

O Enviesamento Intrínseco e Escondido em Big Data

Nos últimos anos, o uso de *big data* vem criando raízes na cultura produtiva mundial, sendo corporações querendo atingir consumidores ou sociólogos ao estudar interações humanas. No entanto, ordenar e selecionar as melhores opções - ou as mais lucrativas - significa gerar um modelo de vencedores e perdedores. Por isso, é necessário levar em consideração como essas decisões podem desfavorecer e afetar classes de grupos humanos.

Todo e qualquer algoritmo de aprendizagem de máquina está limitado aos dados que é alimentado. Isso significa que as inferências tomadas por humanos ao final do processo está comprometida e associada com a coleta de dados iniciais, mesmo que a discriminação causada não seja intencional, intuitiva ou previsível pelos *designers*. O uso de *data mining*, se abordado com descautela, pode reproduzir padrões existentes de discriminação, e isso é crucial pois, por priorizar lucros, oportunidades são limitadas aos grupos que já possuem pouca representatividade. Ademais, porque a natureza dos algoritmos são misteriosas por definição, a tarefa de encontrar suscetibilidade das partes envolvidas mostra-se árdua.

A lei de antidiscriminação americana - particularmente as descritas sobre o *Title VII*, que descrevem práticas discriminatórias para emprego em empresas - já possui medidas para defender grupos pouco representados. No entanto, a lei falha em analisar novos casos envolvendo o uso de *data mining* e algoritmos de classificação, e algumas brechas tornam-se suscetíveis ao replicamento histórico de preconceito da sociedade ocidental. Sobre o entendimento de que o resultado das análises estão intrinsecamente conectados com os dados, é possível estudar cada etapa do algoritmo para entender e encontrar suas vulnerabilidades aos viés discriminatórios.

Cabe ao *designer* do algoritmo **classificar seus dados** sobre algum rótulo. E alguns casos, esta etapa é trivial e facilmente distinguível - por exemplo, classificar um *email* como *spam* é uma tarefa binária, *ou é ou não é*. No entanto, determinar se um candidato é apto para ser empregado, ou até se um empregado existente pode ser considerado um bom modelo, é mais abstrato. Esta etapa, inclusive, pode carregar uma discriminação que já foi utilizada previamente como medida de contratação. Então, não é possível separar o preconceito dos rótulos, mesmo que sejam tomadas medidas imparciais.

Dados desproporcionais em relação a classes também podem apresentar problemas. No caso de pouca representação, o algoritmo tenderá a favorecer objetos de uma única classe, mesmo que *outliers* atendam aos objetivos da mineração. Em certo período, a cidade de Boston enfrentava um problema com buracos em suas rodovias. Para mapear os locais de concentração, o aplicativo

StreetBump, que utiliza tecnologias como o acelerômetro e GPS dos dispositivos *smartphones*, foi desenvolvido. No entanto, a equipe de desenvolvimento falhou ao considerar que a posse de *smartphones* fosse ubíqua em todas as regiões da cidade, o que mostrou não ser o caso. Enquanto as áreas mais ricas possuíam mais população de informações, as mais pobres pouco apareciam no mapa.

É possível notar como, para ambos os casos citados acima, decisões podem ser tomadas deliberadamente ou não para que ocorra discriminação. No caso do *StreetBump*, o problema só foi encontrado uma vez que pode ser observada a disparidade no mapa da cidade, e modificar o algoritmo envolveria um grande custo. Já no caso da empregabilidade, a discriminação intrínseca pode não ser intencional por parte dos empregadores, mas devem ser levadas como suscetibilidade.

Para melhor atender o enviesamento humano, as leis precisam se adequar às tecnologias de atuais mineração de dados. Além disso, a busca pelo equilíbrio entre precisão do algoritmo sobre os dados analisados e a amenização da discriminação em *data mining* deve servir como objeto primário de estudo dos *designers* e analistas de *big data*.