revista Fronteiras – estudos midiáticos
24(3):47-61 setembro/dezembro 2022
Unisinos – doi: 10.4013/fem.2022.242.05

Codificação de vieses no processo de modelagem algorítmica: formas de opacidade e obscurecimento a partir do estudo de caso da base de dados *Boston Housing*

Coded biases in algorithmic modeling: forms of opacity and obscuration from the case study of the Boston Housing database

Andre Pecini[*]
pecini@gmail.com

Denise Tsunoda[*]
dtsunoda@gmail.com

RESUMO

A multiplicação de atividades mediadas por processos algorítmicos levanta questões sobre seu papel na transformação da sociabilidade, seja sobre na mudança das formas de interação e ação, seja na manutenção de preconceitos e estereótipos, compreendidos como vieses sociais. A opacidade desses processos dificultam o aprofundamento de pesquisas sobre o tema. Este trabalho visa contribuir com a investigação das formas pelas quais esses vieses podem ser codificados e obscurecidos em cada etapa do processo sociotécnico de modelagem algorítmica. O caso em questão é a base de dados *Boston Housing*, cujo atributo "B" carrega não só um viés racial explícito, mas também um processo de obscurecimento. A partir de pesquisas que tratam de vieses sociais – por exemplo, o *redlining* e seus desdobramentos, como a demarcação tecnológica – busca-se articular perspectivas sobre opacidade e obscurecimento com a investigação das diversas formas de codificação de vieses. A fim de ilustrar a dimensão técnica e compor a perspectiva interdisciplinar deste trabalho, descreve-se os principais passos envolvidos no treinamento de um modelo simplificado de regressão linear e o papel do atributo em questão nos resultados. Por fim, é realizada uma revisão de literatura a fim de verificar como este atributo foi tratado na produção acadêmica

ABSTRACT

The multiplication of algorithmically mediated activities raises questions about their role in transforming sociability, whether it be about changing forms of interaction and action, or maintaining prejudices and stereotypes, understood as social biases. The opacity of these processes make it difficult to deepen research on the subject. This work aims to contribute to the investigation of the ways in which these biases can be codified and obscured in each step of the sociotechnical process of algorithmic modeling. The case studied is the Boston Housing database, whose "B" attribute carries not only an explicit racial bias, but also an obscuring process. Based on research dealing with social biases – for example, redlining and its consequences, such as technological demarcation – we seek to articulate perspectives on opacity and obscuration with investigation of the diverse ways of coding biases. In order to illustrate the technical dimension and compose the interdisciplinary perspective of this work, the main steps involved in training a simplified linear regression model and the role of the attribute in question in the results are described. Finally, a literature review is carried out in order to verify

[*] Universidade Federal do Paraná (UFPR). R. Bom Jesus, 650 - Juvevê, Curitiba - PR.

recente. Após este percurso, observa-se que o acesso aos dados e ao treinamento dos modelos permitem abrir o que muitas vezes se apresenta como caixas-pretas, possibilitando a investigação de métodos, decisões e resultados desse processo. O modelo treinado com os dados da base *Boston Housing* ilustra didaticamente o aprendizado de máquina e mostra que o atributo B pode ser até mesmo prescindível. A revisão de literatura mostra que quase a totalidade dos trabalhos produzidos com esta base de dados abdicam de se aprofundar sobre as implicações dos dados nela contidos.

Palavras-chave: vieses algorítmicos; bases de dados; aprendizado de máquina.

how this attribute has been approached in recent academic production. After this path, it is observed that access to data and training of models allow opening what often appears as a black boxes, allowing the investigation of methods, decisions and results of this process. The model trained with data from the Boston Housing database didactically illustrates machine learning and shows that attribute B can even be dispensed. The literature review shows that almost all of the works produced with this database do not delve into the implications of the data contained in it..

Keywords: coded biases; dabatases; machine learning.

Introdução

Ao buscar presentes para as sobrinhas usando as palavras "meninas negras" em 2010, a pesquisadora Safiya Noble se deparou com conteúdo pornográfico nas primeiras posições dos resultados de pesquisa. Embora este problema tenha sido sanado com uma mudança no algoritmo do Google em 2012, ainda afetava outras meninas não brancas, segundo a autora. Em 2015, diante de matérias sobre a possível violação dos termos de uso do Facebook, Brittany Kaiser, na época funcionária da Cambridge Analytica, aponta que a empresa tinha dados de milhões de cidadãos, "os quais incluíam informações do Facebook que tinham em média 570 pontos de dados por pessoa, de mais de 30 milhões de indivíduos" (KAISER, 2020, p. 149).

Esses dois casos identificam duas dimensões centrais da mediação digital contemporânea: algoritmos e dados. Pontos de dados são o conjunto de informações que uma plataforma, uma empresa ou uma instituição tem de cada indivíduo. Algoritmos e os modelos gerados com eles são propriedades das empresas, também inacessíveis. As plataformas digitais oferecem possibilidades antes vedadas às pessoas comuns: audiência, clientela, conexões, alcance. Em contrapartida, ganham a possibilidade de moldar interações – restringir ou permitir, incentivar ou punir – e os próprios atores, em um sentido ontológico, pois estes se tornam perfis cuja "influência, durabilidade e visibilidade" em cada plataforma é diferente (SCHWARZ, 2017). Além de se posicionarem como intermediárias incontornáveis para a atuação em certos setores da sociedade, as plataformas coletam um vasto conjunto de dados sobre interações e interagentes que promovem a atividade por meio de suas infraestruturas. Parte relevante da atividade social e econômica trafega por esse dispositivo que tem em seu cerne algoritmos alimentados por dados (VAN DIJCK; POELL; DE WAAL, 2018).

Um dos principais desafios na pesquisa dos vieses algorítmicos é a opacidade de seus mecanismos. "É impossível saber quando e o que influencia o design privado de algoritmos, além de que seres humanos e que eles não estão submetidos ao debate público, exceto quando nos engajamos em análise crítica e protesto" (NOBLE, 2021, p. 12). A inacessibilidade exige que muitas das pesquisas realizadas sobre o tema sejam baseadas em seus resultados ou *outputs*. Além do caso que abre este artigo, há outros citados por Noble (2021), como a marcação de fotos de afrodescendentes como macacos pelo serviço Google Fotos. E pode-se encontrar exemplos adicionais em publicações sobre o tema: o caso de Tay, o chatbot da Microsoft lançado no Twitter que em questão de horas começou a publicar textos simpáticos ao nazismo (NEFF, 2016), a recomendação de canais com conteúdo não confiável pelo YouTube em 55% dos acessos, constatado em levantamento realizado recentemente (NETLAB UFRJ, 2022), o caso de Catherine Taylor, que descobriu erros em bases de dados pessoais amplamente usadas por empresas quando foi rejeitada para uma vaga de emprego (ONEIL, 2021, pp. 237-8) e os casos de racismo algorítmico listados por Tarcízio Silva (2020, p. 138; 2022).

Um trabalho exemplar do potencial de pesquisas com acesso aos dados é o projeto *Gender Shades*, desenvolvido a partir da base pública *Labeled Faces in the Wild* (LFW), com rostos de celebridades, considerada um padrão de excelência para o reconhecimento facial. A base era composta por 77,5% de rostos de homens e 83,5%, de brancos. Pesquisa realizada com 3 algoritmos de classificação de gênero apontou que todos tinham mais precisão ao classificar pessoas mais claras e homens em

geral e tiveram pior desempenho com mulheres de pele mais escura (BUOLAMWINI; GEBRU, 2018). Outro caso, escolhido para análise no presente artigo, foi a identificação por Michael Carlisle de um atributo com viés racial em uma pequena base de dados chamada *Boston Housing*, usada para testes e ensino de algoritmos. Em texto intitulado "destruição racista de dados?" (CARLISLE, 2019, tradução nossa), o pesquisador denuncia o a aplicação de uma função matemática que impede a recuperação dos dados originais e, ao mesmo tempo, traça níveis de segregação racial que teriam impacto nos preços dos imóveis.

Finalmente, extensa revisão de literatura realizada por Kordzadeh e Ghasemaghaei (2022) sobre o tema constatou que a maioria das publicações não-técnicas sobre vieses algorítmicos se dedica a discutir problemas e necessidades de pessoas afetadas por eles e identificar preocupações e áreas de atenção. Identifica-se, portanto, a utilidade de um estudo interdisciplinar que combina a perspectiva crítica e analítica da implantação dos modelos e algoritmos com o detalhamento técnico dos meandros desse processo.

Partindo da dificuldade de acesso a dados e modelos e beneficiando-se da pesquisa de Carlisle em uma base de dados disponível, este artigo tem o objetivo geral de discutir a codificação de vieses sociais nas etapas da produção de modelos a partir de algoritmos com uma perspectiva interna a esse processo, desde a produção de uma base de dados até o treinamento do modelo. Em outras palavras, abrir o que normalmente se apresenta como uma caixa-preta, nos termos de Latour (2001), para trazer à luz atores que muitas vezes são invisíveis.

Este artigo se apoia em dois conjuntos de perspectivas teóricas, partindo do contexto de opacidade dos sistemas algorítmicos e dos vieses codificados apresentado nesta introdução (SILVA, 2020, 2022; NOBLE, 2021; ONEIL, 2021). O primeiro conjunto trata de diferentes abordagens do processo de modelagem algorítmica (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996; AZEVEDO; SANTOS, 2008) e a codificação de vieses em cada uma das etapas desse processo (SURESH; GUTTAG, 2019; ONEIL, 2021; RUBACK; AVILA; CANTERO, 2021), que compõem a perspectiva sob qual se dá o estudo de caso. A seguir, arregimenta-se referências da ciência da informação na apresentação dos passos envolvidos no treinamento de um modelo, desde a transformação da base de dados e seleção de atributos (CARLISLE, 2019; SHETYE, 2019), com definição de conceitos da mineração de dados (CASTRO; FERRARI, 2016), passando por métricas de aferimento de desempenho (QUININO, REIS, BESSEGATO, 1991; MARTINS, 2018), treinamento (CHEIN, 2019) e implementações usadas como referência para o modelo deste artigo (LEYREZOV, 2017; AGARWAL, 2018; ÇEPEL, 2019).

O percurso metodológico usado para esta abordagem interdisciplinar envolve três etapas. A primeira delas é o desenvolvimento da discussão conceitual descrita acima e sua relação com o estudo de caso da base de dados *Boston housing*, na qual houve a codificação de um viés racial; a segunda é o desenvolvimento de um modelo simplificado de regressão linear, aproveitando a simplicidade da base para ilustrar o processo didaticamente; por fim, faz-se uma revisão sistemática de publicações que mencionam a base de dados na busca de informações sobre os modelos gerados e a percepção dos problemas éticos deste atributo.

Ao examinar detalhadamente a transformação ocorrida na base de dados, discutir o processo de modelagem e apresentar a implementação de um modelo simples, espera-se jogar luz neste processo usualmente opaco e muitas vezes inacessível, e assim contribuir com as pesquisas sobre o tema geral do papel da mediação algorítmica da sociedade contemporânea e fornecer insumos àquelas sobre a codificação de vieses sociais nesse processo. Finalmente, a revisão de literatura objetiva verificar empiricamente o uso da base de dados em publicações revisadas por pares e a atenção dada a este atributo nos trabalhos em questão.

O processo de modelagem algorítmica e a codificação de vieses sociais

As definições de algoritmo englobam uma série de componentes e processos, comportando definições que vão desde as mais técnicas, que os tratam como a formalização de problemas em modelos matemáticos em um vocabulário simbólico específico (BERLINSKI *in* GRILL, 2016), até aquelas que consideram o termo um símbolo de processos sociotécnicos, ou uma sinédoque usada para definir processos complexos de modelagem algorítmica (GILLESPIE, 2016). Os exemplos listados na introdução deste trabalho são resultados da implementação desses modelos. Nesta seção, formalizações desse processo serão articuladas com perspectivas sobre a codificação de vieses em cada uma das etapas a fim de introduzir o estudo de caso que é objeto central deste artigo.

Há mais de uma década, Azevedo e Santos (2008) identificaram o interesse crescente pela mineração de dados levou à multiplicação de esforços para estabelecer padrões na área. O processo de Descoberta de Conhecimento em

Bancos de Dados (*Knowledge Discovery in Databases*, ou KDD) é composto por cinco etapas: 1) Seleção de dados, que inclui a escolha de variáveis e definição do conjunto de dados; 2) pré-processamento, ou a limpeza dos dados a fim de que se tornem consistentes; 3) transformação, com métodos para adaptar os dados às formas de aprendizado desejadas e a redução de dimensionalidade; 4) mineração de dados, ou a busca de padrões nos dados com o uso de algoritmos; 5) interpretação e avaliação dos resultados (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996; AZEVEDO; SANTOS, 2008).

Dois outros métodos mais recentes são o SEMMA – acrônimo para Amostra, Exploração, Modificação, Modelagem e Avaliação (*Sample, Explore, Modify, Model, Assess*) –, desenvolvido pelo SAS Institute e o CRISP-DM (*CRoss-Industry Standard Process for Data Mining* , ou Processo Padrão Multissetorial para Mineração de Dados, em tradução livre), por uma união de esforços de empresas, entre elas a DaimlerChrysler e a SPSS. A análise comparativa desses processos mostrou que, embora sejam compostos por diferentes etapas, Azevedo e Santos (2008) concluem que são idênticos, pois embora tenham denominações diferentes, os passos seguidos em cada um deles envolve as atividades escolhidas para definir os nomes das etapas dos outros.

O objetivo desta breve descrição dos métodos de implantação de sistemas de mineração de dados em empresas foi detalhar o que foi denominado anteriormente por processo de modelagem algorítmica, que consiste em um conjunto de diferentes ações, estratégias, escolhas e esforços combinados, que vão muito além da exposição de um modelo matemático a uma base de dados a fim de que aprendam padrões gerais aplicáveis a dados inéditos, como resumem Suresh e Guttag (2019). Este trabalho se detém principalmente nas etapas de transformação a mineração dos dados, a fim de discuti-lo na condição de um processo sociotécnico.

O trajeto aqui delineado se baseia na noção de obscurecimento, definida por Latour (2001, p. 353) como "a maneira como o trabalho técnico e científico torna-se invisível decorrente de seu próprio êxito". Esta definição será usada de forma bastante limitada à sua aplicação aos objetos técnicos que, nos termos do autor, são caixas-pretas capazes de revelar uma série de entidades que permaneciam invisíveis – ou inacessíveis – quando abertas (LATOUR, 2001, pp. 212-213).

A primeira etapa do KDD é denominada "aprendizado sobre o domínio de aplicação" por Fayyad et. al. (1996, p. 30). Trata-se de obter conhecimento sobre o campo no qual será realizado o desenvolvimento dos modelos e identificação dos objetivos da tarefa. Aqui surge o risco de enviesamento histórico, pois mesmo sob amostragem estatisticamente precisa, os dados podem refletir preconceitos e injustiças (SURESH; GUTTAG, 2019). O enviesamento histórico diz respeito à estrutura social na qual os dados foram gerados e armazenados, que muitas vezes fica obscurecida. Para além da opacidade de fiscalização e responsabilização tratada por Pasquale, Silva (2022, p. 64) argumenta que o racismo algorítmico consiste em "uma camada adicional do racismo estrutural" e identifica uma união das "tradições de ocultação e exploração, tanto nas relações raciais quanto nas decisões ideológicas que definem o que é tecnologia e o que é inovação desejável" (SILVA, 2022, p.14), que denomina de dupla opacidade. Esta perspectiva pode ser estendidas aos vieses sociais, termo usado para dar conta das mais variadas formas de preconceitos, discriminações e estereótipos (FISKE in KORDZADEH; GHASEMAGHAEI, 2022, p. 3). As etapas descritas nesta seção visam discutir o processo pelo qual esses vieses podem ser codificados em sistemas técnicos e, consequentemente, se tornarem obscurecidos.

O segundo tipo de viés é definido por Suresh e Guttag (2019) como viés de representação. Apontam que determinadas populações podem ser significativamente sub-representadas em bases de dados, mesmo que não haja seleção ou filtragem prévia. Mudanças na população entre o treinamento do modelo e sua aplicação podem gerar este tipo de viés e minorias que componham menos de 5% do total das bases podem ser sub-representadas. Relacionado com ele, há o risco do viés de mensuração, especificamente quando se usa dados *proxy*, ou substitutos. O exemplo usado pelos autores é a medição da taxa de encarceramento como *proxy* para taxa de crimes. Citam pesquisas que mostram a relação entre policiamento ostensivo em áreas de moradia de minorias e o maior encarceramento de seus moradores. Em outro caso que corrobora a hipótese, homens jovens afro-americanos e latinos compunham 85% das pessoas que foram objeto de abordagens policiais em Nova Iorque, e mesmo que apenas 0,1% deles estivesse envolvido com crimes violentos, muitos outros eram pegos por infrações menores, ocasionando uma sobrerrepresentação dessa população entre os indivíduos fichados (ONEIL, 2021, pp. 145-147).

Suresh e Guttag (2019) enumeram duas outras possíveis fontes de vieses. Primeiro, de agregação dos dados, em que as distribuições de grupos entre atributos pode diferir substancialmente, levando modelos a se ajustar mais fortemente a populações maiores do

conjunto de dados, ou mesmo ter desempenho prejudicado. E por fim, na avaliação dos modelos, nos casos em que os dados de avaliação (*benchmark*) não representam a população que é objeto do modelo.

Em sua releitura do conjunto de vieses de Suresh e Guttag (2019) aplicada ao reconhecimento facial, Ruback et. al. (2021) discorrem sobre a criação do modelo, na qual os algoritmos são escolhidos e testados com a base de dados a fim de se obter o "melhor desempenho". Contudo, o melhor desempenho estatístico pode ser obtido exatamente pela confirmação de vieses, um risco identificado por Kordzadeh e Ghasemaghaei (2022). Os autores citam o exemplo dos anúncios de vagas de emprego nas áreas de ciência, engenharia, tecnologia e matemática, que obtêm melhores resultados quando são exibidos mais frequentemente para homens do que para mulheres (LAMBRECHT; TUCKER *in* KORDZADEH; GHASEMAGHAEI, 2022). Ao descrever os sistemas algorítmicos de direcionamento de anúncios online em sua palestra no TED, Zeynep Tufekci (2017) usa um caso hipotético de publicidade de passagens aéreas para Las Vegas, que provavelmente teriam bons resultados quando exibidos para apostadores compulsivos ou pessoas com transtorno bipolar em episódios de mania. O caso se torna mais interessante quando a pesquisadora conta que, após uma das palestras, um(a) cientista da computação a contou que havia desenvolvido uma forma de identificar episódios de mania a partir de publicações em redes sociais mesmo antes de diagnósticos clínicos, mas não conseguiu publicar o estudo. Este risco decorre de características identificadas pelo processo de treinamento dos modelos e produzem resultados estatisticamente precisos, porém socialmente indesejáveis.

A investigação de Buolamwini e Gebru (2018) identificou um viés decorrente da base de dados nos modelos de reconhecimento facial e levantou questões importantes sobre transparência, responsabilidade e definições de precisão dos modelos por demografia. Na década de 1990, Lorna Roth (2016) investigou a má qualidade da reprodução de peles escuras em fotografias tiradas com filme e reveladas em laboratório. Descobriu que a fórmula química das emulsões dos filmes tendiam a reproduzir melhor peles claras, além de os equipamentos dos laboratórios de revelação acompanharem cartões com escalas de cinza, mas também fotos de mulheres (de pele clara) com roupas de cores fortes para indicar o que seria o resultado esperado após sua calibragem. A emulsão dos filmes Kodak teria sido alterada para aumentar a gama de tons marrons não por causa de sua baixa qualidade ao reproduzir pessoas negras, mas após reivindicação de fabricantes de chocolates e móveis nos anos 1960-70. Ainda segundo a autora, câmeras digitais carregaram problemas semelhantes, da incapacidade de detectar rostos de pele escura por um equipamento da HP ao alerta de que pessoas asiáticas teriam piscado ao tirar as fotos, por uma câmera Nikon.

A composição química da emulsão fotográfica, os ajustes dos equipamentos de revelação e os códigos digitais se aproximam, assim, ao promover ou participar do obscurecimento de vieses sociais estruturais. O caso estudado na próxima seção se alinha com os exemplos anteriores. A condição explícita do atributo da base de dados que diz respeito a negros auxilia a exposição didática do problema em questão. Porém, este atributo também é exemplo de obscurecimento, na forma de uma função matemática.

A base de dados *Boston Housing*

Pacotes de algoritmos e repositórios acadêmicos possuem *toy datasets,* pequenas bases de dados que são usadas para exemplificar e testar o funcionamento dos algoritmos de aprendizagem de máquina sem a necessidade de se pesquisar e fazer o download de arquivos externos (SCIKIT-LEARN, 2022). A base de dados *Boston Housing* é um deles. Possui apenas 506 linhas ou registros individuais com 14 colunas ou atributos. Está disponível em pacotes de algoritmos de aprendizagem de máquina como o scikit-learn e o TensonFlow, que referenciam o repositório da Universidade Carnegie Mellon (CARNEGIE MELLON UNIVERSITY, s/d). A utilidade de um *toy dataset* deriva das características de cada atributo – se possui dados categóricos ou contínuos, por exemplo (CASTRO; FERRARI, 2016, p. 50) – mais do que dos dados em si.

A base em questão foi usada principalmente em trabalhos que tratam de problemas de regressão, segundo a página do pacote scikit-learn (SCIKIT-LEARN, 2022). Os atributos e suas descrições traduzidas estão na **Tabela 1**.

A base de dados hospedada na Universidade Carnegie Mellon não possui data, mas a versão disponível na Universidade da Califórnia, Irvine possui a data 7 de julho de 1993 (UNIVERSITY OF CALIFORNIA IRVINE, s/d). Seu conteúdo foi retirado de um artigo de 1978 que propõe um modelo de precificação hedônica para estimar a disposição dos habitantes de pagar por melhoria na qualidade do ar na região metropolitana de Boston, a partir de dados do censo de 1970. O procedimento usado no artigo mostrava que o impacto negativo nos preços dos imóveis aumentariam com os níveis de poluição do ar e renda domiciliar (HARRISON JR; RUBINFELD, 1978, p. 98).

Apenas em 2019 um pesquisador chamou atenção para uma das colunas desta base de dados: "B" (CARLISLE, 2019). Este atributo diz respeito à proporção de negros em cada cidade. A mera inclusão desta característica na base de

Atributo	Descrição
CRIM	Taxa de crime *per capita* por cidade
ZN	Proporção de zonas residenciais por lote acima de 25.000 pés quadrados
INDUS	Proporção de acres comerciais com negócios não-varejistas por cidade
CHAS	Variável *dummy* para o Rio Charles (= 1 se o terreno tem fronteira com o rio, 0 caso contrário
NOX	Concentração de óxido nítrico (partes por 10 millhões)
RM	Número médio de cômodos por habitação
AGE	Proporção de unidades ocupadas pelos proprietários construídas antes de 1940
DIS	Distâncias ponderadas para cinco centros de emprego de Boston
RAD	Índice de acessibilidade a rodovias radiais
TAX	Preço total da taxa de propriedade por US$10.000
PTRATIO	Relação aluno-professor por cidade
B	1000(Bk – 0.63)^2, onde Bk é a proporção de negros por cidade
LSTAT	Percentual da população com "baixo status"
MEDV	Valor médio das casas ocupadas por proprietários em milhares de dólares

Tabela 1. Atributos da base de dados *Boston Housing* (CARNEGIE MELLON UNIVERSITY, s/d, tradução nossa).

dados merece discussão, mas neste caso, há um problema adicional. Não se trata apenas da proporção de negros, mas de uma função desenvolvida de modo que a proporção de negros em uma cidade tem impacto negativo nos preços dos imóveis até certo nível, quando passam a ter impacto positivo. A descrição do atributo no artigo é a seguinte:

> *Proporção de negros na população. Em níveis baixos ou moderados de B, um aumento em B deveria ter uma influência negativa nos valores dos imóveis se negros são considerados vizinhos indesejáveis por brancos. No entanto, a discriminação no mercado significa que os valores dos imóveis são mais altos com níveis muito altos de B. Espera-se, então, uma relação parabólica entre a proporção de negros em uma vizinhança e os valores dos imóveis (HARRISON JR; RUBINFELD, 1978, p. 96, tradução nossa).*

Em primeiro lugar, os dados deste atributo são estruturados a partir do pressuposto que vizinhos negros seriam considerados indesejáveis por brancos – cabe lembrar que a Lei dos Direitos Civis, que proibiu a discriminação racial nos EUA (antes permitida pelas chamadas Leis Jim Crow) foi assinada em 1964 (DA SILVA, 2021). Em seguida, modela matematicamente o que Carlisle (2019) chama de "efeito gueto", traçando um limite de segregação racial a partir do qual o valor dos imóveis deixa de sofrer impacto negativo e passa a sofrer impacto positivo. E faltam evidências de que a segregação racial teria impacto direto no valor dos imóveis. O pesquisador cita estudo do *National Bureau of Economic Research* de 1975 apontando que seria praticamente impossível medir o papel da segregação racial nos valores, apesar de reconhecer que ainda havia, na época, resistências de brancos em relação à integração (NBER *in* CARLISLE, 2019). Os dados referentes à raça compõem, portanto, um exemplo de viés histórico.

Mas o ponto que mais chama atenção de Carlisle é a transformação de dados realizada por Harrison Jr. e Rubinfeld (1978). Os autores aplicaram uma função não-inversível aos dados, impedindo que se retorne aos valores originais – para alguns registros, há dois valores possíveis no censo para cada valor entre 0 e 136,9, um acima e outro abaixo do limite de 63% de segregação responsável pelo efeito gueto descrito acima. É este o ponto no qual ocorre o obscurecimento na forma de uma operação matemática. Há apenas 36 registros com essas condições, que podem até mesmo ser eliminados da base sem grandes prejuízos quantitativos. Mas com esta operação, os autores impossibilitaram a recuperação dos dados originais, anteriores à transformação submetida por eles.

Carlisle (2019) buscou os dados originais e encontrou somente 20 desses 36 valores. Sua investigação termina com este problema e algumas questões, como a real necessidade de manter esses dados na base, mas também sobre o próprio papel do censo na democracia e as formas como foi usado em relação a minorias. Por fim, lembra da prática de *redlining*, o uso de marcadores raciais e de classe para delimitar territórios urbanos e restringir a oferta de serviços como empréstimos aos seus habitantes – que ganha novas dimensões na atualidade (NOBLE, 2021, P. 8). A base continua disponível, mas será removida na próxima atualização do scikit-learn (2022), que emite um aviso quando é carregada.[1]

A partir da análise desse processo, Carlisle reforça dois pontos essenciais para o trabalho com dados e algoritmos. Primeiro, deve-se publicar apenas dados originais, "crus". Qualquer transformação deve ser feita na etapa de modelagem e anotada a fim de que se torne compreensível. Termina seu texto com um alerta mais amplo: os cientistas de dados (mas, pode-se acrescentar, todos os profissionais) devem se questionar se os dados com os quais trabalham fazem sentido. A fim de compreender o papel do atributo B em modelos de regressão, na próxima seção serão apresentados testes feitos com o pacote de algoritmos scikit-learn em diferentes cenários.

Aprendizado de máquina: o treinamento de um modelo algorítmico

Nesta seção, será apresentada a etapa de treinamento de um modelo de aprendizado de máquina. Nesta tarefa, os atributos usados como *inputs* para os modelos são chamados variáveis preditoras (FERREIRA, s/d) e o atributo que contém o dado a ser estimado – neste caso, o valor médio dos imóveis (coluna MEDV) – é o que se denomina "variável alvo" ou "variável meta" (CASTRO; FERRARI, 2016, p. 49). Nos modelos de regressão linear simples, este atributo é a "variável dependente ou variável endógena, y, aquela cujo comportamento será explicado pela variável x, chamada de variável explicativa, regressor ou variável independente" (CHEIN, 2019, p. 11, grifos da autora). Na base Boston housing há 13 atributos independentes – coeficientes parciais de regressão (CHEIN, 2019, p. 33).

Há inúmeras implementações de modelos de aprendizado de máquina disponíveis online. A fim de ilustrar o processo de forma mais simplificada e didática, foram selecionados três exemplos que tratam da escolha de atributos (SHETYE, 2019) e do treinamento de um modelo com um algoritmo de regressão linear (AGARWAL, 2018; ÇEPEL, 2019). O processo descrito a seguir se limita às formas mais simples de cada etapa, com a finalidade de privilegiar a clareza e de acordo com os limites propostos para esta seção.

A escolha de atributos é uma etapa fundamental de qualquer projeto de aprendizado de máquina, pois busca-se eliminar dados irrelevantes ou redundantes a fim de agilizar o processamento e reduzir ruídos. No entanto, diferentes métodos resultam em diferentes conjuntos de atributos. Shetye (2019) descreve três formas de realizar esta tarefa. A mais simples é pelo coeficiente de correlação de Pearson e seleção dos atributos que tenham maior correlação com a variável-alvo MEDV. Os níveis de correlação entre os atributos são facilmente visualizáveis em um mapa de calor (**Figura 1**), onde se pode notar que atributo B tem correlação relativamente baixa não apenas com o valor dos imóveis, mas com todas as outras variáveis.

Três atributos que possuem coeficiente de correlação superior a 0,5 com MEDV: o número médio de cômodos (RM), o percentual da população com "com *status* baixo" (LSTAT) – atributo que também gera questões – e a relação aluno-professor por cidade (PTRATIO); contudo, devido à alta correlação entre RM e LSTAT são escolhidos apenas os atributos LSTAT e PTRATIO como variáveis explicativas por meio deste método.

A segunda forma de seleção de atributos é por meio de um *wrapper*, que usa algoritmos com diferentes combinações de atributos, que são adicionados ou removidos

1 – O texto do aviso é: "The Boston housing prices dataset has an ethical problem. You can refer to the documentation of this function for further details. The scikit-learn maintainers therefore strongly discourage the use of this dataset unless the purpose of the code is to study and educate about ethical issues in data science and machine learning".

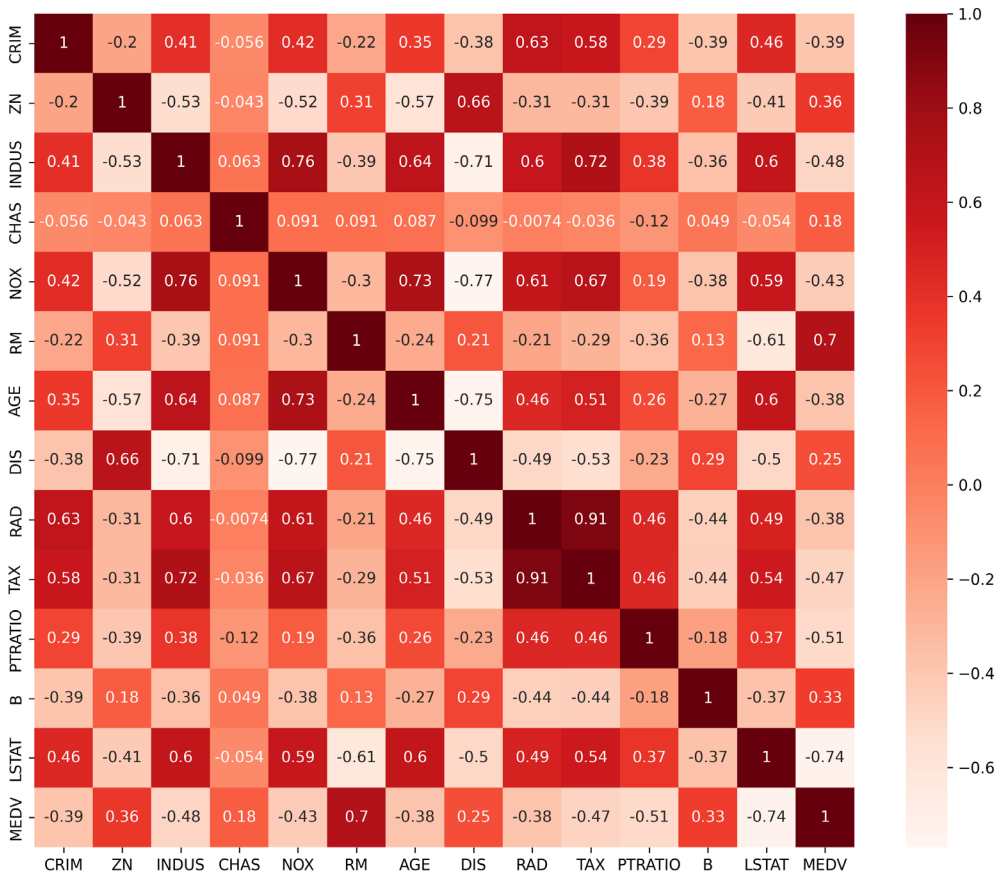

Gráfico 1. Mapa de calor dos atributos da base Boston housing (Shetye, 2019).

a partir dos resultados obtidos. Este processo é repetido com dois métodos. No primeiro (*Backward Elimination* com medição pelo p-valor), são selecionados 11 dos 13 atributos (CRIM, ZN, CHAS, NOX, RM, DIS, RAD, TAX, PTRATIO, B, LSTAT); no segundo (*Recursive Feature Elimination*), restam 10 colunas (CRIM, ZN, INDUS, CHAS, NOX, RM, DIS, RAD, PTRATIO, LSTAT).

A terceira e última forma de seleção de atributos apresentada por é o método *Embedded*, iterativo, por meio da regressão Lasso, que leva à exclusão de variáveis preditoras (FERREIRA, s/d) e resulta também em 10 atributos (RM, DIS, PTRATIO, LSTAT, CRIM, RAD, ZN, TAX, AGE, B). O atributo B não é excluído, mas tem o menor coeficiente entre os selecionados.

As duas formas de seleção de atributos consideradas mais precisas mantiveram conjuntos semelhantes de atributos, embora B tenha sido substituído por INDUS em um dos casos. Ao apresentar diferentes modos de seleção de atributos, busca-se explicitar que os procedimentos adotados nessa etapa do processo podem alterar o conjunto dos dados selecionados para treinar o modelo.

A etapa seguinte é o chamado treinamento do modelo, que neste exemplo é realizada com o módulo de regressão linear "LinearRegression()" do pacote scikit-learn e os atributos selecionados pelos métodos acima. O processo consiste em dividir a base em dois conjuntos de registros, um para treinamento e outro para teste, onde o modelo é aplicado e os resultados são verificados. A fim de simplificar a visualização do processo, o treinamento do modelo foi feito com um único atributo de cada vez: LSTAT e PTRATIO, selecionados pela correlação, RM e B, para comparação.

O **Gráfico 2** (formatação obtida em Kim, 2019) mostra os dados da parcela da base usada para teste com pontos e a linha resultante do modelo de regressão a partir de cada atributo selecionado, com uma divisão de 70% dos registros para treinamento e 30% para teste. A linha indica onde os valores dos imóveis são estima-

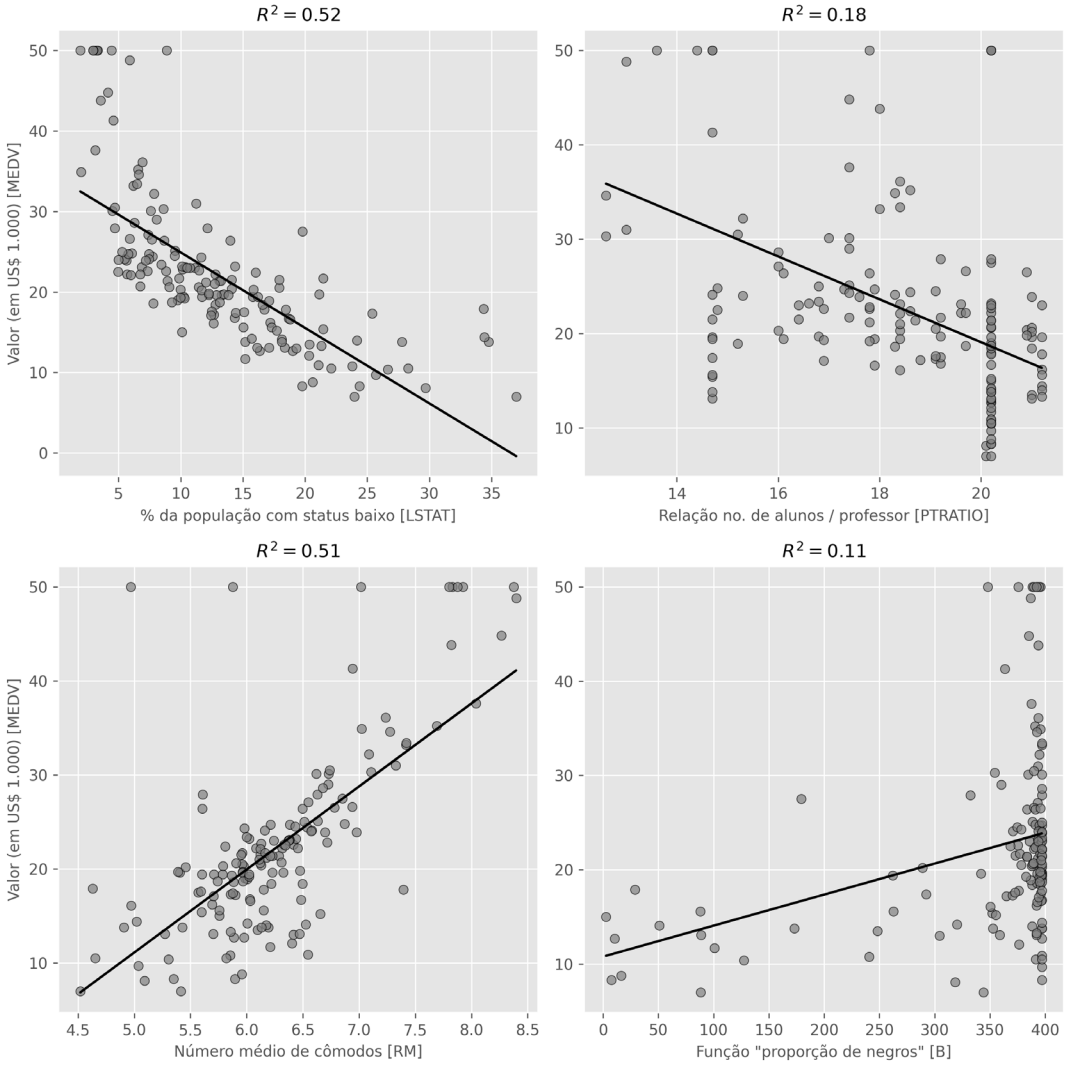

Gráfico 2. Modelos gerados a partir de atributos selecionados.

dos (no eixo vertical) e os pontos são os dados reais de valores de imóveis na porção de testes da base.

Os resultados, medidos pelo coeficiente de determinação (r^2), indicam que os atributos LSTAT (0,52) e RM (0,51) obtiveram maior precisão, enquanto PTRATIO (0,18 mesmo com maior correlação com MEDV) e B (0,11) geraram modelos menos preciso para estimar preços dos imóveis.[2] Outras formas de visualização dos resultados são as equações geradas com o treinamento e a estimativa individual de valores para cada nível do atributo preditor.

Por fim, espera-se que a combinação de atributos aumente a complexidade e a precisão dos modelos, especialmente em uma base usada para testes e aprendizado. Os resultados estão listados na **Tabela 2**.

2 – O coeficiente de determinação "dá a percentagem de variabilidade dos ys (variável a prever) que fica explicada em função da variabilidade dos xs. Assim, um valor de r 2 ≈ 1 significa que, em princípio, a nuvem de pontos apresentada no diagrama de dispersão está próxima da reta de regressão, considerada para o modelo de regressão" (MARTINS, 2018). Apesar de haver críticas a seu uso como medida de avaliação da qualidade do modelo (QUININO; REIS; BESSEGATO, 1991), é uma medida útil para a verificação da adequação do modelo aos dados.

Método de seleção	Atributos	r^2
Correlação > 0.5 com MEDV	LSTAT, PTRATIO	0,51
Correlação > 0.5 com MEDV + B	LSTAT, RM, B	0,52
Backward Elimination	CRIM, ZN, CHAS, NOX, RM, DIS, RAD, TAX, PTRATIO, B, LSTAT	0,68
Recursive Feature Elimination	CRIM, ZN, INDUS, CHAS, NOX, RM, DIS, RAD, PTRATIO, LSTAT	0,68
Embedded	RM, DIS, PTRATIO, LSTAT, CRIM, RAD, ZN, TAX, AGE, B	0,65

Tabela 2. Resultados dos modelos treinados a partir de atributos selecionados por diferentes métodos.

A Tabela 2 mostra que as bases de dados com mais atributos geraram modelos mais precisos. Neste exemplo, não foram removidos os valores atípicos, nem foi feita nenhuma transformação ou normalização nos dados. Implementações mais complexas chegam a coeficientes de até 0,89 (ÇEPEL, 2019). O objetivo desta seção foi ilustrar a etapa de modelagem, ou de treinamento e teste dos modelos gerados a partir de algoritmos e, ao mesmo tempo, testar a influência do atributo B em um modelo simples de regressão linear. Foi possível verificar que o atributo em questão não é um dos que possui maior importância em diversos métodos de seleção de atributos e que a precisão das estimativas feitas a partir dele são relativamente baixas. O fato de a base *Boston Housing* ter um número reduzido de registros (506) é um limitador da precisão de estimadores. Contudo, a questão sobre a necessidade deste atributo na base de dados ganha uma dimensão adicional quando se verifica que consta entre os atributos menos relevantes e não influenciou significativamente nos resultados.

Revisão Sistemática de Literatura

Uma Revisão Sistemática de Literatura (RSL) é uma forma de pesquisa que possui características e objetivos próprios. Mais do que as revisões de literatura sem regras definidas, a RSL

> *busca entender e dar alguma logicidade a um grande corpus documental, especialmente, verificando o que funciona e o que não funciona num dado contexto. Está focada no seu caráter de reprodutibilidade por outros pesquisadores, apresentando de forma explícita as bases de dados bibliográficos que foram consultadas, as estratégias de busca empregadas em cada base, o processo de seleção dos artigos científicos, os critérios de inclusão e exclusão dos artigos e o processo de análise de cada artigo (GALVÃO; RICARTE, 2019, pp. 58-9).*

Entre os tipos de RSL, a revisão narrativa é considerada adequada "quando os estudos quantitativos a serem considerados empregam diversas metodologias ou partem de diferentes conceituações teóricas, construtos e/ou relacionamentos" e "podem ser usadas para fornecer uma descrição histórica do desenvolvimento da teoria e da pesquisa sobre um tópico" (GALVÃO; RICARTE, 2019, p. 59). Esta revisão, embora bastante simplificada em relação às estratégias disponíveis, objetiva avaliar se o atributo B é identificado, de que forma é descrito e qual seu papel no desenvolvimento de modelos com o uso da base de dados *Boston Housing*. Apesar de limitada, consiste em um esforço para estender à produção acadêmica a questão sobre a atenção aos dados com que se trabalha que Carlisle direciona a cientistas de dados.

Foi feita uma pesquisa pelo termo composto ("boston housing" AND "machine learning") em todos os campos das publicações dos anos 2017 a 2022 das bases científicas Scopus e Web of Science no dia 02/08/2022. Retornaram 123 na base Web of Science e 5 artigos na Scopus. A leitura foi feita a partir das publicações mais recentes.

A revisão teve início pelas 14 publicações do ano de 2022. A grande maioria apenas cita outros artigos que contêm as palavras "Boston housing". Cinco dessas publicações citam De Cock (2011), que propõe um substituto para a base de dados Boston, o *Ames dataset* (ou "*Ames housing data*") – essencialmente por causa da defasagem de preços ocorrida no período, tornando os valores irreais. Não há menção ao atributo B nesses textos, que foram excluídos da revisão. Também foram excluídos dois outros

artigos que citam publicação de Murakami (2017), que apresenta um pacote com funções para estimar modelos de regressão espacial que usa apenas sete atributos da base *Boston housing*, entre os quais não está B, e o método utilizado para seleção de atributos não é identificado.

Dentre as publicações de 2022 que citam diretamente a base de dados *Boston housing*, Wang et al. (2022) usam a base entre outras sem relação com ela, não identificam os atributos ou os descrevem. O mesmo se verifica em Heo et al. (2022), onde não há descrição dos atributos, nem de seu papel nos resultados finais. Em Bai (2022), somente algumas linhas da base de dados são exibidas em uma imagem com ilustrativa do processo proposto, na qual o atributo B aparece com o título *black*, sem indicação do que cada atributo significa. Na descrição dos resultados, apenas aspectos das fórmulas são mencionados, como a acurácia, a habilidade de lidar com erros e formas de melhorar as soluções.

A *Boston housing* é uma entre as cinco bases de dados usadas por Artelt et al. (2022) para avaliar o método proposto para explicar adaptações e diferenças nos modelos quando aplicados a novos conjuntos de dados. É usada especificamente em um experimento que divide os registros em dois grupos a partir do valor do atributo NOX. O atributo B aparece no gráfico da direita na **Figura 2** da publicação, mas como não sofre mudanças, não é indicado na explicação dos resultados, nem discutido, apesar da preocupação do texto com a transparência e a interpretabilidade dos modelos.

Por fim, o artigo de Mashhadi, Zolyomi e Quedado (2022) é dedicado ao estudo da integração de ferramentas de visualização como um recurso para o aprendizado de conceitos de justiça algorítmica. A *Boston housing* é mencionada apenas como uma das bases de dados disponíveis na ferramenta do projeto Fairlearn, em cujo site há uma explicação detalhada sobre a base de dados. Seu potencial didático reside exatamente na transparência com que se pode identificar os vieses presentes nos dados, que além de incompletos, têm baixa qualidade, citando Carlisle (2019) ao apontar que qualquer modelo que use esta base incorpora racismo codificado aos seus resultados (FAIRLEARN, s/d).

Dada a ausência de descrições e discussões sobre o atributo em questão entre os resultados da pesquisa pelo termo em todos os campos dos artigos, a estratégia foi alterada para concentrar a avaliação nos resultados da pesquisa pelos mesmos termos e nas mesmas bases de dados, apenas em títulos, resumos e palavras-chave, mas em todo o período disponível. Os retornos totalizaram 20 artigos na base Scopus e 13 na Web of Science. Destes, 11 eram duplicados, restando 22 publicações, 9 delas no período entre 2017 e 2022. Três dessas publicações foram removidas por terem acesso restrito e uma delas é Bai (2022), analisada acima. As cinco publicações restantes ofereceram mais informações de interesse para o objetivo desta revisão e são apresentadas em ordem cronológica inversa abaixo.

Em Adetunji et al., (2021), O atributo B é destacado entre aqueles que possuem alta distorção, juntamente com CRIM e ZN. A descrição estatística deste último mostra que, embora haja valores entre 0.32 e 396.90, o primeiro quartil (25% menores valores) é 375.3775 e o terceiro (75%) é 396.2250, ou seja, uma concentração muito alta de valores que pode ser observada na Figura 2 do presente artigo (treinamento do modelo) e também foi notada por Shetye (2019). Contudo, não há discussão sobre o caráter do atributo. Mesa et al. (2019) aplicam o método Lasso (*Least Absolute Shrinkage and Selection Operator*) para seleção de variáveis, no qual o atributo B é excluído (cf. pp. 635-636). Por fim, Nalenz e Villani (2018) listam 16 bases de dados para propor um modelo bayesiano para regressão não linear e classificação usando agrupamentos de árvores de decisão. Ao contrário das outras publicações avaliadas nesta revisão, os autores apresentam a base de dados Boston, suas dimensões ecológicas (NOX e PART) e socioeconômicas (B, LSTAT e CRIM). De acordo com o método *HorseRule* de cálculo de importância dos atributos, B aparece em 11º lugar entre os 13 atributos da base (Fig 8, p. 2400).

A baixa correlação do atributo B e sua recorrente exclusão em diversos métodos de seleção de atributos denota que pode ser considerada até mesmo prescindível na base, o que torna ainda mais importante o questionamento sobre as razões pelas quais foi incluído, ou não foi removido anteriormente. Esta busca bastante específica pode ter sido limitada pelo caráter técnico dos textos recuperados a partir da busca pela base de dados. No entanto, corroboram o argumento de que o trabalho da modelagem deve envolver a atenção aos dados a fim de mitigar vieses, inclusive alguns facilmente identificáveis como o caso da base Boston. Todos os textos listados acima, exceto Nalenz e Villani (2018), foram publicados após a investigação de Carlisle (2019), enquanto o aviso sobre o problema ético na base e sua futura remoção do pacote scikit-learn só foi adicionado em agosto de 2021, segundo o histórico do GitHub do pacote (GRISEL, 2021). Conclui-se que o questionamento e o alerta de Carlisle pode ser estendido às publicações acadêmicas e é necessário esforço a fim de que as características dos dados utilizados em modelos sejam avaliadas mesmo nas pesquisas de cunho mais técnico.

Considerações finais

À medida que cada vez mais atividades são mediadas por processos algorítmicos, torna-se mais relevante seu escrutínio. No entanto, os modelos e os dados usados para seu treinamento são, usualmente, inacessíveis a cidadãos, órgãos reguladores e pesquisadores. O ponto de partida deste artigo é a investigação por Michael Carlisle (2019) de um atributo com viés racista em uma base de dados de treinamento que é amplamente utilizada por cientistas de dados para ensino de ciência de dados, testes de modelos e, conforme foi identificado nesta pesquisa, para avaliação de modelos em publicações acadêmicas. Assim como no trabalho de Buolamwini e Gebru (2018), uma base de dados pública é um objeto privilegiado para a avaliação de etapas que são obscurecidas e só podem ser inferidas em pesquisas baseadas nos resultados desses modelos. A fim de contribuir com o conjunto de investigações sobre o tema, este artigo deriva do questionamento sobre as formas pelas quais viesse sociais, na forma de preconceitos e estereótipos e outros, podem ser codificados no processo de desenvolvimento de modelos com algoritmos de aprendizado de máquina.

Partindo do arcabouço teórico que trata dos vieses históricos e sociais (SILVA, 2020, 2022; NOBLE, 2021; ONEIL, 2021), a pesquisa busca evidenciar sua codificação concentrando-se nas etapas que vão desde a seleção dos dados até a avaliação dos modelos. O primeiro passo desta tarefa foi apresentar e contrastar diferentes abordagens sobre esse processo – KDD, CRISP-DM, SEMMA (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996; AZEVEDO; SANTOS, 2008) – e as possíveis formas de codificação de vieses em cada uma delas (SURESH; GUTTAG, 2019; ONEIL, 2021; RUBACK; AVILA; CANTERO, 2021). A base de dados na qual um desses vieses já havia sido investigado surge como um objeto privilegiado para a discussão. De início, pode-se apontar o viés racial na mera inclusão deste atributo em uma base de dados que visa treinar modelos capazes de estimar o valor de imóveis. Porém, a descoberta de uma transformação desses dados com uma função não inversível (CARLISLE, 2019) evidencia o processo de obscurecimento (LATOUR, 2001) que pode ocorrer nessa etapa do processo.

Em seguida, buscou-se descrever em detalhes a etapa de modelagem, com a apresentação de diferentes métodos para seleção de variáveis ou atributos (SHETYE, 2019) e a observação de que resultam em diferentes conjuntos de dados. Para o desenvolvimento do modelo, foi usado um algoritmo de regressão linear e a apresentação de seus resultados foi feita a partir de apenas um atributo para que a visualização dos seus resultados se tornasse mais simples. Neste caso, nota-se que o impacto final do atributo B não é significativo em comparação com outros. Ressalta-se que o modelo gerado neste artigo visa apenas ilustrar o processo, tendo em vista que a base de dados é relativamente pequena e o processo foi simplificado, sem a utilização de métodos adicionais que podem aumentar sua precisão, como a validação cruzada ou a normalização dos dados (LEYREZOV, 2017, ÇEPEL, 2019). Porém, o objetivo desta seção do trabalho consistiu em evidenciar o papel de cada atributo no treinamento de um modelo e o produto final da etapa de modelagem se realizada com diferentes conjuntos de dados.

Por fim, estende-se à produção acadêmica a questão de Carlisle (2019) acerca do cuidado com os dados usados no treinamento dos modelos por parte de cientista de dados. Por meio de uma revisão sistemática de literatura com foco bastante específico, realizou-se pesquisa por publicações nas quais a base de dados *Boston Housing* foi usada, mencionada ou discutida. Os resultados mostram que os autores e as autoras que usam esta base, entre outras bases de dados diversas, muitas vezes não identificam nem descrevem seus atributos. Corroborando o impacto relativamente baixo do atributo B identificado pelos métodos de seleção de variáveis da seção anterior, este atributo não é descrito (WANG et al., 2022; HEO et al., 2022), é descartado (MURAKAMI, 2017); MESA et al., 2019), figura entre os atributos com pouco potencial (NALENZ; VILLANI, 2018; ADETUNJI et al., 2021), ou não figura entre aqueles significativos para os resultados (ARTELT et al., 2022). Mesmo que tenham apenas mencionado a base de dados em sua pesquisa sobre ferramentas para ensino de responsabilidade na Inteligência Artificial (IA), Mashhadi, Zolyomi e Quedado (2022) citam o projeto Fairlearn (s/d) que, por sua vez, alerta para seus os problemas.

Espera-se, assim, ter oferecido com este artigo uma contribuição para a pesquisa sobre a codificação de vieses sociais por meio de uma perspectiva interdisciplinar que buscou articular a pesquisa crítica do papel dos modelos algorítmicos na atualidade com a apresentação e discussão dos métodos e das técnicas que compõem este processo. Vislumbra-se desdobramentos para este trabalho, desde a pesquisa dos vieses históricos e sociais presentes nos dados públicos (NOBLE, 2021), como o caso do próprio censo (CARLISLE, 2019), passando pela aplicação de métodos mais complexos no treinamento de modelos com esta base de dados, pela pesquisa e avaliação de técnicas para mitigação de vieses, como aquelas propostas por Suresh e Guttag (2019), até a verificação das potencialidades e limitações para a tradução e implementação de ferramentas de ensino de IA Responsável no Brasil.

Referências

ADETUNJI, A. B. et al. House Price Prediction using Random Forest Machine Learning Technique. Department of Computer Science, Faculty of Computing and Informatics, Ladoke Akintola University of Technology, Nigeria: Elsevier B.V., 2021. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85124949478&doi=10.1016%2Fj.procs.2022.01.100&partnerID=40&md5=06ff451087e838e7c76eb2fb93bd1764>. Acesso em: 15/03/2022.

AGARWAL, A. *Linear Regression on Boston Housing Dataset*. Disponível em: <https://towardsdatascience.com/linear-regression-on-boston-housing-dataset-f409b7e4a155>. Acesso em: 03/12/2022.

ARTELT, A. et al. Contrasting Explanations for Understanding and Regularizing Model Adaptations. *Neural Processing Letters*, 2022. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85129255368&doi=10.1007%2Fs11063-022-10826-5&partnerID=40&md5=346419ac0f0bc85123dfba6d33df0060>. Acesso em: 18/08/2022.

AZEVEDO, A.; SANTOS, M. F. KDD, SEMMA and CRISP-DM: a parallel overview. *IADS-DM*, 2008.

BAI, S. Boston house price prediction: machine learning. In: Troy High School, Fullerton, CA, United States. *Anais*... Troy High School, Fullerton. 2022. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85131831328&doi=10.1109%2FICSP54964.2022.9778372&partnerID=40&md5=bb9836f9cd2ea81bec3792cb66b5425b>. Acesso em: 18/08/2022.

BUOLAMWINI, J.; GEBRU, T. Gender shades: Intersectional accuracy disparities in commercial gender classification. In: Conference on fairness, accountability and transparency, *Anais*...PMLR, 2018.

CARLISLE, M. *racist data destruction?. a Boston housing dataset controversy | by M Carlisle | Medium*. 2019. Disponível em: <https://medium.com/@docintangible/racist-data-destruction-113e3eff54a8>. Acesso em: 04/04/2022.

CASTRO, J. C. L. de. Plataformas algorítmicas: interpelação, perfilamento e performatividade (Algorithmic Platforms: Interpellation, Profiling and Performativity). *Revista Famecos, Porto Alegre (RS)*, v. 26, n. 3, p. 1–24, 2019.

CASTRO, L. N. de; FERRARI, D. G. Introdução à mineração de dados: conceitos básicos, algoritmos e aplicações. *São Paulo: Saraiva*, v. 5, 2016.

ÇEPEL, T. *Boston Housing Regression Analysis*. 2019. Disponível em: <https://www.kaggle.com/code/tolgahancepel/boston-housing-regression-analysis/notebook>. Acesso em: 5/12/2022.

CHEIN, F. Introdução aos modelos de regressão linear: um passo inicial para compreensão da econometria como uma ferramenta de avaliação de políticas públicas. 2019.

DA SILVA, W. B. C. A LUTA PELOS DIREITOS CIVIS NOS ESTADOS UNIDOS. *Revista Ibero-Americana de Humanidades, Ciências e Educação*, v. 7, n. 9, p. 414–423, 2021.

DE COCK, D. Ames, Iowa: Alternative to the Boston housing data as an end of semester regression project. *Journal of Statistics Education*, v. 19, n. 3, 2011.

FAIRLEARN. *Revisiting the Boston Housing Dataset*. Disponível em: <https://fairlearn.org/main/user_guide/datasets/boston_housing_data.html>. Acesso em: 02/12/2022.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM*, v. 39, n. 11, p. 27–34, 1996.

FERREIRA, E. V. *Regularização*. Disponível em: <http://cursos.leg.ufpr.br/ML4all/apoio/Regularizacao.html#__regularização__>. Acesso em: 01/12/2022.

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, 2019.

GILLESPIE, T. Algorithm. In: PETERS, B. (Ed.). *Digital Keywords*. 1. ed. Princeton: Princeton University Press, 2016.

GRILL, G. *Critical Algorithm Studies*. Disponível em: <https://algorithmstudies.files.wordpress.com/2016/05/critical-algorithm-studies-introduction-1-1.pdf>. Acesso em: 28/11/2022.

GRISEL, O. *DOC add warning regarding the load_boston function*. s.l: s.n.. Disponível em: <https://github.com/scikit-learn/scikit-learn/commit/c592361e536fbc84c59935b7b4c659e8ab38737c>. Acesso em: 29/11/2022.

HARRISON JR, D.; RUBINFELD, D. L. Hedonic housing prices and the demand for clean air. *Journal of environmental economics and management*, v. 5, n. 1, p. 81–102, 1978.

HEO, J. P. et al. Shallow Fully Connected Neural Network Training by Forcing Linearization into Valid Region and Balancing Training Rates. *Processes*, v. 10, n. 6, 2022. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85132274595&doi=10.3390%2Fpr10061157&partnerID=40&md5=870c4957b12ed3e0e41c6a2d092413ac>. Acesso em: 29/08/2022.

IRVINE, U. of C. *Boston Housing Data*. Disponível em: <https://archive.ics.uci.edu/ml/machine-learning-databases/housing/housing.names>. Acesso em: 26/11/2022.

KAISER, B. *Manipulados: como a Cambridge Analytica e o Facebook invadiram a privacidade de milhões e botaram a democracia em xeque*. Rio de Janeiro: Harlequin, 2020.

KIM, E. *Multiple Linear Regression and Visualization in Python*. Disponível em: <https://aegis4048.github.io/mutiple_linear_regression_and_visualization_in_python>. Acesso em: 25/11/2022.

KORDZADEH, N.; GHASEMAGHAEI, M. Algorithmic bias: review, synthesis, and future research directions. *European Journal of Information Systems*, v. 31, n. 3, p. 388–409, 2022.

LATOUR, B. *A esperança de Pandora*. Bauru: EDUSC, 2001.

LEYREZOV, O. *Model Evaluation and Validation: Predicting Boston Housing Prices*. 2017.Disponível em: <https://olegleyz.github.io/boston_housing.html>. Acesso em: 25/11/2022.

MARTINS, E. G. M. Coeficiente de determinação. *Revista Ciência Elementar*, v. 6, n. 1, p. 24, 2018.

MASHHADI, A.; ZOLYOMI, A.; QUEDADO, J. A Case Study of Integrating Fairness Visualization Tools in Machine Learning Education. In: Computing and Software Systems, University of Washington Bothell, Bothell. 2022. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85129702990&doi=10.1145%2F3491101.3503568&partnerID=40&md5=9775e8a5ecb6ca3890505c9b23e320e2>. Acesso em: 30/11/2022.

MESA, D. A. et al. A Distributed Framework for the Construction of Transport Maps. *Neural Computation*, v. 31, n. 4, p. 613–652, 2019. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85062970757&doi=10.1162%2Fneco_a_01172&partnerID=40&md5=e438cf22f9653b8138d1d367b4dd0ebf>. Acesso em: 28/08/2022.

MURAKAMI, D. Spatial regression modeling using the spmoran package: Boston housing price data examples. *arXiv preprint arXiv:1703.04467*, 2017. Acesso em: 21/08/2022.

NALENZ, M.; VILLANI, M. Tree Ensembles with Rule Structured Horseshoe Regularization. *Annals of Applied Statistics*, v. 12, n. 4, p. 2379–2408, 2018. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85057187571&doi=10.1214%2F18-AOAS1157&partnerID=40&md5=ad3941bd780540d16d1b10234331cdb5>. Acesso em: 28/08/2022.

NEFF, G. Talking to bots: Symbiotic agency and the case of Tay. *International Journal of Communication*, 2016.

Netlab UFRJ. Recomendação no Youtube: o caso Jovem Pan. 5 de Setembro de 2022, Escola de Comunicação da Universidade Federal do Rio de Janeiro, Brasil.

NOBLE, S. U. *Algoritmos da Opressão: Como os mecanismos de busca reforçam o racismo*. Santo André: Editora Rua do Sabão, 2021.

O'NEIL, C. *Algoritmos de destruição em massa*. Santo André: Editora Rua do Sabão, 2021.

QUININO, R. C.; REIS, E. A.; BESSEGATO, L. F. O coeficiente de determinação R2 como instrumento didático para avaliar a utilidade de um modelo de regressão linear múltipla. *Belo Horizonte: UFMG*, 1991.

ROTH, L. Questão de pele. *Revista Zum*, n. 10, 2016.

RUBACK, L.; AVILA, S.; CANTERO, L. Vieses no Aprendizado de Máquina e suas Implicações Sociais: Um Estudo de Caso no Reconhecimento Facial. In: Anais do II Workshop sobre as Implicações da Computação na Sociedade, *Anais*...SBC, 2021.

SCHWARZ, J. A. Platform logic: An interdisciplinary approach to the platform-based economy. *Policy & Internet*, v. 9, n. 4, p. 374–394, 2017.

SCIKIT-LEARN. *Toy Datasets - SckitLearn*. Disponível em: <https://scikit-learn.org/stable/modules/generated/sklearn.datasets.make_moons.html>. Acesso em: 26/11/2022.

SHETYE, A. *Feature Selection with sklearn and Pandas*. Disponível em: <https://towardsdatascience.com/feature-selection-with-pandas-e3690ad8504b>. Acesso em: 03/12/2022.

SILVA, T. Racismo Algorítmico em Plataformas Digitais: microagressões e discriminação em código. *Comunidades, algoritmos e ativismos digitais: olhares afrodiaspóricos*, p. 121–135, 2020.

SILVA, Tarcízio. Linha do Tempo do Racismo Algorítmico. Blog do Tarcizio Silva, 2022. Disponível em: <https://tarciziosilva.com.br/blog/posts/racismo-algoritmico-linha-do-tempo>. Acesso em: 03/12/2022.

SILVA, T. *Racismo algorítmico: inteligência artificial e discriminação nas redes digitais*. s.l. Edições Sesc SP, 2022.

SURESH, H.; GUTTAG, J. V. A framework for understanding unintended consequences of machine learning. *arXiv preprint arXiv:1901.10002*, v. 2, 2019.

TUFEKCI, Z. *Were building a dystopia just to make people click on ads*TED Talk, , 2017. . Disponível em: <https://www.ted.com/talks/zeynep_tufekci_we_re_building_a_dystopia_just_to_make_people_click_on_ads/transcript>. Acesso em: 08/12/2022.

UNIVERSITY, C. M. *The Boston house-price data of Harrison, D. and Rubinfeld, D.L.* Disponível em: <http://lib.stat.cmu.edu/datasets/boston>. Acesso em: 26/11/2022.

VAN DIJCK, J.; POELL, T.; DE WAAL, M. *The platform society: Public values in a connective world*. Oxford: Oxford University Press, 2018.

WANG, J. et al. Manifold-Regularized Multitask Fuzzy System Modeling with Low-Rank and Sparse Structures in Consequent Parameters. *IEEE Transactions on Fuzzy Systems*, v. 30, n. 5, p. 1486–1500, 2022. Disponível em: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85102279322&doi=10.1109%2FTFUZZ.2021.3062691&partnerID=40&md5=6f9ea0347d9ef5f82978dd839de7ec20>. Acesso em: 28/08/2022.