nojo, medo, alegria, tristeza e surpresa. SVM obteve melhor desempenho em acurácia.</p>  | <p>Análise de sentimento</p>                           | <p>Naive Bayes (NB); Support Vector Machine (SVM)</p>  | <p>Supervisionado</p> | <p>R</p>  | <p>Pacote R TwttrR;<br/>Pacote R ggplot2;<br/>Pacote R NLP; Pacote R Tm; Pacote R e1071; SentiWordNet; Pacote R coreNLP</p> |
| 49 | Rudolph et al. (2022) | <p>O objetivo da aplicação é demonstrar violações práticas de positividade. Em uma primeira etapa, os autores demonstram como a estimativa do modelo de exposição no nível errado pode induzir violações práticas de positividade analisando o escore de propensão em dois níveis de agregação, estado e condado. Para isso foi utilizado um estimador chamado Super Learner a partir de um ensemble de três modelos de aprendizado de máquina Least Absolute Shrinkage and Selection Operator, Generalized Linear Models e Multiple Additive Regression Splines. Como modelo alternativo para não violação, propõem estimar o efeito de o estado adiar a promulgação da lei por um ano em cada estado. Também foi utilizado o estimador Super Learner com os mesmos modelos.</p> | <p>Balanciamento de covariáveis; Inferência causal</p> | <p>Least Absolute Shrinkage and Selection Operator (LASSO); Generalized Linear Models (GLM); Multiple Additive Regression Splines (MARS)</p> | <p>Supervisionado</p> | <p>R</p>  | <p>Pacote R lmtip</p>   |

| N. | Referência                 | Aplicação   | Aplicação codificada   | Métodos  | Tipo           | Lingua-<br>gem | Pacotes   |
|----|----------------------------|---|------------------------|--|----------------|----------------|---|
| 50 | Shah, Kreif e Jones (2021) | A fim de demonstrar a aplicação de modelos de aprendizado de máquina baseados em Meta-Learners (T-, S- e X-Learners), R-Learner e Causal Forest, os autores realizam a avaliação de efeitos do programa na utilização de serviços de cuidado saúde. O tratamento equivale a estar coberto pelo plano de saúde, com um grupo 475.930 indivíduos, e o controle a não estar coberto com 436.882. A variável de resultado é a quantidade de dias de estadia no hospital no ano anterior. São estimados efeito individual de tratamento (ITE), efeito médio de tratamento (ATE), efeito médio de tratamento nos tratados (ATT) efeito médio de tratamento no controle (ATC), efeito médio de tratamento condicional (CATE), efeito médio de tratamento por grupo (GATE). | Inferência causal      | Meta-Learners (ML); R-Learner (RL); Causal Forest (CF)   | Supervisionado | R              | Pacote R grf; Pacote R causalToolbox; Pacote R rlearner |
| 51 | Sharma e Shekhar (2021)    | Foram treinados cinco modelos de aprendizado de máquina usando diferentes técnicas para classificação de sentimento (positivo, negativo ou neutro) das publicações do Twitter: Support Vector Machine, Logistic Regression, Random Forest, XGBoost e Multilayer Perceptron. O pré-processamento do texto empregou TF-IDF com Particle Swarm Optimization. O modelo que obteve melhor desempenho foi XGBoost. Os dados, incluindo a classificação gerada, foram submetidos a uma análise de textos no Software NVivo para codificação quanto os objetivos do programa.   | Análise de sentimento  | Support Vector Machine (SVM); Logistic Regression (LR); Random Forest (RF); Extreme Gradient Boosting (XGBoost); Multilayer Perceptron (MLP) | Supervisionado | Python         | Software NVivo  |
| 52 | Martins et al. (2021)      | Regressão LASSO foi utilizada para identificar subconjuntos de cláusulas mais informativas para cada período analisado. Na sequência, Random Forest foi aplicado para identificar as cláusulas com com maior poder preditivo e para investigar a interação entre cláusulas.   | Análise dos preditores | Least Absolute Shrinkage and Selection Operator (LASSO); Random Forest   | Supervisionado | R              | -   |

| N. | Referência                                     | Aplicação  | Aplicação codificada                              | Métodos                 | Tipo           | Lingua-gem | Pacotes                            |
|----|--|--|---|-------------------------|----------------|------------|------------------------------------|
| 53 | <a href="#">Singh, Xu e Gretton (2020)</a>     | São propostos estimadores causais baseados em adaptação do modelo de regressão de kernel ridge para estimação não paramétrica de funções causais. A avaliação de política se de com a aplicação dos estimadores no programa Job Corps. O tratamento contínuo é definido pelo total de horas em capacitação acadêmica e vocacional no primeiro ano e o resultado contínuo é a proporção de semanas empregado no ano seguinte. São estimadas as curvas de resposta da dose, de resposta incremental e de resposta heterogênea com relação à idade, além de efeitos discretos de tratamento.  | Inferência causal                                 | Kernel Ridge Regression | Supervisionado | Python; R  | Pacote R npcausal;<br>Pacote R grf |
| 54 | <a href="#">Stetter, Mennig e Sauer (2022)</a> | Os autores aplicaram Causal Forest com a finalidade de estimar os efeitos dos esquemas agroambientais. Inicialmente é aplicado um modelo de aprendizado de máquina para uso do método Recursive Partitioning para imputação de valores, porém o autor não indica qual algoritmo especificamente. Foi treinado um modelo para cada uma das quatro variáveis de resultado, sendo índice Gini-Simpson como proxy para biodiversidade, emissões GHG como proxy para clima, intensidade de fertilizante e de pesticida, ambas como proxy para água/solo. São reportadas as distribuições de efeito de tratamento para cada uma. Para identificar os impulsadores dos efeitos estimados no nível das fazendas, os autores analisam os valores Shapley dos preditores em comparação com efeitos médios de tratamento condicionais (CATE). | Inferência causal; Imputação de valores faltantes | Causal Forest (CF)      | Supervisionado | R          | Pacote R IML                       |

| N. | Referência                     | Aplicação  | Aplicação codificada             | Métodos  | Tipo           | Lingua-gem | Pacotes |
|----|--------------------------------|--|----------------------------------|--|----------------|------------|---------|
| 55 | <a href="#">Valente (2023)</a> | São treinados modelos de aprendizado de máquina baseados em Causal Forest com o objetivo de identificar o efeito do valor de preços de resíduos (há 45 faixas de valores entre 1 e 18 centavos de euros por litro de resíduo não classificado) na quantidade de resíduos reciclados, não classificados e total per capita. São criados modelos por cada uma dessas 3 variáveis de resultado e por ano. São reportados Conditional Average Price Effect, Conditional Average Treatment Effect e Average Price Effect. Os modelos também são utilizados para avaliar o efeito nos custos municipais decorrentes da gestão de resíduos. São estimados os efeitos na demanda e custos municipais para o caso de todos os municípios implementarem a política de preços. Finalmente, os efeitos médios de tratamento são comparados com parados com método R-Learner utilizando LASSO considerando uma variável de tratamento contínua. | Inferência causal                | Least Absolute Shrinkage and Selection Operator (LASSO); Causal Forest (CF); R-Learner | Supervisionado | R          | -       |
| 56 | <a href="#">Viviano (2023)</a> | Os potenciais beneficiários eram submetidos a sessões informativas consecutivas sobre o seguro, sendo que o grupo de tratamento recebia mais informações sobre os benefícios. Para a avaliação, foi considerado que a rede de relacionamentos de toda a população não é observada. O autor propõe um novo método, chamado Network Empirical Welfare Maximization (NEWEM) para estimar o bem-estar como uma função da política utilizando estimadores arbitrários, incluindo baseados em aprendizado de máquina. Nessa aplicação, é utilizado o LASSO. O método estima uma política (regras baseadas nas covariáveis dos beneficiários)   | Otimização de regras de alocação | Least Absolute Shrinkage and Selection Operator (LASSO)                                | Supervisionado | -          | -       |



| N. | Referência                         | Aplicação   | Aplicação codificada           | Métodos       | Tipo           | Lingua-gem | Pacotes |
|----|------------------------------------|---|--------------------------------|---------------|----------------|------------|---------|
| 57 | <a href="#">Wang et al. (2020)</a> | <p>Treinamento de modelo baseado no Random Forest para preenchimento dos valores faltantes de Profundidade Ótica de Aerossóis (AOD), feitas a cada hora do dia, devido a cobertura por nuvens. Com isso, buscou-se eliminar vieses para o cálculo da média diária de AOD. Outro Random Forest foi treinado para estimar a concentração média mensal de PM2.5. Esse modelo foi treinado usando como uma das variáveis o AOD, enquanto outro modelo foi treinado sem AOD a fim de estimar o impacto do preenchimento dos valores faltantes na média. Usando a média anual de concentração de PM2.5 a partir dos valores estimados pelo modelo, foi utilizada regressão linear para avaliar os efeitos das políticas implementadas ao longo do tempo em Fujian e Taiwan.</p>                   | Imputação de valores faltantes | Random Forest | Supervisionado | -          | -       |
| 58 | <a href="#">Wang et al. (2021)</a> | <p>Um modelo de rede neural Conv2LSTM, uma técnica regularmente aplicada em análise de vídeo, é treinada para extrair parâmetros sobre a dinâmica de infecção da Covid-19 a partir de mapas de infecção reportados por instituto alemão. Os mapas diários referentes a 30 dias são submetidos ao modelo para treinamento gerando parâmetros de infecção relativos a taxa de transferência entre localidades, taxa de infecção, taxa de diagnóstico e população não diagnosticada. Esses parâmetros são utilizados para gerar simulações de infecção futura em um modelo baseado em autômato celular. A avaliação das políticas de lockdown restrito, lockdown atenuado e sem restrição de lockdown, se deu com a variação do parâmetro de taxa de transferência no modelo de simulação.</p> | Previsão                       | Conv2LSTM     | Supervisionado | -          | -       |

| N. | Referência                         | Aplicação  | Aplicação codificada                     | Métodos  | Tipo           | Lingua-gem | Pacotes |
|----|------------------------------------|--|--|--|----------------|------------|---------|
| 59 | <a href="#">Wozny (2022)</a>       | Random Forest foi utilizado para preencher dados faltantes das empresas aéreas da base ICAO. A imputação de valores foi feita de forma iterativa, variável a variável, pela ordem de variáveis com menor quantidade de dados faltantes. Dados que haviam mais de 100 variáveis no dataset, foi utilizada eliminação recursiva considerando conjuntos cada vez menores de variáveis, resultando em 11 variáveis para treinamento ao final. O modelo treinado para predição das tarifas aéreas anuais foi o XGBoost. Para cada ano foram utilizados os dados reais dos anos anteriores e otimização bayesiana para seleção de hiperparâmetros. Assim, para 2020 o modelo foi treinado sem conhecer os efeitos da Covid-19 e para efeito de tratamento foi considerada a diferença entre o valor predito e o valor observado. | Imputação de valores faltantes; Previsão | Random Forest; Extreme Gradient Boosting (XGBoost) | Supervisionado | -          | -       |
| 60 | <a href="#">Yang et al. (2023)</a> | Um modelo baseado em rede neural artificial (FF-ANN) é treinado com dados mensais de condições macroeconômicas, ambiente, estrutura de mercado de eletricidade e desenvolvimento corrente de energia renovável para prever a produção anual de eletricidade a partir de recursos renováveis e capacidade de geração de energia renovável (duas saídas). Um modelo é gerado para cada uma das 3 regiões avaliadas com determinantes mais adequados às políticas em vigor em cada região. Esses dados são utilizados em um modelo de difusão tecnológica para prever cenários de tendência na transição para energia renovável, divididos em linha de base, ótimo para o processo e um de acordo com os parâmetros atuais da política.   | Previsão                                 | Feedforward Artificial Neural Network (FF-ANN)     | Supervisionado | -          | -       |

| N. | Referência          | Aplicação  | Aplicação codificada   | Métodos   | Tipo           | Lingua-gem | Pacotes                                  |
|----|---------------------|--|------------------------|---|----------------|------------|--|
| 61 | Yin et al. (2018)   | Modelos anuais para o período de 2000 a 2014 empregando Random Forest foram treinados com dados de satélite da região da Mongólia para classificar cada píxel em quatro classes de uso e cobertura de terra: terra de cultivo, pastagem, área florestal e outros. Em seguida, foi aplicada uma abordagem de detecção de mudança baseada em trajetória com base em uma versão modificada da detecção baseada em Landsat de tendências em distúrbios e algoritmo de recuperação (LandTrendr). Finalmente foram realizadas análises à luz da localidade foco de cada um dos três programas. | Classificação          | Random Forest   | Supervisionado | R          | Software TIME-SAT; Pacote R randomForest |
| 62 | Zare et al. (2021)  | Os dados contém o IMC dos estudantes na educação infantil e no quarto ano. Os modelos de aprendizado de máquina foram treinados para prever o estado de obesidade ( $IMC >= 95$ percentil) no quarto ano. O desempenho também foi comparado com um modelo de regressão logística.  | Análise dos preditores | Decision Tree; Artificial Neural Network; Random Forest | Supervisionado | R          | Pacote R caret; Pacote R randomForest    |
| 63 | Zhang et al. (2021) | Treinamento de um modelo baseado em Random Forest para predição da concentração média diária de PM2.5 em cada estação de monitoramento. Foram utilizadas 40 preditores, incluindo variáveis de estação do ano, uma variável dicotômica indicando se o período era anterior (2014-2015) ou posterior (2016-2018) à mudança da política, profundidade ótica de aerossóis, fatores meteorológicos, população, uso de terra, emissão de substância, econômicos, entre outros.  | Previsão               | Random Forest   | Supervisionado | R          | -  |

| N. | Referência                             | Aplicação   | Aplicação codificada | Métodos                         | Tipo           | Lingua-gem | Pacotes      |
|----|--|---|----------------------|---------------------------------|----------------|------------|--------------|
| 64 | <a href="#">Zhang, Li e Ren (2022)</a> | O algoritmo de Generalized Random Forest é utilizado para estimar efeito médio de tratamento (ATE), efeito médio de tratamento nos tratados (ATE <sub>T</sub> ) e efeito médio de tratamento nos não tratados (ATU). As variáveis de resultado avaliadas foram colisões anuais com ferimentos pessoais e colisões graves e fatais. Ademais, investigou-se os efeitos heterogêneos a partir dos resultados do modelo (CATE). | Inferência causal    | Generalized Random Forest (GRF) | Supervisionado | R          | Pacote R grf |

# APÊNDICE D – Estudos eliminados na elegibilidade

Quadro 12 – Estudos eliminados na elegibilidade

| N. | Referência                                | Título   | Justificativa de exclusão  | Evidenciação  |
|----|---|--|--|---|
| 1  | <a href="#">Amarasinghe et al. (2023)</a> | Explainable Machine Learning for Public Policy: Use Cases, Gaps, and Research Directions   | O foco do estudo são tarefas administrativas, ou seja, de implementação. Nesse sentido, não traz nenhuma aplicação de avaliação.   | <i>Among the broad range of intervention points that the domain of policy presents to ML (e.g., policy design, evaluation, and administration), we focus our attention on policy administration tasks where predictive ML models are used to support human decisions with objectives of improving the efficiency of resource usage, the effectiveness of interventions, and equity of outcomes. Typically, these systems are designed to improve the efficiency of resource utilization, intervention effectiveness, and equity of outcomes</i>   |
| 2  | <a href="#">Andimi et al. (2018)</a>      | Targeting with machine learning: An application to a tax rebate program in Italy           | O objetivo é demonstrar que uma atribuição ao tratamento usando aprendizado de máquina seria mais eficiente que a que foi usada no programa. Consideramos como uma nova proposta hipotética e não avaliação com uso de ML. | <i>This paper shows how machine learning (ML) methods can be used to improve the effectiveness of public schemes and inform policy decisions. Focusing on a massive tax rebate scheme introduced in Italy in 2014, it shows that the effectiveness of the program would have significantly increased if the beneficiaries had been selected according to a transparent and easily interpretable ML algorithm. [...] Using the Bank of Italy's Survey on House hold Income and Wealth (SHIW), we start by showing that reporting difficulties in making ends meet - which we use as a proxy for being consumption constrained - is associated with a larger effect of the bonus on consumption. We therefore use this variable, observable only in the survey, as outcome in our targeting exercise.</i> |
| 3  | <a href="#">Andimi et al. (2022)</a>      | Machine learning in the service of policy targeting: The case of public credit guarantees  | Propõe um novo método de seleção de empresas beneficiárias do programa com base em aprendizado de máquina, a avaliação não envolve aprendizado.  | <i>We use Machine Learning (ML) predictive tools to propose a policy-assignment rule designed to increase the effectiveness of public guarantee programs. This rule can be used as a benchmark to improve targeting in order to reach the stated policy goals</i>   |
| 4  | <a href="#">Stoffi et al. (2019)</a>      | Estimating Heterogeneous Causal Effects in the Presence of Irregular Assignment Mechanisms | Duplicado. Trata-se de uma versão anterior do artigo analisado e complementado, conforme citado na introdução de BargagliGnecco2020.   | <i>(STOFFI; GNECCO, 2020)</i>   |

| N. | Referência                                   | Título   | Justificativa de exclusão  | Evidenciação  |
|----|--|--|--|---|
| 5  | <a href="#">Beitel (2005)</a>                | Applying Artificial Intelligence Data Mining Tools to the Challenges of Program Evaluation | Avaliação de um programa de ensino no âmbito de uma escola.  | <i>The superintendent, like most educational leaders, asked a predictive question regarding the efficacy of a newly implemented reading program in the lower elementary grades.</i>   |
| 6  | <a href="#">Chen (2018)</a>                  | Evaluating Conditional Cash Transfer Policies with Machine Learning Methods                | Apesar de usar dados de uma política públicas, trata-se de uma comparação da capacidade preditiva de modelos de aprendizado de máquina com a predição de um modelo econométrico.   | <i>We have compared the out-of-sample forecast between major machine learning models and the structural econometric model using the Progress-social experiment in this paper. The prediction performance of the machine learning models, in terms of the MAE and RMSE, is better than that found in the previous structural model used by Todd and Wolpin (2006)</i>  |
| 7  | <a href="#">Cortá, Jukic e Decman (2016)</a> | Open data for human development in Mexico: A data science perspective                      | Aborda conceitos relacionados a dados abertos e ciências de dados e como poderiam ser aplicados para aprimoramento das políticas de forma geral  | <i>The article includes: 1) a brief review on OD currently available in agencies; 2) a series of general recommendations to analyze and to model these data by applying statistical and machine-learning methods that are frequently available in free or in low-cost software tools; and 3) a proposal on the application of these methods on policies that are focused on HD.</i>   |
| 8  | <a href="#">Blasio e Dignazio (2021)</a>     | Regional policy in the machine learning era: Opportunities and challenges                  | Os autores apresentam um caso hipotético de atribuição de tratamento por meio de aprendizado de máquina e comparam com o de uma política de crédito. A aplicação de ML está na suposta implementação. Baseado em outros estudos que também foram excluídos pela mesma razão ( <a href="#">ANDINI et al., 2018</a> ; <a href="#">ANDINI et al., 2022</a> ). | <i>To illustrate the benefits of ML targeting for regional policy, we build on Andini et al. (2019) and move from firm-level targeting to a geographical targeting at the local labor system (LLS) level. Our aim is to provide guidance on which LLSs should be targeted by policies to support access to credit. We use the ML targeting rule from Andini et al. (2019) and compute, for the universe of Italian limited companies, the probability that a firm is credit-constrained, and the probability that a firm is credit-worthy.[...] We then investigate whether the ML-based index described above might provide guidance for the policy-maker in order to identify those areas where public credit guarantees would be more effective.</i> |

| N. | Referência               | Título  | Justificativa de exclusão   | Evidenciação   |
|----|--------------------------|---|---|--|
| 9  | Furtado e Andrade (2022) | Machine Learning Simulates Agent-Based Model Towards Policy   | Emulação de um simulador existente de políticas de habitação para estender os resultados para outras regiões.   | <p>PS2 is defined as a primarily endogenous computational agent-based model (ABM) that includes mortgage loans, housing construction, taxes collection and investments, with firms and households interacting in real estate, goods and services, and labor market. The main goal of PS2 is to contrast and compare three competing policy investments alternatives for the case of Brasília. Given low-income households, municipalities may either (a) acquire houses to transfer (Property acquisition); (b) provide rental vouchers (Rent vouchers); or (c) make monetary transfers (Monetary aid). PS2, however, is developed and evaluated for the case of Brasília Metropolitan Region alone. Although the authors also run the model and present comparisons to four other MRs, results were not exhaustively run for all of the 46 available MRs. This is exactly how the surrogate model we present contribute to the analysis. We use the bulk of runs of the sensitivity analysis to expand the results to all 46 MRs, while also expanding the variability contained in the data better exploring the parameter space, and steering results towards optimal policies.</p> |
| 10 | Hatami et al. (2023)     | Non-Linear Associations Between the Urban Built Environment and Commuting Modal Split: A Random Forest Approach and SHAP Evaluation | Avalia o impacto do ambiente construído nos modos de transporte, não é de uma política específica.              | <p>In this paper, we evaluate the impact of built environment variables on commuting modal split, including the four modes of public transit-bus, public transit-rail, walking, and driving.</p>   |
| 11 | Hoffman (2018)           | Essays in Public Economics  | Os dados são uma simulação de choques econômicos a fim de produzir resultados que subsidiem avaliações futuras. | <p>In Section 2.3, we describe our data and the Census updating shock. The sample is a panel of roughly 3,000 U.S. counties from 1987 to 2015, and our outcomes of interest are per capita growth rates in a variety of amenities. Census shocks are abrupt, unanticipated changes to county population estimates generated by a new decennial census. These changes have economic consequences because they affect federal spending (such as block grants and staffing at federal agencies) through funding formulas that depend on local population, and these spending changes in turn affect local income and employment. We follow the methodology of Suarez-Villa and Wingender (2016) in constructing the shocks and, in Section 2.4, in using inverse propensity score weighting and a treatment effects framework to analyze their amenity spillovers.</p>  |

| N. | Referência  | Título  | Justificativa de exclusão   | Evidenciação  |
|----|---|---|---|---|
| 12 | <a href="#">Hozouri, Dong e Freedman (2020)</a>         | Advancing National Health Through Investigations of Nutrition, Medical Expenditure, and Public Insurance Programs | Dos 3 capítulos da tese, o segundo utiliza aprendizado de máquina. Contudo não é para avaliação, mas para prever despesas em saúde. | <i>The Medical Expenditure Panel Survey (MEPS) is a rich data source on American health care consumption over the last two decades. Its Household Component consists medical expenses, sources of payment, types of medical expenses and diagnoses, and demographic information on panel members. [...] In this paper, I review a variety of ML-based techniques which could be applied to data similar to the the Medical Expenditure Panel Survey. My goal is to assess which models perform best compared to the realized data in regard to out-of-sample prediction performance.</i>  |
| 13 | <a href="#">Huang, Hsiang e Gonzalez-Navarro (2021)</a> | Using Satellite Imagery and Deep Learning to Evaluate the Impact of Anti-Poverty Programs                         | Programa privado de uma organização não-governamental dos EUA no Quênia.  | <i>GiveDirectly, a US charity, implemented a randomized controlled trial of unconditional cash transfers to rural households via mobile money, using as sole eligibility criterion whether the household lived under a thatched roof (a low quality roof material that served as a simple meanstest).</i>   |
| 14 | <a href="#">Kallus (2022)</a>                           | Treatment Effect Risk: Bounds and Inference - 2022 ACM Conference on Fairness, Accountability, and Transparency   | Avaliação de uma mudança "hipotética" em um programa.   | <i>Studying a hypothetical change to French job-search counseling services, our bounds and inference demonstrate a small social benefit entails a negative impact on a substantial subpopulation. [...] We consider a hypothetical scenario where the private-run counseling program (A=0) is currently being offered to the unemployed and we consider the change to a public-run program (A=1). 10 We take reemployment within six months as our (binary) outcome.</i>  |
| 15 | <a href="#">Martire (2022)</a>                          | Sickness Absence Among Patients With Chronic Pain in Swedish Specialist Healthcare                                | Criação de modelo para predição de dor crônica, não envolve avaliação.  | <i>Chronic pain beyond three months is a global public health problem. Every third adult suffers from a chronic pain condition, resulting in a socioeconomic burden that corresponds to 3-10% of gross domestic product in western economies. [...] Simultaneously, a detailed overview of sickness absence (SA) associated with chronic pain is complicated by incongruent classification due to conflicting perspectives on the condition as either a symptom or a disease in its own right. [...] The aims were addressed in three register-based studies using microdata from five Swedish national registers. Study I used sequence analysis to describe SA in 44,241 patients over a 7-year period and subsequently developed a machine learning-based model to predict chronic pain-related SA in the final two years.</i> |



| N. | Referência                       | Título   | Justificativa de exclusão  | Evidenciação   |
|----|----------------------------------|--|--|--|
| 16 | López-Andreu et al. (2021)       | Sentinel-2 images and machine learning as tool for monitoring of the common agricultural policy: Calasparra rice as a case study | Modelo para predição de áreas de plantio com o objetivo de apoiar a gestão e monitoramento.  | <i>Based on the new requirements for the CBM of CAP subsidies, this work aims to create a methodology that allows for the identification of rice crops through the use of S2 data and classification techniques based on ML. Considering this, the farmer's obligation to the payment scheme, as monitored in this work, is to carry out the routine work required for rice cultivation and for the rice to reach the flowering stage. Therefore, the management, in this way of the Paying Agencies aid, will be more productive and efficient.</i>   |
| 17 | Nijkamp, Burch e Vindigni (2002) | A Comparative Institutional Evaluation of Public-Private Partnerships in Dutch Urban Land-use and Revitalisation Projects        | Método baseado em rough set, não é aprendizado de máquina  | <i>(KOMOROWSKI et al., 1999)</i>   |
| 18 | Ratledge et al. (2022)           | Using machine learning to assess the livelihood impact of electricity access   | Duplicado. Uma versão mais extensa já está publicada no ArXiv e foi catalogada no projeto.   | <i>(RATLEDGE et al., 2021a)</i>  |
| 19 | Srivastava (2022)                | Challenges for evaluation practices and innovative approaches: Lessons during COVID-19   | Estudo fala sobre desafios para a avaliação durante a pandemia, principalmente pela dificuldade de acesso a dados primários e secundários. Uma das formas de superar seria o uso de inteligência artificial, porém não aborda casos concretos.                     | <i>3.1. Use of Artificial Intelligence (AI) for evaluation. Often, we think of evaluation only in terms of how well AI systems perform, yet it is vital to all stages of research, from early conceptualization to retrospective analyses of series of programs (Cohen &amp; Howe, 1988). AI supportive database will be useful in situations like pandemic or inaccessible terrains...</i>  |
| 20 | Sun et al. (2023)                | Research on the Policy Analysis of Sustainable Energy Based on Policy Knowledge Graph Technology: A Case Study in China          | Graph machine learning é mencionado três vezes no texto: abstract, introdução e conclusão. Contudo, consideramos que os métodos apresentados não de aprendizado de máquina especificamente medidas de centralidade, pagerank e detecção de comunidade com Louvain. | <i>The establishment of the policy knowledge graph based on publicly-open official documents can facilitate the analysis and visualization of sustainable energy policies, providing new ideas for policy research. This paper introduces the knowledge graph, graph machine learning algorithms and big data technology, which can deepen the depth and breadth of peoples research on sustainable energy policy.</i>   |
| 21 | Wang e Sun (2021)                | Reducing Bus Bunching with Asynchronous Multi-Agent Reinforcement Learning   | Propõe um modelo para otimizar o controle de tempo de permanência do tráfego de ônibus. Uso de ML para implementação, a avaliação do modelo não envolve ML.  | <i>We establish an asynchronous multi-agent reinforcement learning (ASMR) framework to optimize holding control policy on a bus route to reduce bus bunching. [...] The key idea of holding control is to let a bus stay longer (i.e., by adding a slack period) at stops in addition to the dwell time for passengers to board/alight. [...] In this paper, we develop an asynchronous multi-agent re-inforcement learning model to improve dynamic bus control and reduce bus bunching. In particular, we propose CAACA credit assignment framework to address the asynchronous issue in multi-agent control problems, which is often over-looked in previous studies.</i> |

| N. | Referência               | Título   | Justificativa de exclusão  | Evidenciação   |
|----|--------------------------|--|--|--|
| 22 | Yang e Guo (2021)        | Research on Diagnostic Test and Treatment for Higher Education System  | Apenas Cadeia de Markov  | Secondly, by using Markov Chain Model for prediction, the periodic changing rule of German educational reform is gained  |
| 23 | Du et al. (2022)         | Leveraging Computational Tools to Inform Public Policy Process for Social Good   | Um capítulo é sobre simulação para entender barreiras dentro de organizações (glass ceiling). O segundo é uma avaliação sobre uma intervenção baseada em seleção algorítmica hipotética de crianças para creche. O último traz uma proposta de alocação de jovens em um programa de capacitação. | First, I utilize an agent-based model to support employment identification of the glass ceiling effect. Second, I utilize forensic social science and a data-driven system dynamic model to support employment of screen-in decisions made in the U.S. foster care system with respect to various goals. Finally, I propose to use a machine learning model to support implementation of policy. To be more specific, I formulate dynamic service allocation in foster care system in U.S. to a novel setup of stochastic multi-armed bandit problem and propose three service allocation strategies to solve it.  |
| 24 | Zheng, Zheng e Ye (2016) | Using machine learning in environmental tax reform assessment for sustainable development: A case study of Hubei Province, China | O autor utiliza o método estatístico Synthetic Controls Method. A aplicação de ML não ficou clara apesar de mencionada várias vezes.   | In this section, we attempt to reveal the causal effect of EFT reform by imputing the missing potential emission intensity reductions, which is the counterfactual outcome that would have prevailed if cities had not implemented the EFT reform. For this purpose, we invoke the synthetic control method, first developed by Abadie and Gardeazabal [29] and then extended by Abadie et al. [31]. [...] For Equation (6), we used the quadratic programming, which is a typical optimization method in machine learning [30], in order to optimize the minimization process of the loss function of the weights. The machine learning method in the synthetic control method enables us to evaluate the model not only with big size of samples, but also with small ones by constructing a virtual synthetic match groups. |

# APÊNDICE E – Novos estudos decorrentes de exploração das citações

Quadro 13 – Novos estudos decorrentes de exploração das citações

| N. | Referência                            | Título   | Citado em                       | Aplicação de (Citado em)                            |
|----|---------------------------------------|--|---------------------------------|---|
| 1  | Athey, Imbens e Wager (2018)          | Approximate Residual Balancing: Debiased Inference of Average Treatment Effects in High Dimensions   | Knaus (2022)                    | Inferência causal; Otimização de regras de alocação |
| 2  | Athey e Wager (2020)                  | Policy Learning with Observational Data  | Colangelo e Lee (2020)          | Inferência causal                                   |
| 3  | Baiardi e Naghi (2021)                | The Value Added of Machine Learning to Causal Inference: Evidence from Revisited Studies   | Knaus (2022)                    | Inferência causal; Otimização de regras de alocação |
| 4  | Bandara et al. (2020)                 | Towards Accurate Predictions and Causal What-if Analyses for Planning and Policy-making: A Case Study in Emergency Medical Services Demand   | Grecov et al. (2022)            | Inferência causal                                   |
| 5  | Belloni, Chernozhukov e Hansen (2014) | High-Dimensional Methods and Inference on Structural and Treatment Effects   | Zhang, Li e Ren (2022)          | Inferência causal                                   |
| 6  | Bertrand et al. (2017)                | Contemporaneous and Post-Program Impacts of a Public Works Program   | Knaus (2022)                    | Inferência causal; Otimização de regras de alocação |
| 7  | Bodory, Busshoff e Lechner (2022)     | High Resolution Treatment Effects Estimation: Uncovering Effect Heterogeneities with the Modified Causal Forest                              | Cockx, Lechner e Bollens (2023) | Inferência causal                                   |
| 8  | Buhl-Wiggers et al. (2022)            | Some children left behind: Variation in the effects of an educational intervention   | Heiler e Knaus (2021)           | Inferência causal                                   |
| 9  | Burlig et al. (2020)                  | Machine Learning from Schools about Energy Efficiency  | Wozny (2022)                    | Imputação de valores faltantes; Previsão            |
| 10 | Chen e Hsiang (2019)                  | Causal Random Forests Model Using Instrumental Variable Quantile Regression  | Shah, Kreif e Jones (2021)      | Inferência causal                                   |
| 11 | Chen, Huang e Tien (2021)             | Debiased/Double Machine Learning for Instrumental Variable Quantile Regressions  | Shah, Kreif e Jones (2021)      | Inferência causal                                   |
| 12 | Chernozhukov et al. (2017)            | Double/Debiased/Neyman Machine Learning of Treatment Effects   | Colangelo (2021)                | Inferência causal                                   |
| 13 | Chernozhukov et al. (2018)            | Double/debiased machine learning for treatment and structural parameters   | Colangelo (2021)                | Inferência causal                                   |
| 14 | Chernozhukov et al. (2023)            | Generic Machine Learning Inference on Heterogenous Treatment Effects in Randomized Experiments, with an Application to Immunization in India | Knaus (2022)                    | Inferência causal; Otimização de regras de alocação |
| 15 | Chernozhukov, Newey e Singh (2022a)   | Automatic Debiased Machine Learning of Causal and Structural Effects   | Colangelo e Lee (2020)          | Inferência causal                                   |
| 16 | Chernozhukov, Newey e Singh (2022b)   | Debiased machine learning of global and local parameters using regularized Riesz representers  | Colangelo e Lee (2020)          | Inferência causal                                   |
| 17 | Chin, Kahn e Moon (2020)              | Estimating the Gains from New Rail Transit Investment: A Machine Learning Tree Approach  | Chen et al. (2021)              | Inferência causal                                   |

| N. | Referência                           | Título   | Citado em                          | Aplicação de (Citado em)                            |
|----|--------------------------------------|--|------------------------------------|---|
| 18 | Christensen et al. (2023)            | Decomposing the Wedge between Projected and Realized Returns in Energy Efficiency Programs   | Valente (2023)                     | Inferência causal                                   |
| 19 | Cockx, Lechner e Bollens (2023)      | Priority to unemployed immigrants? A causal machine learning evaluation of training in Belgium   | Knaus (2022)                       | Inferência causal; Otimização de regras de alocação |
| 20 | Daoud e Johansson (2019)             | Estimating Treatment Heterogeneity of International Monetary Fund Programs on Child Poverty with Generalized Random Forest                             | Shah, Kreif e Jones (2021)         | Inferência causal                                   |
| 21 | Davis e Heller (2017)                | Using Causal Forests to Predict Treatment Heterogeneity: An Application to Summer Jobs   | Kallus e Zhou (2019)               | Inferência causal                                   |
| 22 | Davis e Heller (2020)                | Rethinking the Benefits of Youth Employment Programs: The Heterogeneous Effects of Summer Jobs   | Knaus (2022)                       | Inferência causal; Otimização de regras de alocação |
| 23 | Dueñas et al. (2021)                 | Assessing the Impact of COVID-19 on Trade: a Machine Learning Counterfactual Analysis  | Wozny (2022)                       | Imputação de valores faltantes; Previsão            |
| 24 | Farrell (2015)                       | Robust inference on average treatment effects with possibly more covariates than observations  | Colangelo e Lee (2020)             | Inferência causal                                   |
| 25 | Goller et al. (2020b)                | Does the estimation of the propensity score by machine learning improve matching estimation? The case of Germany's programmes for long term unemployed | Zhang, Li e Ren (2022)             | Inferência causal                                   |
| 26 | Grange e Carslaw (2019)              | Using meteorological normalisation to detect interventions in air quality time series  | Dai et al. (2022)                  | Previsão  |
| 27 | Greco et al. (2021)                  | Causal Inference Using Global Forecasting Models for Counterfactual Prediction   | Greco et al. (2022)                | Inferência causal                                   |
| 28 | Halvorsen et al. (2016)              | Reducing Subway Crowding: Analysis of an Off-Peak Discount Experiment in Hong Kong   | Chen e Zhou (2022)                 | Classificação                                       |
| 29 | Hill (2011)                          | Bayesian Nonparametric Modeling for Causal Inference   | McFowland, Somanchi e Neill (2023) | Inferência causal                                   |
| 30 | Hill, Weiss e Zhai (2011)            | Challenges With Propensity Score Strategies in a High-Dimensional Setting and a Potential Alternative  | Goller et al. (2020a)              | Balanceamento de covariáveis                        |
| 31 | Hsu et al. (2022)                    | Testing Monotonicity of Mean Potential Outcomes in a Continuous Treatment with High-Dimensional Data   | Colangelo e Lee (2020)             | Inferência causal                                   |
| 32 | Huber et al. (2020)                  | Direct and indirect effects of continuous treatments based on generalized propensity score weighting   | Colangelo e Lee (2020)             | Inferência causal                                   |
| 33 | Imai e Ratkovic (2013)               | Estimating treatment effect heterogeneity in randomized program evaluation   | Zhang, Li e Ren (2022)             | Inferência causal                                   |
| 34 | Jarvis, Deschenes e Jha (2022)       | The Private and External Costs of Germany's Nuclear Phase-Out  | Abrell, Kosch e Rausch (2022)      | Previsão  |
| 35 | Kallus, Mao e Uehara (2022)          | Localized Debiased Machine Learning: Efficient Inference on Quantile Treatment Effects and Beyond  | Knaus (2022)                       | Inferência causal; Otimização de regras de alocação |
| 36 | Knaus, Lechner e Strittmatter (2020) | Machine learning estimation of heterogeneous causal effects: Empirical Monte Carlo evidence  | Knaus (2022)                       | Inferência causal; Otimização de regras de alocação |

| N. | Referência   | Título   | Citado em                                       | Aplicação de (Citado em)                            |
|----|--|--|---|---|
| 37 | <a href="#">Knaus, Lechner e Strittmatter (2022)</a> | Heterogeneous Employment Effects of Job Search Programs  | <a href="#">Knaus (2022)</a>                    | Inferência causal; Otimização de regras de alocação |
| 38 | <a href="#">Kreiff et al. (2020)</a>                 | Who benefits from health insurance? : Uncovering heterogeneous policy impacts using causal machine learning.                                   | <a href="#">Chen et al. (2021)</a>              | Inferência causal                                   |
| 39 | <a href="#">Kube, Das e Fowler (2019)</a>            | Allocating Interventions Based on Predicted Outcomes: A Case Study on Homelssness Services   | <a href="#">Kallus e Zhou (2019)</a>            | Inferência causal                                   |
| 40 | <a href="#">Lechner (2019)</a>                       | Modified Causal Forests for Estimating Heterogeneous Causal Effects  | <a href="#">Goller et al. (2020a)</a>           | Balanciamento de covariáveis                        |
| 41 | <a href="#">Lechner e Mareckova (2022)</a>           | Modified Causal Forest   | <a href="#">Cockx, Lechner e Bollens (2023)</a> | Inferência causal                                   |
| 42 | <a href="#">Linden e Yarmold (2016b)</a>             | Using data mining techniques to characterize participation in observational studies  | <a href="#">Linden e Yarmold (2016a)</a>        | Balanciamento de covariáveis; Inferência causal     |
| 43 | <a href="#">Linden e Yarmold (2016c)</a>             | Using machine learning to assess covariate balance in matching studies   | <a href="#">Linden e Yarmold (2018)</a>         | Balanciamento de covariáveis; Inferência causal     |
| 44 | <a href="#">Linden e Yarmold (2016d)</a>             | Using machine learning to identify structural breaks in singlegroup interrupted time series designs  | <a href="#">Linden e Yarmold (2018)</a>         | Balanciamento de covariáveis; Inferência causal     |
| 45 | <a href="#">Louizos et al. (2017)</a>                | Causal Effect Inference with Deep Latent-Variable Models   | <a href="#">Stetter, Mennig e Sauer (2022)</a>  | Inferência causal; Imputação de valores faltantes   |
| 46 | <a href="#">Lovrić et al. (2021)</a>                 | Understanding the true effects of the COVID-19 lockdown on air pollution by means of machine learning  | <a href="#">Dai et al. (2022)</a>               | Previsão  |
| 47 | <a href="#">Ma et al. (2020)</a>                     | Behavioral response to promotion-based public transport demand management: Longitudinal analysis and implications for optimal promotion design | <a href="#">Chen e Zhou (2022)</a>              | Classificação                                       |
| 48 | <a href="#">McGaffrey, Ridgeway e Morral (2004)</a>  | Propensity score estimation with boosted regression for evaluating causal effects in observational studies                                     | <a href="#">Chang (2020)</a>                    | Balanciamento de covariáveis; Inferência causal     |
| 49 | <a href="#">Miller (2020)</a>                        | Causal forest estimation of heterogeneous and time-varying environmental policy effects  | <a href="#">Zhang, Li e Ren (2022)</a>          | Inferência causal                                   |
| 50 | <a href="#">Mullainathan e Spiess (2017)</a>         | Machine Learning: An Applied Econometric Approach  | <a href="#">Wozny (2022)</a>                    | Imputação de valores faltantes; Previsão            |
| 51 | <a href="#">Mullally e Chakravarty (2018)</a>        | Are matching funds for smallholder irrigation money well spent?  | <a href="#">Stetter, Mennig e Sauer (2022)</a>  | Inferência causal; Imputação de valores faltantes   |
| 52 | <a href="#">O'Neill e Weeks (2019)</a>               | Causal Tree Estimation of Heterogeneous Household Response to Time-Of-Use Electricity Pricing Schemes  | <a href="#">Rana et al. (2022)</a>              | Previsão  |
| 53 | <a href="#">Poulos e Zeng (2021)</a>                 | RNN-Based Counterfactual Prediction, With an Application to Homestead Policy and Public Schooling  | <a href="#">Grecov et al. (2022)</a>            | Inferência causal                                   |
| 54 | <a href="#">Prest (2020)</a>                         | Peaking interest: How awareness drives the effectiveness of time-of-use electricity pricing  | <a href="#">Valente (2023)</a>                  | Inferência causal                                   |

| N. | Referência                        | Título   | Citado em                  | Aplicação de (Citado em)                            |
|----|-----------------------------------|--|----------------------------|---|
| 55 | Qu et al. (2020)                  | Evaluating the meteorological normalized PM2.5 trend (20142019) in the 2+26 region of China using an ensemble learning technique                       | Dai et al. (2022)          | Previsão  |
| 56 | Rana e Miller (2019)              | Machine learning to analyze the social-ecological impacts of natural resource policy: insights from community forest management in the Indian Himalaya | Rana et al. (2022)         | Previsão  |
| 57 | Sasaki, Ura e Zhang (2022)        | Unconditional Quantile Regression with High Dimensional Data   | Colangelo e Lee (2020)     | Inferência causal                                   |
| 58 | Shalit, Johansson e Sontag (2017) | Estimating individual treatment effect: generalization bounds and algorithms   | Shah, Kreif e Jones (2021) | Inferência causal                                   |
| 59 | Shi et al. (2021)                 | Abrupt but smaller than expected changes in surface air quality attributable to COVID-19 lockdowns   | Guo et al. (2022)          | Previsão  |
| 60 | Steinkraus e others (2019)        | Estimating Treatment Effects With Artificial Neural NetsA Comparison to Synthetic Control Method   | Grecov et al. (2022)       | Inferência causal                                   |
| 61 | Strittmatter (2021)               | What Is the Value Added by Using Causal Machine Learning Methods in a Welfare Experiment Evaluation?   | Knaus (2022)               | Inferência causal; Otimização de regras de alocação |
| 62 | Su et al. (2012)                  | Facilitating Score and Causal Inference Trees for Large Observational Studies  | Stoffi e Gnecco (2020)     | Inferência causal                                   |
| 63 | Toth e Laan (2016)                | TMLE for marginal structural models based on an instrument   | Kreif e DiazOrdaz (2019)   | Inferência causal                                   |
| 64 | Vu et al. (2019)                  | Assessing the impact of clean air action on air quality trends in Beijing using a machine learning technique   | Dai et al. (2022)          | Previsão  |
| 65 | Zhang et al. (2020)               | Significant Changes in Chemistry of Fine Particles in Wintertime Beijing from 2007 to 2017: Impact of Clean Air Actions                                | Guo et al. (2022)          | Previsão  |