

2020. Algoritmo de Destrução em Massa por Cathy O' Neil

Grafia atualizada segundo o Acordo Ortográfico da Língua Portuguesa de 1990,  
que entrou em vigor no Brasil em 2009.

Tradução: Rafael Abraham

Revisão: Ana Helena Oliveira

Preparação: Leonardo Garzaro

Edição: Felipe Damorim e Leonardo Garzaro

Direção de Arte: Vinicius Oliveira

Imprensa: Beatriz Reingenheim

Colaboração: Laeticia Monteiro, Fernanda Mota e Danilo Abraham

Conselho Editorial: Felipe Damorim, Leonardo Garzaro, Lígia Garzaro,

Vinicius Oliveira e Ana Helena Oliveira

Dados Internacionais de Catalogação na Publicação (CIP)  
(Câmara Brasileira do Livro, SP, Brasil)

O'Neil, Cathy

Algoritmos de destruição em massa : como o big data aumenta  
a desigualdade e ameaça a democracia / Cathy O'Neil ; tradução Rafael  
Abraham. -- 1. ed. -- Santo André, SP : Editora Rua do Sabão, 2020.

Título original: Weapons of math destruction : how big data  
increases inequality and threatens democracy

ISBN 978-65-86460-02-5

1. Algoritmos 2. Antropologia - Estudo de casos 3. Acessibilidade  
4. Democracia 5. Desigualdade social 6. Redes sociais online -  
Aspectos sociais I. Abraham, Rafael. II. Título.

20-46463

CDD-305.5

Índices para catálogo sistemático:

1. Desigualdade social : Sociologia 305.5  
Aline Graziele Benitez - Bibliotecária - CRB-1/3129



Todos os direitos desta edição reservados à:  
Editora Rua do Sabão  
Rua da Fonte, 275 sala 62B  
09040-270 - Santo André, SP.

[www.editoraruadosabao.com.br](http://www.editoraruadosabao.com.br)  
[facebook.com/editoraruadosabao](https://facebook.com/editoraruadosabao)  
[instagram.com/editoraruadosabao](https://instagram.com/editoraruadosabao)  
[twitter.com/edit\\_ruadosabao](https://twitter.com/edit_ruadosabao)  
[youtube.com/editoraruadosabao](https://youtube.com/editoraruadosabao)  
[pinterest.com/editorarua](https://pinterest.com/editorarua)

# ALGORITMOS DE DESTRUIÇÃO EM MASSA

COMO O BIG DATA AUMENTA A DESIGUALDADE E AMEAÇA A DEMOCRACIA

Cathy O' Neil

Traduzido do inglês por Rafael Abraham

encontrar e manter um emprego. Todas essas esferas da vida são cada vez mais controladas por modelos secretos exercendo punições arbitrárias.

Bem-vindos ao lado sombrio do Big Data.

## COMPONENTES DA BOMBA

### O Que é um Modelo?

Era uma tarde quente de agosto em 1946. Lou Boudreau, o jogador e técnico do time de beisebol Cleveland Indians, estava tendo um dia péssimo. No primeiro jogo de um vai e volta, Ted Williams tinha aniquilado o time de Boudreau praticamente sozinho. Williams, talvez o maior rebatedor da época, havia marcado três home runs e rebatido oito bolas fortes retas. Os Indians acabaram perdendo por 11 a 10.

Boudreau precisava agir. Quando Williams entrou em campo pela primeira vez no segundo

jogo, os jogadores dos Indians começaram a se mover. Boudreau, na posição entre a segunda e terceira base, fixou-se na segunda. O homem de segunda base, por sua vez, recuou para o campo direito. O homem de terceira base moveu-se para a esquerda, na posição original de Boudreau. Ficou claro que ele, talvez por desespero, estava mudando toda a orientação de sua defesa na tentativa de fazer Williams errar.

Em outras palavras, ele estava pensando como um cientista de dados. Ele havia analisado dados brutos, a maioria por observação: Ted Williams geralmente rebatia a bola para o campo direito. Então a defesa foi reconfigurada. E funcionou. Os campistas pegaram mais bolas fortes retas de Williams do que antes (mesmo que nada pudessem fazer com os home runs que passavam voando sobre suas cabeças).

Se você for ver um jogo da Série A hoje, verá que as defesas tratam quase todos os jogadores adversários como Ted Williams. Enquanto Boudreau meramente observou para onde Williams costumava rebater a bola, os técnicos hoje sabem precisamente onde cada jogador rebateu cada bola na última semana, no último mês, ao longo de sua carreira, contra canhotos, quando já tem dois strikes e por aí vai. Usando esses dados de histórico, eles analisam a situação e calculam o posicionamento com a maior probabilidade de su-

cesso. E isso às vezes envolve mover jogadores por grandes distâncias.

Mudar a posição da defesa é só uma peça de uma questão maior: quais passos um time de beisebol pode dar para maximizar as chances de vencer? Na busca por respostas, estatísticos do jogo examinaram todas as variáveis que puderam quantificar e deram um valor a cada uma. Quanto uma rebatida dupla vale mais que uma rebatida simples? Quando, se alguma vez, é válido fazer um bunt de sacrifício para um jogador correr da primeira à segunda base?

As respostas de todas essas questões são misturadas e combinadas dentro dos modelos matemáticos daquele esporte. Esses são universos paralelos do mundo do beisebol, cada qual uma trama complexa de probabilidades. São incluídas cada relação mensurável entre cada componente único do esporte, de walks e home runs aos jogadores em si. O propósito do modelo é rodar diferentes cenários em cada conjuntura, procurando pela combinação ideal. Se o NY Yankees escalar um arremessador destro para enfrentar o rebatedor de alto aproveitamento Mike Trout, do LA Angels, em comparação a deixar em campo o atual arremessador, qual a probabilidade de eliminá-lo? E como isso irá afetar as chances globais de vencer a partida?

O beisebol é o lugar ideal para modelos matemáticos de predição. Como Michael Lewis escreveu em seu bestseller de 2003, Moneyball, por toda sua história o esporte atraiu nerds de dados. Em décadas passadas, os fãs colocavam as estatísticas no verso das figurinhas de beisebol, analisando os padrões de home runs de Carl Yastrzemski ou comparando os totais de strikeout de Roger Clemens e Dwight Gooden. Mas a partir da década de 1980, estatísticos sérios começaram a investigar o que esses números, junto de uma nova avalanche de outros, realmente significava: como se traduziam em vitórias, e como executivos poderiam maximizar o sucesso com o mínimo de dólares.

“Moneyball” é hoje sinônimo de qualquer abordagem estatística em áreas antes dirigidas por instinto. Mas o beisebol representa um saudável estudo de caso — e serve como um contraste útil aos modelos tóxicos, ou ADMs, que estão pipocando em tantas áreas de nossas vidas. Os modelos do beisebol são justos, em parte, por serem transparentes. Todo mundo tem acesso às estatísticas e consegue entender mais ou menos como são interpretadas. É claro, o modelo de um time pode dar mais valor a jogadores que fazem home runs, e outros reduzi-lo um pouco, já que rebatedores de alto aproveitamento também correm mais risco de serem eliminados. Em todo caso, os

números de home runs e strikeouts estão lá para qualquer um ver.

O beisebol tem também rigor estatístico. Seus gurus têm um imenso conjunto de dados à mão, quase na maioria das vezes diretamente relacionados à performance dos jogadores em campo. Além disso, esses dados são altamente relevantes para os resultados que estão tentando prever. Pode soar óbvio, mas como veremos ao longo deste livro, o pessoal criando ADMs habitualmente carece de dados para os comportamentos em que têm mais interesse. Então trocam por dados substitutos — ou proxies, indicadores aproximados. Eles traçam correlações estatísticas entre o CEP residencial de alguém ou padrões de linguagem e seu potencial de pagar um empréstimo ou conseguir dar conta de um emprego. Essas correlações são discriminatórias e, algumas delas, ilegais. Os modelos de beisebol, em geral, não usam proxies porque usam entradas pertinentes como bolas, strikes e acertos.

Além disso, os dados do esporte chegam de modo constante, com novas estatísticas de uma média de doze ou treze jogos chegando diariamente de abril a outubro. Os estatísticos podem comparar os resultados desses jogos com as predições de seus modelos, e podem ver onde erraram. Talvez tenham previsto que um arremessador reserva canhoto se daria mal contra um rebatedor

destro — e no entanto aconteceu o oposto. Assim, a equipe de dados precisa ajustar seu modelo e investigar o porquê do erro. A nova bola curva do arremessador afetou suas estatísticas? Ele arremessa melhor à noite? O que quer que aprendam, podem realimentar ao modelo, refinando-o. É assim que modelos de confiança operam. Mantêm um vai-e-vem constante com aquele objeto que estão tentando entender ou prever. Se as condições mudam, também muda o modelo.

Ora, você pode ver o modelo do beisebol, com seus milhares de variáveis mutáveis, e perguntar-se como poderíamos sequer usá-lo de comparação com o modelo de avaliação de professores de Washington, D.C. Em um deles, todo um esporte é modelado rigorosamente e atualizado de forma constante. No outro, apesar de coberto em mistério, parece basear-se pesadamente num punhado de notas de provas de um ano a outro. Isso é mesmo um modelo?

A resposta é sim. Um modelo, afinal de contas, nada mais é do que a representação abstrata de algum processo, seja um jogo de beisebol, a cadeia logística de uma petroleira, as ações de um governo estrangeiro, ou o público de um cinema. Esteja ele rodando dentro de um computador ou na nossa cabeça, o modelo pega o que sabemos e usa isso para prever respostas em situações variadas. Todos nós carregamos milhares de modelos

*Definição de  
modelo*

em nossas cabeças. Eles nos dizem o que esperar, e guiam nossas decisões.

Aqui vai um modelo informal que uso todos os dias. Como mãe de três, cozinho as refeições em casa — meu marido, abençoado seja, não consegue se lembrar de salgar a água do macarrão. Todas as noites, quando começo a fazer a refeição da família, intuitiva e internamente eu modelo o apetite de cada um. Sei que um filho ama frango (mas odeia hambúrguer), enquanto outro vai comer só o macarrão (com muito queijo parmesão ralado). Mas também tenho de levar em conta que os apetites das pessoas variam de um dia para outro, então uma mudança pode pegar meu modelo de surpresa. Há alguma incerteza inevitável envolvida.

As entradas, ou inputs, do meu modelo interno de culinária são as informações que tenho sobre minha família, os ingredientes que tenho às mãos ou que posso conseguir, e minha própria energia, tempo e ambição. As saídas, ou outputs, são o que e como decido cozinhar. Avalio o sucesso de uma refeição por quão satisfeita minha família parece ao terminar, quanto comeram, e quão saudável a comida era. Ver se a refeição foi bem recebida e quanto foi aproveitada me permite atualizar o modelo para a próxima vez que cozinhar. As atualizações e ajustes fazem dele o que os estatísticos chamam de “modelo dinâmico”.

Ao longo dos anos, fiquei bastante boa em fazer refeições para minha família, tenho orgulho em dizer. Mas e se meu marido e eu viajarmos por uma semana, e tenho de explicar meu sistema para minha mãe quando ela vier ajudar? Ou se minha amiga que tem filhos quiser saber meus métodos? É aí que eu começaria a formalizar meu modelo, fazendo-o muito mais sistemático e, de certo modo, matemático. E, se estivesse me sentindo ambiciosa, poderia colocá-lo dentro de um programa de computador.

Idealmente, o programa incluiria todas as opções disponíveis de comida, seu valor nutricional e preço, e uma base de dados completa dos gostos da família: cada preferência e aversão individual. Seria difícil, porém, sentar e invocar toda essa informação de cabeça. Tenho diversas lembranças de gente repetindo o aspargo ou deixando a vagem de lado. Mas são memórias embaralhadas e difíceis de formalizar numa lista detalhada.

A melhor solução seria treinar o modelo com o tempo, entrando com os dados diariamente sobre o que comprei e cozinhei e anotando as reações de cada membro da família. Também incluiria parâmetros, ou restrições. Posso limitar frutas e vegetais aos da estação e distribuir uma certa quantia de docinhos, mas só o suficiente para evitar uma franca rebelião. Também adicionaria um número de regras. Este aqui gosta de carne, este

outro de pão e macarrão, e aquele bebe muito leite e insiste em passar Nutella em absolutamente tudo.

Se eu fizer deste trabalho minha prioridade, dentro de alguns meses pode ser que eu venha a criar um modelo muito bom. Teria transformado o controle de comida que tenho na cabeça, meu modelo interno informal, em um modelo externo formal. Ao criar meu modelo, estenderia meu poder e influência sobre o mundo. Estaria criando um eu automatizado que pode ser implementado pelos outros, mesmo que eu não esteja aqui.

Sempre haveria erros, entretanto, porque modelos são, por sua própria natureza, simplificações. Nenhum modelo consegue incluir toda a complexidade do mundo real ou as nuances da comunicação humana. Inevitavelmente alguma informação importante fica de fora. Posso ter falhado em informar a meu modelo que as regras de junk food devem ser relaxadas em dias de aniversário, ou que cenouras cruas fazem mais sucesso que a cozida.

Para criar um modelo, então, fazemos escolhas sobre o que é importante o bastante para ser incluído, simplificando o mundo numa versão de brinquedo que possa ser facilmente entendida, e a partir da qual possamos inferir fatos e ações importantes. Esperamos que o modelo lide com apenas um trabalho e aceitamos que irá ocasionalmente julgar fatos e prisões de maneira errada. Seu juízo é limitado, mas é o que queremos.

nalmente agir como uma máquina ignorante com enormes pontos cegos.

Às vezes esses pontos cegos não importam. Quando pedimos o caminho no Google Maps, ele modela o mundo como uma série de ruas, túneis e pontes. Prédios são ignorados, já que não são relevantes à tarefa. Quando um software de aviação guia o avião, ele modela o vento, a velocidade e a pista de pouso, mas não ruas, túneis, prédios e pessoas.

Os pontos cegos de um modelo refletem o julgamento e prioridades de seus criadores. Ainda que as escolhas no Google Maps e nos softwares de aviação pareçam claras e inequívocas, outras são mais problemáticas. O modelo de valor agregado nas escolas de Washington, D.C., para voltar àquele exemplo, avalia os professores basicamente com as notas das provas dos alunos, à medida que ignora quanto os professores os envolvem e cativam, trabalham em tarefas específicas, lidam com o controle da sala ou os ajudam com problemas pessoais ou familiares. É por demais simples, sacrificando a precisão e critério em troca de eficiência. Mas do ponto de vista do gestor, o modelo fornece uma ferramenta efetiva de identificar centenas de supostos professores insatisfatórios, mesmo sob risco de alguns resultados falso-positivos.

Aqui vemos que os modelos, apesar de sua reputação de imparcialidade, refletem objetivos e ideologias. Quando tirei a possibilidade de comer doces em todas as refeições, estava impondo a minha ideologia ao modelo de refeições. É algo que fazemos sem pensar duas vezes. Nossos próprios valores e desejos influenciam nossas escolhas, dos dados que optamos por coletar às perguntas que fazemos. Modelos são opiniões embutidas em matemática.

Se um modelo funciona ou não também é questão de opinião. Afinal, um componente-chave de todo modelo, seja formal ou informal, é sua definição de sucesso. Este é um ponto importante ao qual voltaremos conforme exploramos o mundo obscuro das ADMs. Em cada caso, devemos nos perguntar não somente quem desenhou o modelo, mas também o que aquela pessoa ou empresa está tentando alcançar. Se o governo da Coréia do Norte criar um modelo para as refeições da minha família, por exemplo, ele poderá ser otimizado para nos manter acima do limite da fome ao menor custo possível com base no estoque de comida disponível. Gostos pessoais contariam pouco ou nada. Em contrapartida, se meus filhos estivessem criando o modelo, sucesso seria sorvete todos os dias. Meu próprio modelo tenta mesclar um pouco da gestão de recursos norte-coreana com a felicidade dos meus filhos, junto com minhas próprias prioridades de saúde, conveniência, diver-

sidade de experiência e sustentabilidade. Como resultado, é muito mais complexo. Mas ele ainda reflete minha própria realidade pessoal. É um modelo criado para hoje irá funcionar um pouco pior amanhã. Ficará obsoleto se não for atualizado de forma constante. Os preços mudam, assim como os gostos pessoais. Um modelo criado para uma criança de seis anos não funciona para um adolescente.

Isso é verdadeiro para modelos internos também. Podemos ver problemas quando avós visitam seus netos depois de algum tempo sem vê-los. Na visita anterior, reuniram informação sobre o que a criança sabe, o que a faz rir, quais programas de TV ela gosta, e criaram um modelo (inconsciente) de se relacionar com essa criança de quatro anos em particular. Ao encontrá-la um ano depois, poderão passar algumas horas desconfortáveis porque seu modelo está obsoleto. Thomas e Seus Amigos já não é mais legal. Leva-se algum tempo para agregar novos dados sobre a criança e ajustar o modelo.

Não é que bons modelos não possam ser primitivos ou rudimentares. Alguns dos mais efetivos dependem de uma única variável. O modelo mais comum de detecção de incêndio em casas e edifícios mede somente uma única variável fortemente relacionada: a presença de fumaça. Geralmente basta. Mas os modeladores se deparam com pro-

blemas — ou nos sujeitam a problemas — quando põem outros seres humanos em foco com modelos tão simples quanto um alarme de incêndio.

Racismo, no nível individual, pode ser visto como um modelo de previsão zunindo em bilhões de mentes humanas ao redor do mundo. É construído a partir de dados incompletos, defeituosos ou generalizados. Quer venha de experiência própria ou de boatos, os dados indicam que certos tipos de pessoa se comportaram mal. Isso gera uma previsão binária de que todas as pessoas daquela raça irão se comportar da mesma forma.

Desnecessário dizer, racistas não gastam muito tempo buscando dados confiáveis para treinar seus modelos deturpados. E uma vez que esse modelo se transforma em crença, torna-se programado, encucado. Ele gera suposições tóxicas, mas raramente as testa, assentando-se em dados que parecem confirmá-las e fortalecê-las. Portanto, racismo é o mais desleixado dos modelos de previsão. Ele é alimentado por coleta irregular de dados e correlações espúrias, reforçado por injustiças institucionais e contaminado por viés de confirmação. Desse modo, curiosamente, o racismo opera como muitas das ADMs que serão descritas neste livro.



Em 1997, um assassino condenado, um homem afro-americano chamado Duane Buck, ficou diante do júri no Condado de Harris, Texas. Buck havia matado duas pessoas, e o júri tinha de decidir se ele seria sentenciado à morte ou à prisão perpétua com chance de liberdade condicional. A promotora defendeu a pena de morte, argumentando que Buck poderia matar novamente caso fosse solto.

O advogado de defesa de Buck trouxe à frente uma testemunha especialista, um psicólogo chamado Walter Quijano, que não ajudou o caso da defesa nem um pouco. Quijano, que havia estudado taxas de reincidência no sistema prisional do Texas, fez referência à raça de Buck, e durante o interrogatório o promotor não deixou isso escapar.

“Você estabeleceu que (...) o fator raça, negro, aumenta a periculosidade futura por várias razões complexas. Está correto?”, perguntou o promotor.

“Sim”, Quijano respondeu. A promotora salientou esse depoimento em suas considerações finais, e o júri sentenciou Buck à morte.

Três anos depois, o Procurador-Geral do Texas, John Cornyn, descobriu que o mesmo psicólogo havia dado depoimentos similares, baseados em raça, em outros seis casos de pena capital, a

maioria deles enquanto trabalhava com a acusação. Cornyn, que seria eleito para o Senado dos EUA em 2002, requisitou novas audiências para os sete detentos — sem a questão de raça, desta vez. Em nota para a imprensa, ele declarou: “É inapropriado permitir que raça seja considerada como um fator em nosso sistema judicial. (...) O povo do Texas quer e merece um sistema proporcione imparcialidade a todos”.

Seis dos detentos obtiveram novas audiências, mas foram novamente sentenciados à morte. O depoimento prejudicial de Quijano, deliberou a corte, não havia sido decisivo. Buck não ganhou uma nova audiência, talvez porque foi sua própria testemunha quem havia levantado a questão da raça. Ele ainda aguarda no corredor da morte.

Independentemente de a questão racial aparecer ou não de forma explícita em julgamento, ela tem sido há muito um fator decisivo em sentenças. Um estudo da Universidade de Maryland mostrou que no Condado de Harris, que inclui Houston, quarta cidade mais populosa dos EUA, os promotores eram três vezes mais propensos a solicitar pena de morte para afro-americanos, e quatro vezes para hispânicos, em comparação a brancos condenados pelas mesmas acusações. Esse padrão não é exclusivo ao Texas. De acordo com a Associação Americana para Liberdades Civis (ACLU), as sentenças impostas

Raça  
como  
fator  
decisivo

## Modelos de Reincidência

em homens negros no sistema federal são cerca de 20% maiores do que condenados brancos por crimes similares. E apesar de serem apenas 13% da população, negros preenchem 40% das vagas em presídios nos EUA.

Então você pode pensar que modelos de risco computadorizados alimentados por dados reduziriam o papel do preconceito em sentenças e contribuiriam para tratamentos mais justos. Com essa esperança, tribunais em 24 estados optaram pelos assim chamados modelos de reincidência. Eles ajudam juízes a avaliar o perigo oferecido por cada condenado. E, por muitas medidas, são uma melhoria. Eles mantêm sentenças mais consistentes e menos propensas a serem influenciadas pelos humores e vieses dos juízes. Também economizam dinheiro ao encurtar a duração média das sentenças. (Custa em média 31 mil dólares ao ano para abrigar um detento, e o dobro disso em estados dispendiosos como Connecticut e New York).

A questão, no entanto, é se de fato eliminamos o viés humano ou simplesmente o camuflamos com tecnologia. Os novos modelos de reincidência são complicados e matemáticos. Mas embutidos dentro desses modelos estão um conjunto de pressupostos, alguns deles prejudiciais. E enquanto as palavras de Walter Quijano foram transcritas para o registro, que mais tarde poderia ser lido e contestado no tribunal, o funcionamen-

to de um modelo de reincidência está escondido em algoritmos, comprehensível somente para uma pequena elite.

Um dos modelos mais populares, conhecido como LSI-R, ou Level of Service Inventory – Revised, inclui um longo questionário para o detento preencher. Uma das perguntas — “Quantas condenações anteriores você já teve?” — é altamente relevante para o risco de reincidência. Outras também são claramente relacionadas: “Qual o papel outros desempenharam no delito? Qual foi o papel de drogas e álcool?”.

Mas conforme as questões continuam indo mais a fundo na vida da pessoa, é fácil imaginar como detentos de origem privilegiada iriam responder de um jeito e aqueles vindos das árduas ruas do centro da cidade de outro. Pergunte a um criminoso que cresceu num confortável bairro nobre sobre “a primeira vez em que você se envolveu com a polícia”, e ele pode não ter um único incidente a relatar além daquele que o fez ser preso. Homens negros jovens, em contrapartida, provavelmente já foram parados pela polícia dezenas de vezes mesmo sem ter feito nada de errado. Um estudo de 2013 da Associação Nova-Iorquina para Liberdades Civis apontou que enquanto homens negros e latinos entre 14 e 24 anos correspondiam a apenas 4,7% da população da cidade, eles respondem por 40,6% das paradas policiais com

revista. Mais de 90% dos revistados eram inocentes. Alguns do restante podem ter bebido antes da idade mínima legal ou levado um baseado no bolso. E, ao contrário da maioria da garotada rica, eles eram penalizados por isso. Então se “envolvimento” precoce com a polícia indica reincidência, pessoas pobres e minorias raciais parecem muito mais perigosas.

As perguntas dificilmente param por aí. Os detentos também são perguntados sobre antecedentes criminais de parentes e amigos. Novamente, faça essa pergunta a um criminoso condenado criado num bairro de classe-média, e as chances de uma resposta negativa são muito maiores. O questionário, sim, evita perguntas sobre raça, que são ilegais. Mas com a riqueza de detalhes fornecidos por cada detento, essa única pergunta ilegal é quase supérflua.

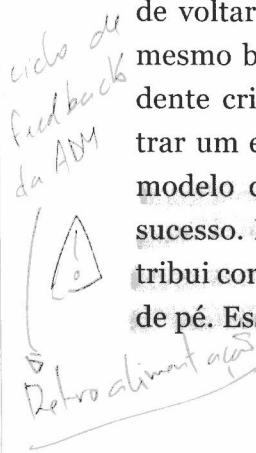
O questionário LSI-R foi dado a milhares de detentos desde sua invenção em 1995. Estatísticos usaram esses resultados para criar um sistema em que as respostas altamente correlacionadas com reincidência ganhavam mais peso e valiam mais pontos. Depois de responder ao questionário, os condenados são categorizados em risco alto, médio ou baixo com base nos pontos que acumularam. Em alguns estados, como Rhode Island, esses testes são usados somente para incluir aqueles com pontuação de alto risco em programas antir-

reincidência enquanto encarcerados. Mas em outros, incluindo Idaho e Colorado, juízes usam essa pontuação como guia em suas sentenças.

Isso é injusto. O questionário inclui circunstâncias de nascimento e criação de um criminoso, incluindo a família, vizinhança e amigos. Esses detalhes não deveriam ser relevantes a um caso criminal ou à sentença. Certamente, se um promotor tentasse manchar um réu ao mencionar os antecedentes criminais do irmão dele ou a alta taxa de crimes de seu bairro, um advogado de defesa decente urraria “protesto, Meritíssimo!” e um juiz sério manteria o protesto. Essa é a base do nosso sistema legal. Somos julgados pelo que fazemos, não por quem somos. E apesar de não sabermos quais pesos exatamente são atribuídos a essas partes do teste, qualquer peso acima de zero é irracional.

Muitos destacariam que sistemas estatísticos como o LSI-R são efetivos em medir o risco de reincidência – ou pelo menos mais precisos do que o palpite aleatório de um juiz. Mas mesmo que coloquemos de lado, por um instante, a questão crucial da imparcialidade, nos vemos caindo num nocivo ciclo de feedback da ADM. É mais provável que uma pessoa marcada como de “alto risco” esteja desempregada e venha de um bairro em que muitos de seus amigos e familiares tenham tido algum problema com a lei. Em parte

graças à alta pontuação resultante da avaliação, ele ganha uma sentença maior, prendendo-o por mais anos numa prisão em que é rodeado por colegas detentos – o que aumenta a probabilidade de voltar a ser preso. Ele é finalmente solto no mesmo bairro pobre, desta vez com um antecedente criminal, o que torna mais difícil encontrar um emprego. Se ele cometer outro crime, o modelo de reincidência pode reivindicar outro sucesso. Mas na verdade o próprio modelo contribui com um ciclo destrutivo e ajuda a mantê-lo de pé. Essa é a qualidade típica de uma ADM.



Neste capítulo, vimos três tipos de modelos. Os modelos do beisebol, em sua maioria, são saudáveis. São transparentes e continuamente atualizados, e tanto os pressupostos quanto as conclusões são claras e visíveis a todos. Os modelos se alimentam de estatísticas do jogo em questão, não de proxies. E as pessoas sendo modeladas compreendem o processo e compartilham o objetivo do modelo: vencer o campeonato mundial. (O que não significa dizer que muitos jogadores, ao término do contrato, não se queixam das valorações de um modelo: “Claro que eu fui eliminado duzentas vezes, mas olhe os meus home runs...”).

Do meu ponto de vista, certamente não há nada de errado com o segundo modelo que discutimos, o modelo hipotético de refeições familiares. Se meus filhos questionassem os pressupostos que o fundamentam, sejam financeiros ou alimentares, eu ficaria feliz em mostrá-los. E mesmo que às vezes lamentem diante de um prato com algo verde, é provável que admitam, se pressionados, que compartilham os objetivos de conveniência, economia, saúde e bom gosto — mesmo que talvez deem a cada um pesos diferentes em seus próprios modelos. (E estarão livres para criá-los quando passarem a comprar a própria comida).

Devo acrescentar que meu modelo muito dificilmente ganharia escala. Não vejo o Walmart, o Ministério da Agricultura dos EUA ou qualquer outro titã adotar meu app e impô-lo sobre centenas de milhões de pessoas, como ocorre com algumas das ADMs sobre as quais debateremos. Não, meu modelo é benigno, em especial porque provavelmente nunca vai sair da minha cabeça e ser formalizado em código de programação.

O exemplo de reincidência no fim do capítulo, entretanto, é uma outra história. Exala um cheiro conhecido e nocivo. Então façamos um rápido exercício de taxonomia de ADM para ver onde se encaixa.

Primeira pergunta: mesmo que o participante saiba que está sendo modelado, ou qual o

① O modelo é opaco ou invisível?

propósito do modelo, ele é opaco ou mesmo invisível? Bem, a maioria dos detentos preenchendo questionários obrigatórios não são idiotas. Eles ao menos têm motivos para suspeitar que as informações que fornecerem serão usadas contra eles para controlá-los quando presos, e talvez os manter lá por mais tempo. Eles conhecem o jogo. Mas os agentes penitenciários também conhecem. E não falam a respeito do propósito do questionário LSI-R. De outro modo muitos detentos tentariam manipulá-lo, dando respostas que os fizessem parecer cidadãos exemplares. Então os detentos são “mantidos no escuro” tanto quanto possível e não sabem quais são suas pontuações de risco.

Nisto, dificilmente estão sozinhos. Modelos opacos e invisíveis são a regra, e os transparentes a exceção. Somos modelados enquanto compradores e preguiçosos de sofá, enquanto pacientes médicos e requerentes de empréstimo, e disso vemos muito pouco — mesmo em aplicativos em que alegremente nos cadastramos. Mesmo quando tais modelos se comportam bem, a opacidade pode dar uma sensação de injustiça. Se ao entrar num espetáculo a céu aberto o funcionário lhe dissesse que não é permitido sentar-se nas dez primeiras fileiras, você acharia descabido. Mas se lhe fosse explicado que as dez primeiras fileiras são reservadas para pessoas com cadeira de rodas, pode fazer uma boa diferença. Transparência é importante.

Uma justiça?

E, no entanto, muitas empresas se esforçam para esconder os resultados de seus modelos ou mesmo a existência deles. Uma justificativa comum é de que o algoritmo constitui um “mílho secreto” crucial ao negócio. É propriedade intelectual, e deve ser defendida, caso necessário, com legiões de advogados e lobistas. No caso de gigantes da web como Google, Amazon e Facebook, esses algoritmos precisamente talhados valem sozinhos centenas de bilhões de dólares. As ADMs são, por projeto, caixas-pretas impenetráveis. Isso torna mais difícil ainda responder a segunda pergunta: o modelo funciona contra os interesses do sujeito? Em resumo, é injusto? Ele danifica ou destrói vidas?

Aqui, o LSI-R outra vez facilmente se qualifica como uma ADM. As pessoas que o projetaram na década de 1990 sem dúvida o viam como uma ferramenta para trazer imparcialidade e eficiência ao sistema de justiça criminal. Também poderia ajudar criminosos não perigosos a conseguir sentenças mais leves. Isso se traduziria em liberdade para eles e enorme economia para o pagador de impostos, que custeia uma conta prisional de 70 bilhões de dólares ao ano nos EUA. Contudo, por conta do questionário julgar um detento por detalhes que não seriam admitidos em tribunal, é injusto. Embora muitos possam se beneficiar disso, traz sofrimento aos demais.

③ ganha escala?

Um fator crucial desse sofrimento é o ciclo nocivo de feedback, ou retroalimentação. Como vimos, modelos de sentenças que perfilam uma pessoa pelas suas circunstâncias ajudam a criar o ambiente que justifica suas premissas. Esse ciclo destrutivo completa uma volta e depois outra, e no processo o modelo se torna mais e mais injusto.

A terceira pergunta é se um modelo tem a capacidade de crescer de forma exponencial. Como um estatístico diria, ele ganha escala? Isso pode soar como uma trivialidade nerd de um matemático. Mas escala é o que transforma ADMs de um pequeno incômodo para algo com a força de um tsunami, um que define e delimita nossas vidas. Como veremos, as ADMs sendo desenvolvidas nas áreas de recursos humanos, saúde e sistema bancário, só para citar algumas, estão rapidamente estabelecendo normas gerais que exercem sobre nós algo muito próximo do poder de lei. Se por exemplo o modelo de um tomador de empréstimo de alto risco, feito por um banco, for aplicado a você, o mundo irá tratá-lo como apenas um caloteiro — mesmo que seja uma terrível incompreensão. E quando esse modelo ganha escala, como o modelo de crédito tem ganhado, ele afeta toda a sua vida — define se você consegue ou não um apartamento, um emprego ou um carro para ir de um ao outro.

Quando se trata de escala, o potencial da modelagem de reincidência continua a crescer. Ela já

A  
- elementos de uma ADM  
- Opacidade  
- Escala  
- Dano

é usada na maioria dos estados, e o LSI-R é a ferramenta mais comum, usada em ao menos vinte e quatro deles. Além do LSI-R, os presídios abrigam um mercado vívido e movimentado para os cientistas de dados. O sistema penitenciário está repleto de dados, especialmente porque condenados desfrutam de menos direito à privacidade que o resto de nós. E mais, o sistema é tão deplorável, superpopuloso, ineficiente, caro e desumano que clama por melhorias. Quem não iria querer uma solução barata como essa?

A reforma penal é uma raridade no mundo politicamente polarizado de hoje, uma questão sobre a qual há consenso entre progressistas e conservadores. No início de 2015, os conservadores irmãos Koch, Charles e David, juntaram-se a um think tank progressista, o Center for American Progress, para defender a reforma prisional e diminuir a população carcerária. Mas minha suspeita é essa: esse esforço bipartidário para reformar presídios, junto com diversos outros, é quase certo de levar à eficiência e considerada imparcialidade de uma solução baseada em dados. É a era em que vivemos. Mesmo que outras ferramentas substituam o LSI-R como sua principal ADM, o sistema prisional provavelmente será um potente incubador de ADMs em larga escala.

Então, resumindo, esses são os três elementos de uma ADM: Opacidade, Escala e Dano. To-

dos estarão presentes, em menor ou maior grau, nos exemplos que serão abarcados aqui. Sim, haverá minúcias. Poderia-se argumentar, por exemplo, que as pontuações de reincidência não são totalmente opacas, já que dão pontuações que os detentos, em alguns casos, podem ver. Mas são repletas de mistério, já que eles não conseguem ver como suas respostas produzem a pontuação. O algoritmo de pontuação é escondido. Algumas outras ADMs podem não preencher os pré-requisitos de escala. Elas não são enormes, pelo menos não ainda. Mas representam espécies perigosas prontas para ganharem corpo, talvez de forma exponencial. Então as incluo. E, por fim, você poderá notar que nem todas essas ADMs são danosas universalmente. Afinal de contas, elas podem enviar pessoas à Harvard, arranjar-lhes empréstimos baratos ou bons empregos, e reduzir tempo de prisão de alguns réus sortudos. Mas a questão não é se alguém se sairá beneficiado. É que fazem muitos sofrer. Esses modelos, movidos por algoritmos, fecham portas na cara de milhões de pessoas, muitas vezes pelas mais frágeis das razões, e não oferecem recurso ou apelação. São injustos.

E aqui vai mais uma coisa a respeito de algoritmos: eles podem pular de uma área a outra, e muitas vezes o fazem. Pesquisa em epidemiologia pode gerar compreensões para previsão de bilheteria; filtros de spam estão sendo reorganizados para identificar o vírus da AIDS. Isso é verdadeiro

para ADMs também. Então se modelos matemáticos em presídios parecerem estar fazendo um bom trabalho — o que na verdade não passa de gerenciar pessoas de modo eficaz — eles podem se espalhar para o resto da economia junto de outras ADMs, nos deixando como dano colateral.

Esse é o meu argumento. Essa ameaça está crescendo. E o mundo das finanças nos proporciona uma história para servir de lição.