



Big Data, Machine Learning, and the Social Sciences: Fairness, Accountability, and Transparency

Allyson Manoel - amnv



Objetivo

Justiça (fairness)

Responsabilidade (accountability)

Transparência (transparency)

O que é big data?

Um largo, diverso, complexo, longitudinal, e/ou distribuído conjunto de dados gerado a partir de instrumentos, sensores, interação com a internet, email, vídeo, fluxo de cliques, e/ou qualquer outra fonte digital disponível.

O que é big data?

É o acumulado de enormes quantidades de informações estatísticas sobre informações sociais e tendências econômicas e comportamento humano



Ciência social computacional.

Estatísticas

Ciência da computação

Ciências sociais



Dois fatores sobre big data

Granularidade

Heterogeneidade



Questions

Uma boa granularidade, talvez não seja possível de alcançar apenas utilizando técnicas computacionais

O que perguntar

Onde encontrar respostas



Data first e Method first

Bases de dados concentram um conjunto específico da sociedade

Modelos que levam em consideração majoritariamente um único subconjunto da população



Como fazer?

Priorizar as questões sociais acima da viabilidade dos dados

Dados podem não ser tão fáceis de achar

Técnicas de agregação e curadoria

Dados de melhor qualidade que permitem uma análise que reflete mais eficazmente o mundo real



Models

Predição

Explicação

Exploração



Modelo exploratório

Análise inicial dos dados

Busca por informações que poderão ser extraídas



Modelo de explanação

Explicar “o porquê” do comportamento mostrado através dos dados

Utilização de teorias sociais



Modelo de predição

Tentam predizer uma informação com base nos dados de uma base

Algo que ainda vai pra acontecer

Preencher dado faltante em um base.



Por que ligar para modelos?

Análises explanatórias e exploratórias tendem a anteceder a predição

Viés e injustiças repassadas para predição



Findings

Importância de um especialista no domínio

Cientistas sociais

Evitar estereótipos

Análise enviesada

Evitar fatores humanos no processo e na conclusão