CS 181
DISCRIMINATION

Lyndal Grant

MIT

Suppose that the age at which someone starts computer programming is strongly correlated with future success as a software engineer at Google.

On average, boys tend to start programming earlier than girls.

Q: Would it be discriminatory for Google to use the age at which someone starts programming as part of their basis for deciding which software engineers to hire?

1) Would this illegal?
2) Would this be unfair?

WHAT IS DISCRIMINATION?

Discrimination (neutral sense) vs. *wrongful* discrimination (moral sense)

Discrimination is *just one way* for a decision procedure or outcome to be unfair.

Discrimination consists of actions, practices or policies that impose a *relative disadvantage* to certain individuals based on *social group membership*.

- Discrimination is *comparative*. Duties of nondiscrimination are duties to treat people in certain ways defined by reference to the ways others are treated.

- What are the relevant social groups?
  - Eg. Race, gender, religion, LGBTQ+ status

KEY QUESTIONS

- What's wrong with discrimination?
- When is it wrong to make predictions/decisions on the basis of (protected) social group membership?
- Where might discrimination arise in ML contexts?

DISPARATE TREATMENT DISCRIMINATION

Related to *direct discrimination.*

A engages in direct racial discrimination against *B* when *A* treats *B* less favorably than she treats or would treat comparator individual(s) *C*, (partly) on the basis of *B*'s membership of racial group *G*

DISPARATE TREATMENT DISCRIMINATION

- Essentially a matter of reasons or motives.
- Need not be intentional (eg. unconscious, implicit bias).
- Need not reflect animus; may reflect indifference, or simple bias.

Q: Why is this kind of discrimination wrong?

DISPARATE IMPACT DISCRIMINATION

- Related to *indirect discrimination.*
- Imposes a disproportionate disadvantage on members of protected social groups.
- Need not involve any kind of discriminatory intent.

WHAT'S WRONG WITH DISCRIMINATION?

2 kinds of answers:

- It is *procedurally* unfair
- It is *substantively* unfair

Procedural unfairness concerns the *process* via which goods and ills are allocated to different people; substantive unfairness concerns the resulting *pattern* of distribution of these goods and ills.

WHAT'S WRONG WITH DISCRIMINATION?

Discrimination is procedurally unfair:

- It involves treating people differently on the basis of morally irrelevant characteristics
- It involves treating people differently on the basis of *group characteristics*, rather than treating them as individuals.

"There's no black male my age, who's a professional, who hasn't come out of a restaurant and is waiting for their car and somebody didn't hand them their car keys.”

Former President Barack Obama

“RACE-NEUTRAL” DATA?

PredPol:

- Used by LAPD to predict where and when crimes will occur.
- Uses “only three data points in making predictions: past type of crime, place of crime and time of crime. It uses no personal information about individuals or groups of individuals, eliminating any personal liberties and profiling concerns.”
- Trained on past police reports

Q: Is PredPol free from discrimination?

Where might discrimination arise in the use/ construction of PredPol?

“RACE-NEUTRAL” DATA?

“Even if code is modified with the aim of securing procedural fairness, however, we are left with the deeper philosophical and political issue of whether neutrality constitutes fairness in background conditions of pervasive inequality and structural injustice. Purportedly neutral solutions in the context of widespread injustice risk further entrenching existing injustices…

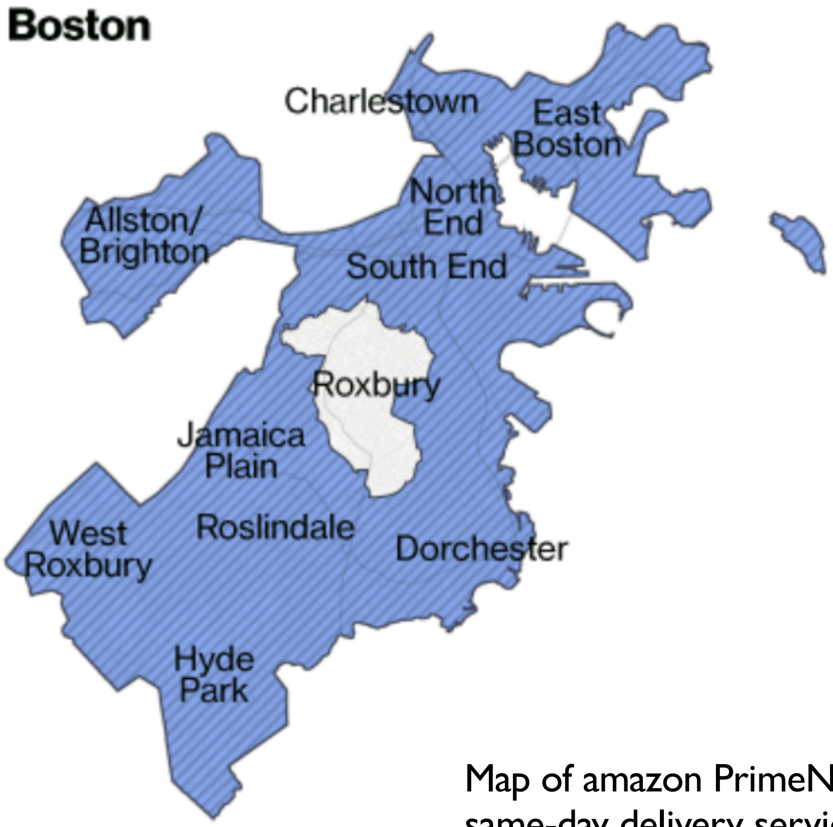
Neutral solutions might well secure just outcomes in a *just* society, but only serve to preserve the status quo in an unjust one.”

-Zimmerman, Di Rosa and Kim, “Technology Can’t Fix Algorithmic Injustice” *Boston Review*

WHAT'S WRONG WITH DISCRIMINATION? (2)

Discrimination is substantively unfair:

- contributes to the under(over)-representation of some racial groups in the most (dis)advantaged positions in society.

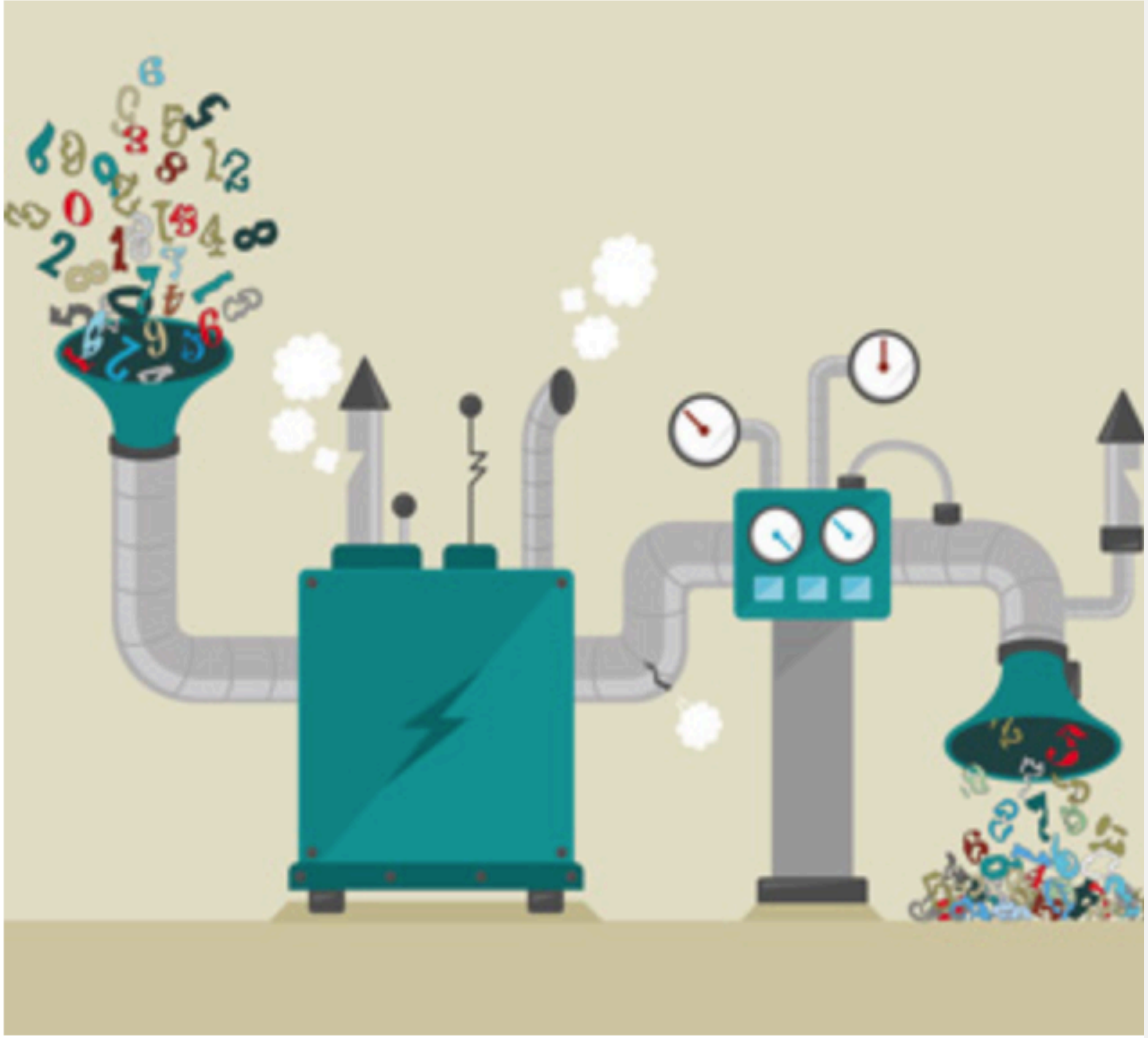


Map of amazon PrimeNow
same-day delivery service.

ACCURACY AND DISCRIMINATION

"Garbage in, garbage out."

What does this mean?



“GARBAGE IN” (WHAT IS “BIASED DATA”?)

2 ways data might be “garbage”:

- Fails to accurately represent the world.
- Accurately represents an *unjust* world.

“GARBAGE IN”

Q: How might data fail to accurately represent the world?

- Poor proxies
  - Eg. StreetBump uses *reported* potholes; PREDPOL uses *reported* crime; tech companies use referrals; NLP programs use twitter (yikes!)

Other ways?

"GARBAGE IN"

Q:What does it mean to say that data accurately represents an *unjust* world?

The data accurately captures existing patterns of prejudice/ injustice.

- Eg. Hostile workplaces like BonAppetit

“GARBAGE OUT”

What is “garbage” out?

- Bad (inaccurate) predictions
- Bad (unfair) decisions

We have epistemic reasons to avoid discrimination
when it leads to less accurate predictions

We *may* have epistemic reasons to take social group membership into account when it leads to more accurate predictions.

Eg. Healthcare.

Tempting thought: treating people fairly just requires making decisions on the basis of accurate predictions.

Accurate predictions-> fair decisions

Q: Is this right?

Accurate predictions =/= fair decisions

SURVEY

https://forms.gle/qpCCvuC7KU8VqSU39