

Valores, Aplicações e Ética na Análise de Dados

Bruno Henrique Lima Vieira de Melo

Renan de Freitas Lins

O uso de dados para o aumento da assertividade e do desenvolvimento de produtos cada vez mais precisos não é novidade, porém, a forma que grandes companhias encaram essas informações, sim. Pela busca incessante de formas de extrair valor desses dados, muito desse investimento se relaciona ao suposto comportamento antiético e a quebra de privacidade dos usuários.

O escândalo Cambridge Analytica é um exemplo perfeito de um comportamento extremamente antiético, onde acaba envolvendo pessoas e privacidade. O caso da Cambridge ocorreu devido a extração de dados de mais de 50 milhões de usuários do Facebook de forma irregular. A violação foi significativa por incitar o público a discussão sobre normas éticas para empresas de mídias sociais.

Subjacente a tudo isso, precisa ser resolvido o que deve ser a ética da ciência da computação e como ela deve ser aplicada. Sistemas precisam ser construídos por um consenso entre as partes interessadas no campo, com foco entre os desenvolvedores e o público, que são os mais afetados. Isso deve ser feito com especial atenção à diversidade de representação. Na ciência da computação, mais do que qualquer outro campo, as falhas do sistema tendem a afetar as pessoas em diferentes contextos sociais (raça, gênero, classe, geografia, deficiência) de maneiras diferentes. A familiaridade com os detalhes da vida real nesses diferentes contextos é necessária para evitar o desastre.

O comportamento antiético no ramo de big data e análise de dados pode ser visto de diferentes formas. Uma delas é relacionada a discriminação algorítmica. O General Data Protection Regulation é um regulamento da União Européia que tem como objetivo discutir sobre possíveis medidas para resolver com sucesso o problema da discriminação algorítmica.

A discriminação algorítmica é um conceito que corresponde ao ato de algoritmos tomarem atitudes discriminatórias ou exclusórias em relação a seres humanos. Estas atitudes podem ser desde simples erros em detecções faciais, até a condenação de um indivíduo por algoritmos jurídicos baseado em suas características raciais.

O GDPR possui dois princípios básicos que são a sanitização de dados e transparência do algoritmo para lidar com o problema da discriminação algorítmica, porém não são todos os casos que são resolvidos com esses dois princípios. Dessa forma, uma outra potencial solução proposta é a implementação de auditorias de algoritmos, inspeções de terceiros de algoritmos, descobrindo e reduzindo os efeitos discriminatórios, onde, quando implementadas adequadamente, poderiam ex-

exercer papel crítico em tornar os algoritmos menos discriminatórios e mais responsáveis.

Algoritmos e os dados desempenham um papel cada vez mais importante nas decisões, estando presentes em quase todas as interações digitais realizadas pelo usuário. Têm-se assim uma consequência significativa para os seres humanos. Esta tendência requer maior responsabilidade na concepção e implementação de algoritmos, preocupando-se com o surgimento da discriminação algorítmica. Mas, uma vez surgida, como fazer para descobrir a responsabilidade do algoritmo?

Há muitas influências humanas embutidas nesses algoritmos. Logo eles carregam conceitos mantidos pelas pessoas que o desenvolvem. Embora a automação tente passar a imagem de que esses conceitos não poluem as saídas esperadas pelos algoritmos.

Diversas metodologias estão sendo desenvolvidas para descobrir qual é a responsabilidade de fato do algoritmo, ou quem é responsável pelo dano causado pelo mesmo. Por exemplo a citada no artigo Algorithmic Accountability Reporting, onde primeiro se identifica qual algoritmo fez a decisão errada que resultou nos danos, e o que pode estar causando essa inconsistência. Depois, o investigador, procura dados significativos para alimentar o algoritmo e observar suas saídas. E, então, busca insights, procurando comparações, seja estatística ou subjetivas.