

# Introduction to Machine Learning for Social Science

## Class 16: Algorithmic Bias

Rochelle Terman

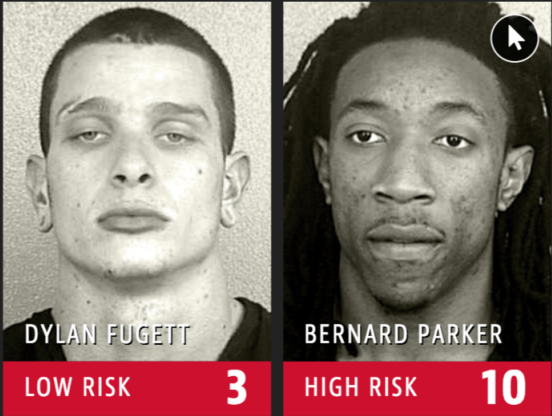
Postdoctoral Fellow  
Center for International Security Cooperation  
Stanford University

March 8, 2018

Training data mirrors social biases.

# Criminal Justice

Predictive model used predict future criminals is biased against black people.



The image displays two mugshot-style photographs side-by-side. The left photograph is of a white man with short hair, identified as Dylan Fugett. Below his photo is a red bar with the text 'LOW RISK' and a large white number '3'. The right photograph is of a black man with dreadlocks, identified as Bernard Parker. Below his photo is a red bar with the text 'HIGH RISK' and a large white number '10'. A mouse cursor icon is visible in the top right corner of the right photograph. At the bottom of the image, there is a line of italicized text.

**DYLAN FUGETT**  
**LOW RISK 3**

**BERNARD PARKER**  
**HIGH RISK 10**

*Fugett was rated low risk after being arrested with cocaine and marijuana. He was arrested three times on drug charges after that.*

# Criminal Justice

Questioned used in the risk score algorithm:

- Have you ever been arrested?
- Was one of your parents ever sent to jail or prison?
- How many of your friends/acquaintances are taking drugs illegally?
- How often did you get in fights while at school?
- How much do you agree with the following statement: ?A hungry person has a right to steal?

# Employment Discrimination

Job recruitment and hiring algorithms discriminate against protected classes, violates ADA.



Minorities under / misrepresented in data.

# Image Recognition Reproduces Sexist Stereotypes

Images of shopping and washing are linked to women, while coaching and shooting are tied to men.



# Google Translate has a Gender Problem

Google translates gender neutral languages in ways that reproduce sexist stereotypes.

**Stereotypes in Google Translate**

**Translate**

French English Turkish Detect language

He is a babysitter  
She is a doctor

English French Turkish Translate

O bir bebek bakıcısı  
O bir doktor


**Translate**


French English Turkish Detect language

O bir bebek bakıcısı  
O bir doktor

English French Turkish Translate

She's a babysitter  
He is a doctor

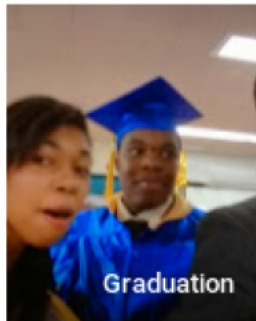
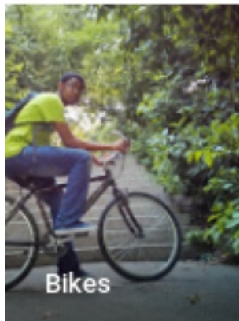
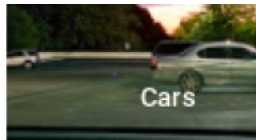
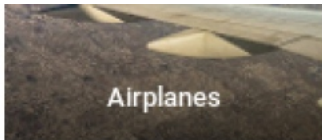
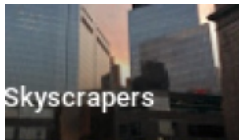
 PAGE 9 | GRACE HOPPER CELEBRATION FOR WOMEN IN COMPUTING 2017  
PRESENTED BY THE ANITA BORG INSTITUTE AND THE ASSOCIATION FOR COMPUTING MACHINERY

 #GHC17



# Google Photos Mirror Racial Stereotypes

Google Photos labeled black people 'gorillas'.



# Interaction Bias in Pokemon Go

Pokestops are overwhelmingly concentrated in majority-white neighborhoods, based on crowdsourced database of historical markers that are contributed disproportionately by young, white males,



Highly targeted data enables digital redlining.

# Price Discrimination

Uber and Amazon charge different prices based on customer's web traffic, location, purchasing habits.



# Micro-targeting

Facebook lets housing advertisers exclude users by race and other protected categories under the Fair Housing Act.

Detailed Targeting ⓘ INCLUDE people who match at least ONE of the following ⓘ

Behaviors > Residential profiles

**Likely to move**

Interests > Additional Interests

**Buying a House**

**First-time buyer**

**House Hunting**

Add demographics, interests or behaviors | **Suggestions** | Browse

Narrow Audience

EXCLUDE people who match at least ONE of the following ⓘ

Demographics > Ethnic Affinity

**African American (US)**

**Asian American (US)**

**Hispanic (US - Spanish dominant)**

Propaganda / emergent bias.

# Micro-targeting in Political Campaigns

Facebook enables advertisers to reach “Jew Haters”

