Algoritmos e Discriminação

Alexsandro Vitor - avsc
Jeffson Simões - jcss3

Roteiro

- Introdução
- Definição
- Causas
- Não-Discriminação
- Medidas

Introdução

Com a intenção de agilizar, reduzir custos e aumentar a eficiência de tomadas de decisão, essas decisões têm sido automatizadas.

Porém, os algoritmos que automatizam essas decisões podem ser tão discriminatórios quanto seres humanos.

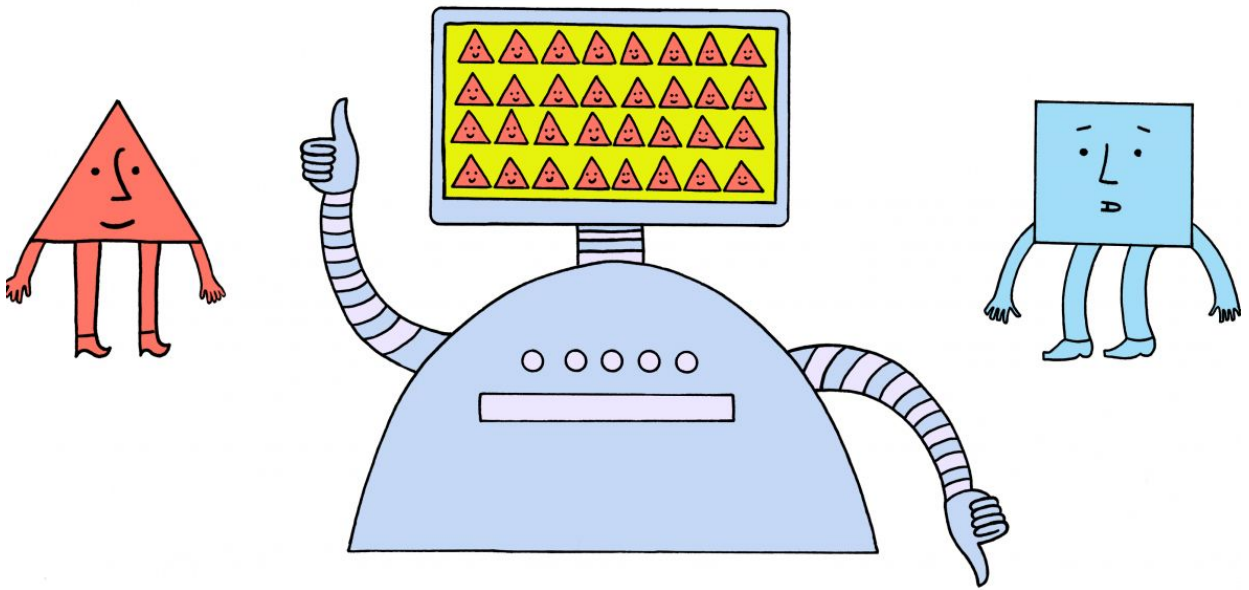
Discriminação

"Tratamento desfavorável a um certo grupo, definido por característica(s) arbitrária(s), como raça ou gênero, no lugar de características economicamente relevantes, como a produtividade."

- Discriminação Direta
- Discriminação Indireta

Discriminação

- Algoritmos Discriminatórios

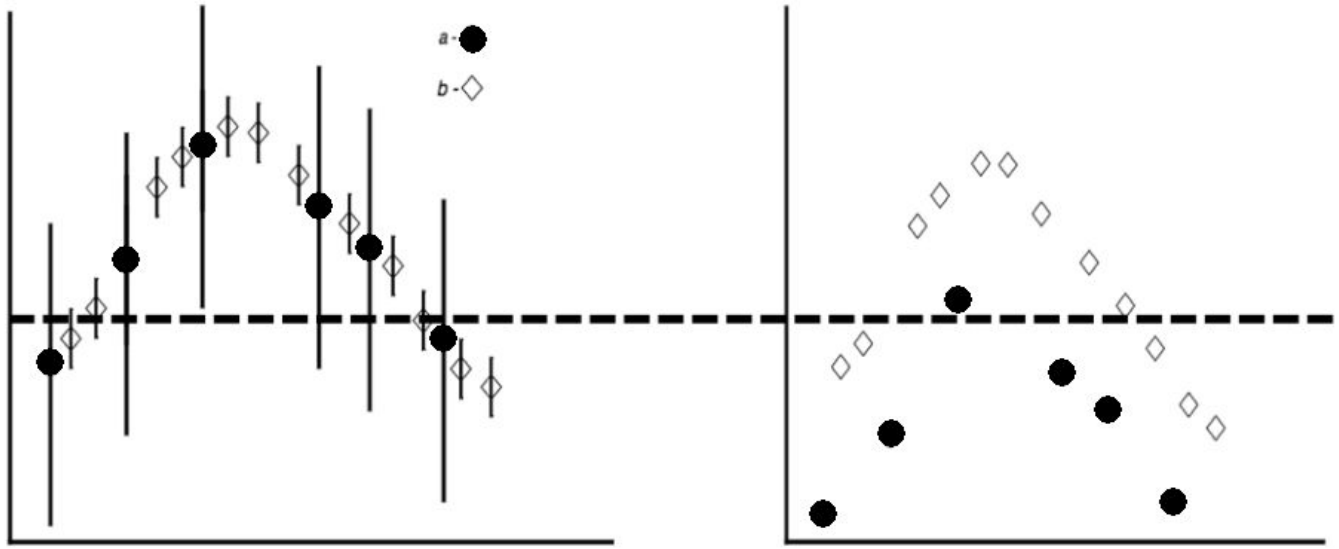


Causas - Base de dados

Algoritmos de aprendizagem procuram imitar classificações anteriores. Isso faz com eles imitem discriminações se elas existem na base de dados.

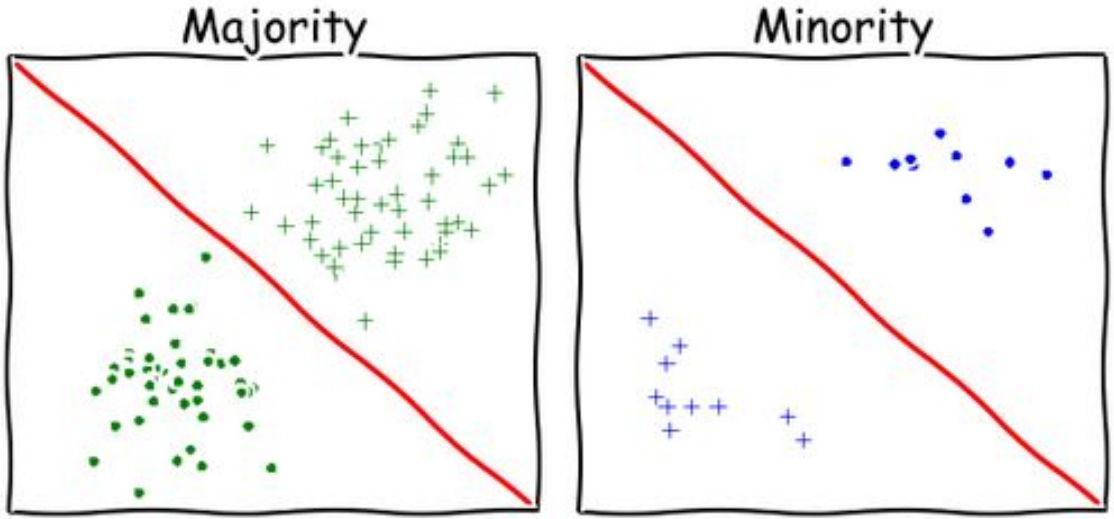
Causas - Subrepresentação

Menos informações sobre um grupo incentivam abordagens conservadoras para reduzir riscos



Causas - Subrepresentação

Relações entre variáveis podem ser diferentes em grupos diferentes



Discriminação - Humanos x Máquinas

Humanos

- Racismo, machismo, LGBTfobia, etc.
- Similaridade, "bolha" social

Máquinas

- Histórico discriminatório
- Precisão (ou falta dela)

Não-Discriminação

- Descoberta :

“A descoberta da discriminação visa encontrar padrões discriminatórios nos dados usando métodos de mineração de dados. A abordagem de mineração de dados para a descoberta da discriminação tipicamente minera as regras de associação e classificação dos dados e, em seguida, avalia essas regras em termos de potencial de discriminação.”

- Prevenção:

“O objetivo é ter um modelo (regras de decisão) que obedeça a restrições de não discriminação, que normalmente estão diretamente relacionadas à medida de discriminação selecionada. ”

Não-Discriminação

- “Pessoas que são semelhantes em termos de características não protegidas devem receber previsões semelhantes.”
- “Diferenças nas previsões entre grupos de pessoas só pode ser tão grandes quanto justificado por características não protegidas.”

Medidas de Discriminação

- Estatística:
  - “Indicam presença ou ausência de discriminação em um conjunto de dados.”
- Absoluta:
  - “Medidas absolutas capturam a magnitude da discriminação sobre um conjunto de dados levando em conta a característica protegida e a decisão de predição; nenhuma outra característica dos indivíduos é considerada”
- Condicional:
  - “As medidas condicionais capturam a magnitude da discriminação, que não pode ser explicada por nenhuma característica não protegida dos indivíduos.”
- Estrutural:
  - “As medidas estruturais não medem a magnitude da discriminação, mas a disseminação da discriminação, ou seja, uma parcela de pessoas no conjunto de dados que são afetadas pela discriminação direta.”

Perguntas?

Fim