
The Greater Responsibility of Great AI Power

Roland Maio
Columbia University

Happy Hacking!



Outline

- AI for Social Good.
 - Intro to Fair Machine Learning.
 - How to Theory (or How / Theory).
-

AI For Social Good

- Data *are* biased.
- AI is not plug-and-play.
- AI cannot solve everything.

Data Are Biased.

Remember George Floyd.

Investigative Update on Critical Incident

May 26, 2020 (MINNEAPOLIS) As additional information has been made available, it has been determined that the Federal Bureau of Investigations will be a part of this investigation.

###

Man Dies After Medical Incident During Police Interaction

May 25, 2020 (MINNEAPOLIS) On Monday evening, shortly after 8:00 pm, officers from the Minneapolis Police Department responded to the 3700 block of Chicago Avenue South on a report of a forgery in progress. Officers were advised that the suspect was sitting on top of a blue car and appeared to be under the influence.

Two officers arrived and located the suspect, a male believed to be in his 40s, in his car. He was ordered to step from his car. After he got out, he physically resisted officers. Officers were able to get the suspect into handcuffs and noted he appeared to be suffering medical distress. Officers called for an ambulance. He was transported to Hennepin County Medical Center by ambulance where he died a short time later.

At no time were weapons of any type used by anyone involved in this incident.

The Minnesota Bureau of Criminal Apprehension has been called in to investigate this incident at the request of the Minneapolis Police Department.

No officers were injured in the incident.

Body worn cameras were on and activated during this incident.

The GO number associated with this case is 20-140629.

###

Source Internet Archive:
<https://web.archive.org/web/20200526121443/https://www.insidempd.com/2020/05/26/man-dies-after-medical-incident-during-police-interaction/>

Remember George Floyd.

Investigative Update on Critical Incident

May 26, 2020 (MINNEAPOLIS) As additional information has been made available, it has been determined that the Federal Bureau of Investigations will be a part of this investigation.

###

Man Dies After Medical Incident During Police Interaction

Man Dies After Medical Incident During P

May 25, 2020 (MINNEAPOLIS) On Monday evening, shortly after 8:00 pm, officers from the Minneapolis Police Department responded to the 3700 block of Chicago Avenue South on a report of a forgery in progress. Officers were advised that the suspect was sitting on top of a blue car and appeared to be under the influence.

Two officers arrived and located the suspect, a male believed to be in his 40s, in his car. He was ordered to step from his car. After he got out, he physically resisted officers. Officers were able to get the suspect into handcuffs and noted he appeared to be suffering medical distress. Officers called for an ambulance. He was transported to Hennepin County Medical Center by ambulance where he died a short time later.

At no time were weapons of any type used by anyone involved in this incident.

The Minnesota Bureau of Criminal Apprehension has been called in to investigate this incident at the request of the Minneapolis Police Department.

No officers were injured in the incident.

Body worn cameras were on and activated during this incident.

The GO number associated with this case is 20-140629.

###

Source Internet Archive:
<https://web.archive.org/web/20200526121443/https://www.insidempd.com/2020/05/26/man-dies-after-medical-incident-during-police-interaction/>

Remember George Floyd.

Investigative Update on Critical Incident

May 26, 2020 (MINNEAPOLIS) As additional information has been made available, it has been determined that the Federal Bureau of Investigations will be a part of this investigation.

###

Man Dies After Medical Incident During Police Interaction

Man Dies After Medical Incident During P

May 25, 2020 (MINNEAPOLIS) On Monday evening, shortly after 8:00 pm, officers from the Minneapolis Police Department responded to the 3700 block of Chicago Avenue South on a report of a forgery in progress. Officers were advised that the suspect was sitting on top of a blue car and appeared to be under the influence.

Two officers arrived and located the suspect in the car. After he got out, he physically resisted officers. Officers were able to get the suspect into handcuffs and noted he appeared to be suffering medical distress. Officers called for an ambulance. He was transported to Hennepin County Medical Center by ambulance where he died a short time later.

Two officers arrived and located the suspect, a male believed to be in his 40s, in his car. He was ordered to step from his car. After he got out, he physically resisted officers. Officers were able to get the suspect into handcuffs and noted he appeared to be suffering medical distress. Officers called for an ambulance. He was transported to Hennepin County Medical Center by ambulance where he died a short time later.

At no time were weapons of any type used by anyone involved in this incident.

The Minnesota Bureau of Criminal Apprehension has been called in to investigate this incident at the request of the Minneapolis Police Department.

No officers were injured in the incident.

Body worn cameras were on and activated during this incident.

The GO number associated with this case is 20-140629.

###

Source Internet Archive:
<https://web.archive.org/web/20200526121443/https://www.insidempd.com/2020/05/26/man-dies-after-medical-incident-during-police-interaction/>

Remember George Floyd.

Investigative Update on Critical Incident

May 26, 2020 (MINNEAPOLIS) As additional information has been made available, it has been determined that the Federal Bureau of Investigations will be a part of this investigation.

###

Man Dies After Medical Incident During Police Interaction

Man Dies After Medical Incident During P

May 25, 2020 (MINNEAPOLIS) On Monday evening, shortly after 8:00 pm, officers from the Minneapolis Police Department responded to the 3700 block of Chicago Avenue South on a report of a forgery in progress. Officers were advised that the suspect was sitting on top of a blue car and appeared to be under the influence.

Two officers arrived and located the suspect, a male believed to be in his 40s, in his car. He was ordered to step from his car. After he got out, he physically resisted officers. Officers were able to get the suspect into handcuffs and noted he appeared to be suffering medical distress. Officers called for an ambulance. He was transported to Hennepin County Medical Center by ambulance where he died a short time later.

Two officers arrived and located the suspect, a male believed to be in his 40s, in his car. He was ordered to step from his car. After he got out, he physically resisted officers. Officers were able to get the suspect into handcuffs and noted he appeared to be suffering medical distress. Officers called for an ambulance. He was transported to Hennepin County Medical Center by ambulance where he died a short time later.

At no time were weapons of any type used by anyone involved in this incident.

The Minnesota Bureau of Criminal Apprehension has been called in to investigate this incident at the request of the Minneapolis Police Department.

No officers were injured in the incident.

Body worn cameras were on and activated during this incident.

The GO number associated with this case is 20-140629.

###

Source Internet Archive:
<https://web.archive.org/web/20200526121443/https://www.insidempd.com/2020/05/26/man-dies-after-medical-incident-during-police-interaction/>

MINNEAPOLIS

Minneapolis police cite 'fluid' situation for troubling misinformation released after George Floyd death

The most credible accounts of what happened that night came from bystander video and private surveillance footage.

By Andy Mannix Star Tribune | JUNE 3, 2020 — 5:08AM

GALLERY GRID

1/17



AARON LAVINSKY - STAR TRIBUNE

Gallery: A memorial shown Tuesday outside the Cup Foods store at 38th Street and S. Chicago Avenue where George Floyd was killed while in police custody.

AI Is Not Plug-And-Play

No Free Lunch

AI Systems embody assumptions and value judgements.

New Criminal Activity (maximum total weight = 13 points)	
Age at current arrest	23 or older = 0; 22 or younger = 2
Pending charge at the time of the offense	No = 0; Yes = 3
Prior misdemeanor conviction	No = 0; Yes = 1
Prior felony conviction	No = 0; Yes = 1
Prior violent conviction	0 = 0; 1 or 2 = 1; 3 or more = 2
Prior failure to appear pretrial in past 2 years	0 = 0; 1 = 1; 2 or more = 2
Prior sentence to incarceration	No = 0; Yes = 2

What are the assumptions here?

Public Safety Assessment: Risk Factors and Formula. Laura and John Arnold Foundation.

No Free Lunch

AI Systems embody assumptions and value judgements.

New Criminal Activity (maximum total weight = 13 points)	
Age at current arrest	23 or older = 0; 22 or younger = 2
Pending charge at the time of the offense	No = 0; Yes = 3
Prior misdemeanor conviction	No = 0; Yes = 1
Prior felony conviction	No = 0; Yes = 1
Prior violent conviction	0 = 0; 1 or 2 = 1; 3 or more = 2
Prior failure to appear pretrial in past 2 years	0 = 0; 1 = 1; 2 or more = 2
Prior sentence to incarceration	No = 0; Yes = 2

What are the assumptions here?

What are the value judgements?

Public Safety Assessment: Risk Factors and Formula. Laura and John Arnold Foundation.

AI Redistributes Power

[Digital Services and Device Support](#) › [Alexa Features Help](#) › [Shopping with Alexa](#) ›

Place Orders with Alexa

Ask Alexa to place orders for products from your order history or your Amazon cart.

1. Say, "Order [item]." Alexa adds a selection for that item to your cart. Follow the prompts to place your order. Some items, require completion using the Amazon app or on the Amazon website.
2. Or, say, "Checkout my cart," and follow the prompts to check out your order.

If you no longer want the order, say "Cancel my order."

Was this information helpful?

Yes

No

AI Redistributes Power

[Digital Services and Device Support](#) › [Alexa Features Help](#) › [Shopping with Alexa](#) ›

Place Orders with Alexa

Ask Alexa to place orders for products from your order history or your Amazon cart.

1. Say, "Order [item]." Alexa adds a selection for that item to your cart. Follow the prompts to place your order. Some items, require completion using the Amazon app or on the Amazon website.
2. Or, say, "Checkout my cart," and follow the prompts to check out your order.

If you no longer want the order, say "Cancel my order."

Was this information helpful?

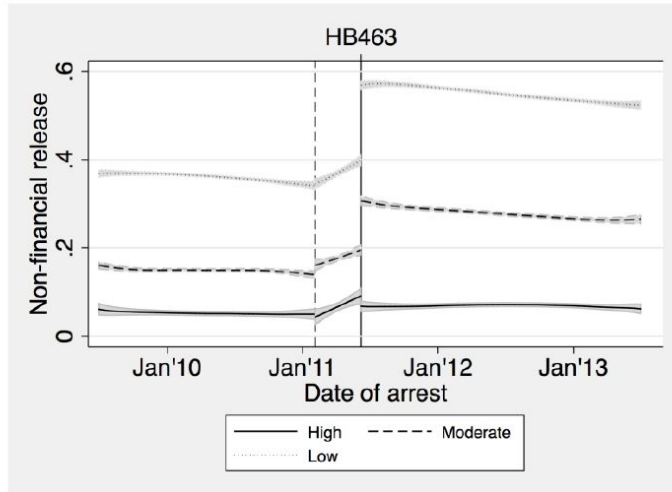
Yes

No

How does Alexa choose which sellers to buy from?

AIs (Don't) Change the System

Figure 2 - How HB 463 Affected Non-financial Release Rates for Defendants at Different Risk Levels



Note: The top, middle, and bottom line indicate the fraction of low, moderate, and high-risk defendants who are granted non-financial release. The dashed vertical line is the date that HB 463 was introduced as legislation; the solid line indicates the date it was implemented.

In 2011, the state of Kentucky passed a law that mandated judges consider risk assessments (read AIs) in their pretrial decision-making which set presumptive default decisions but did not override judge discretion.

AI Cannot Solve Everything

Human vs AI

SHARE

RESEARCH ARTICLE | RESEARCH METHODS



The accuracy, fairness, and limits of predicting recidivism

Julia Dressel and Hany Farid*

+ See all authors and affiliations

Science Advances 17 Jan 2018:
Vol. 4, no. 1, eaao5580
DOI: 10.1126/sciadv.aao5580

Article

Figures & Data

Info & Metrics

eLetters

PDF

Abstract

Algorithms for predicting recidivism are commonly used to assess a criminal defendant's likelihood of committing a crime. These predictions are used in pretrial, parole, and sentencing decisions. Proponents of these systems argue that big data and advanced machine learning make these analyses more accurate and less biased than humans. We show, however, that the widely used commercial risk assessment software COMPAS is no more accurate or fair than predictions made by people with little or no criminal justice expertise. In addition, despite COMPAS's collection of 137 features, the same accuracy can be achieved with a simple linear classifier with only two features.

Mean Accuracy of Human Beings:

Mean Accuracy of COMPASS:

Human vs AI

SHARE

RESEARCH ARTICLE | RESEARCH METHODS



The accuracy, fairness, and limits of predicting recidivism

Julia Dressel and Hany Farid*

+ See all authors and affiliations

Science Advances 17 Jan 2018;
Vol. 4, no. 1, eaao5580
DOI: 10.1126/sciadv.aao5580

Article

Figures & Data

Info & Metrics

eLetters

PDF

Abstract

Algorithms for predicting recidivism are commonly used to assess a criminal defendant's likelihood of committing a crime. These predictions are used in pretrial, parole, and sentencing decisions. Proponents of these systems argue that big data and advanced machine learning make these analyses more accurate and less biased than humans. We show, however, that the widely used commercial risk assessment software COMPAS is no more accurate or fair than predictions made by people with little or no criminal justice expertise. In addition, despite COMPAS's collection of 137 features, the same accuracy can be achieved with a simple linear classifier with only two features.

Mean Accuracy of Human Beings: 62.1%

Mean Accuracy of COMPASS:

Human vs AI

SHARE

RESEARCH ARTICLE | RESEARCH METHODS



The accuracy, fairness, and limits of predicting recidivism

Julia Dressel and Hany Farid*

+ See all authors and affiliations

Science Advances 17 Jan 2018:
Vol. 4, no. 1, eaao5580
DOI: 10.1126/sciadv.aao5580

Article

Figures & Data

Info & Metrics

eLetters

PDF

Abstract

Algorithms for predicting recidivism are commonly used to assess a criminal defendant's likelihood of committing a crime. These predictions are used in pretrial, parole, and sentencing decisions. Proponents of these systems argue that big data and advanced machine learning make these analyses more accurate and less biased than humans. We show, however, that the widely used commercial risk assessment software COMPAS is no more accurate or fair than predictions made by people with little or no criminal justice expertise. In addition, despite COMPAS's collection of 137 features, the same accuracy can be achieved with a simple linear classifier with only two features.

Mean Accuracy of Human Beings: 62.1%

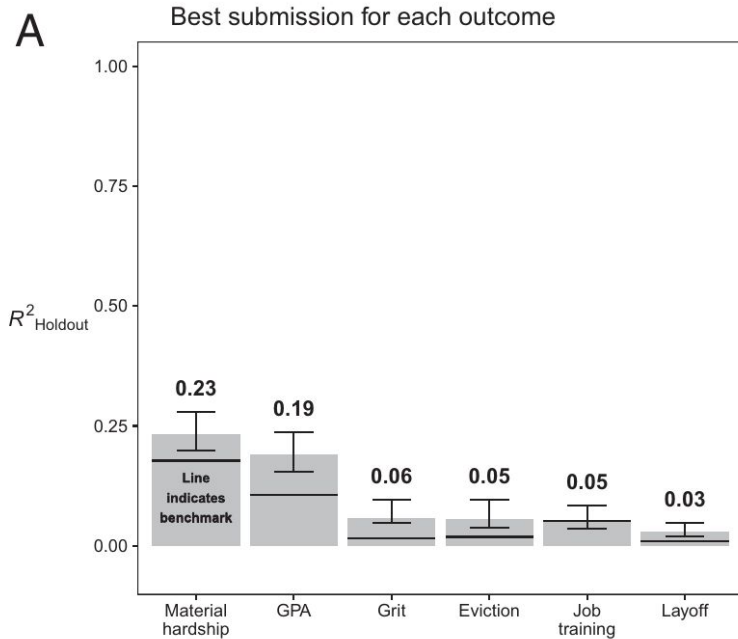
Mean Accuracy of COMPASS: 65.2%

(Not) Predicting Life Outcomes

How predictable are life outcomes? In a recent study, more than 100 teams of top machine learning researchers applied their art to predicting the life outcomes of children using high quality longitudinal data collected by sociologists.

$$R_{\text{Holdout}}^2 = 1 - \frac{\sum_{i \in \text{Holdout}} (y_i - \hat{y}_i)^2}{\sum_{i \in \text{Holdout}} (y_i - \bar{y}_{\text{Training}})^2}.$$

(Not) Predicting Life Outcomes



$$R^2_{\text{Holdout}} = 1 - \frac{\sum_{i \in \text{Holdout}} (y_i - \hat{y}_i)^2}{\sum_{i \in \text{Holdout}} (y_i - \bar{y}_{\text{Training}})^2}.$$

Source: *Measuring the predictability of life outcomes with a scientific mass collaboration*. Salganik et al.

Can Good Prediction be Bad?

Imagine that we lived in a utopia.

Would everything be predictable?

Consider one ideal for social mobility: the socioeconomic status of the family a person is born into should not determine their adult socioeconomic status.

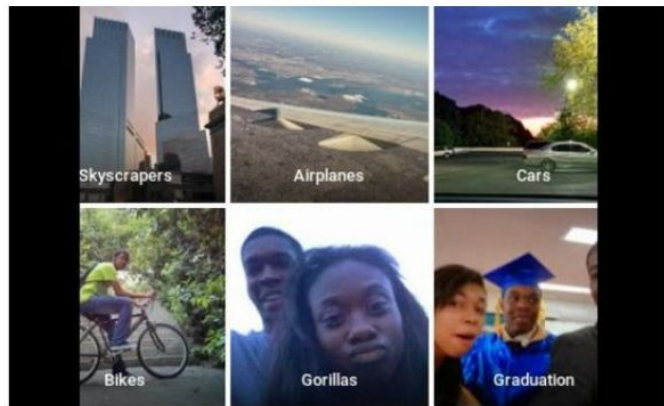
This is fundamentally an *unpredictability* condition.

Intro to Fair Machine Learning

- Motivation
- Fairness Formalized

Motivation

Discrimination in ML Systems



diri noir avec banan @jackyalcine · Jun 29
Google Photos, y'all [redacted] My friend's not a gorilla.




813 394

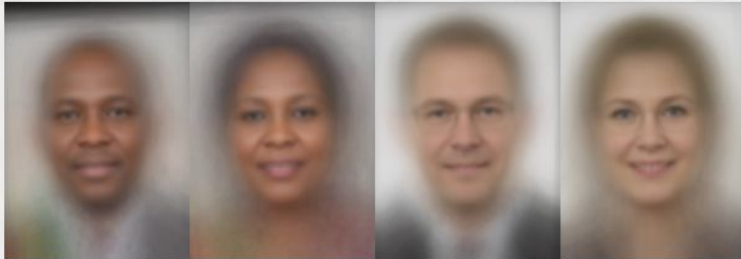
TWITTER

Mr Alcine tweeted Google about the fact its app had misclassified his photo

Source: *Google apologizes for Photos app's racist blunder*, BBC, 2015.

Discrimination in ML Systems

Gender Classifier	Darker Male	Darker Female	Lighter Male	Lighter Female	Largest Gap
 Microsoft	94.0%	79.2%	100%	98.3%	20.8%
 FACE++	99.3%	65.5%	99.2%	94.0%	33.8%
 IBM	88.0%	65.3%	99.7%	92.9%	34.4%






Mr Alcine twee

Source: Go

Source: The Gender Shades Project, gendershades.org.

Discrimination in ML Systems

Gender Classifier	Darker Male	Darker Female	Lighter Male	Lighter Female	Largest Gap
 Microsoft	94.0%	79.2%			
 FACE++	99.3%	65.5%			
 IBM	88.0%	65.3%			



Amazon's Face Recognition Falsely Matched 28 Members of Congress With Mugshots

Jul 26, 2018



By: Jacob Snow [@snowjake](#)

Amazon's face surveillance technology is the target of growing opposition nationwide, and today, there are 28 more causes for concern. In a test the ACLU recently conducted of the facial recognition tool, called "Rekognition," the software incorrectly matched 28 members of Congress, identifying them as other people who have been arrested for a crime.



Source: ACLU.

Source: The Gender Shades Project, gendershades.org.

Mr Alcine twee

Source: Go

Discrimination in ML Systems

Gender Classifier	Darker Male	Darker Female	Lighter Male	Lighter Female	Largest Gap
Microsoft	94.0%	79.2%			
FACE++	99.3%	65.5%			
IBM	88.0%	65.3%			



Amazon's face surveillance tech is the target of growing opposition nationwide, and today, there are many causes for concern. In a test that was recently conducted of the facial recognition tool, called "Rekognition," the system incorrectly matched 28 members of Congress, identifying them as individuals who have been arrested for a crime.

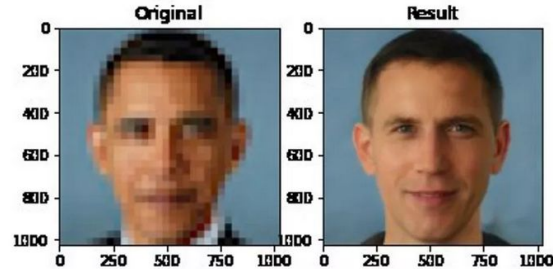
Source: ACLU.

Amazon's Face Recognition Falsely Matched 28 Members of Congress What a machine learning tool that turns Obama white can (and can't) tell us about AI bias

A striking image that only hints at a much bigger problem

By James Vincent | Jun 23, 2020, 3:45pm EDT

f t SHARE



Source: The Verge.

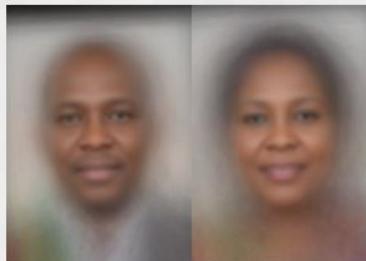
Source: The Gender Shades Project, gendershades.org.

Mr Alcine tweet

Source: Go

Discrimination in ML Systems

Gender Classifier	Darker Male	Darker Female	Lighter Male	Lighter Female	Largest Gap
Microsoft	94.0%	79.2%			
FACE++	99.3%	65.5%			
IBM	88.0%	65.3%			



Amazon's face surveillance tech is the target of growing opposition nationwide, and today, there are several causes for concern. In a test recently conducted of the facial recognition tool, called "Rekognition," the software incorrectly matched 28 members of Congress, identifying them as suspects who have been arrested for a crime.

Source: ACLU.

Amazon's Face Recognition Falsely Matched 28 Members of Congress

What a machine learning tool that turns Obama white can (and can't) tell us about facial recognition

'The Computer Got It Wrong': How Facial Recognition Led To False Arrest Of Black Man

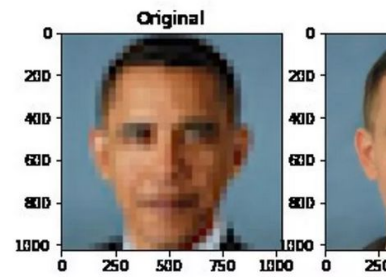
By James Vincent | Jun 23, 2020, 3:45pm EDT

f t SHARE

June 24, 2020 - 8:00 AM ET

BOBBY ALLYN

3-Minute Listen + PLAYLIST



Source: The Verge.

MICHIGAN STATE POLICE
INVESTIGATIVE LEAD REPORT
LAW ENFORCEMENT SENSITIVE

THIS DOCUMENT IS NOT A POSITIVE IDENTIFICATION. IT IS AN INVESTIGATIVE LEAD ONLY AND IS NOT PROBABLE CAUSE TO ARREST. FURTHER INVESTIGATION IS NEEDED TO DEVELOP PROBABLE CAUSE TO ARREST.

BID DIA Identifier: BID-39841-19	Requester: CA Yager, Rathe
Date Searched: 03/11/2019	Requesting Agency: Detroit Police Department
Digital Image Examiner: Jennifer Coulson	Case Number: 1810050167
	File Class/Crime Type: 3000

Michigan State Police ran a facial recognition search for a suspect in Detroit. It suggested the 42-year-old Robert Williams was the suspect. He was arrested and detained. He and his lawyers say the algorithm failed and mistakenly identified him as someone else. Prosecutors have dismissed the case.

Detroit Police Department Incident Report

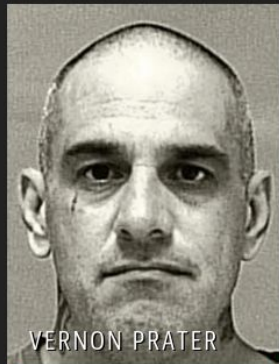
Source: NPR.

Mr Alcine tweet
Source: Go

Source: The Gender Shades Project, gendershades.org.

More Than Biased Data

Two Petty Theft Arrests



VERNON PRATER

LOW RISK

3



BRISHA BORDEN

HIGH RISK

8

Borden was rated high risk for future crime after she and a friend took a kid's bike and scooter that were sitting outside. She did not reoffend.

Source: *Machine Bias*, Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner, ProPublica, 2016.

More Than Biased Data

Two Petty Theft Arrests

Two Petty Theft Arrests

VERNON PRATER

Prior Offenses

2 armed robberies, 1
attempted armed
robbery

Subsequent Offenses

1 grand theft

LOW RISK

3

BRISHA BORDEN

Prior Offenses

4 juvenile
misdemeanors

Subsequent Offenses

None

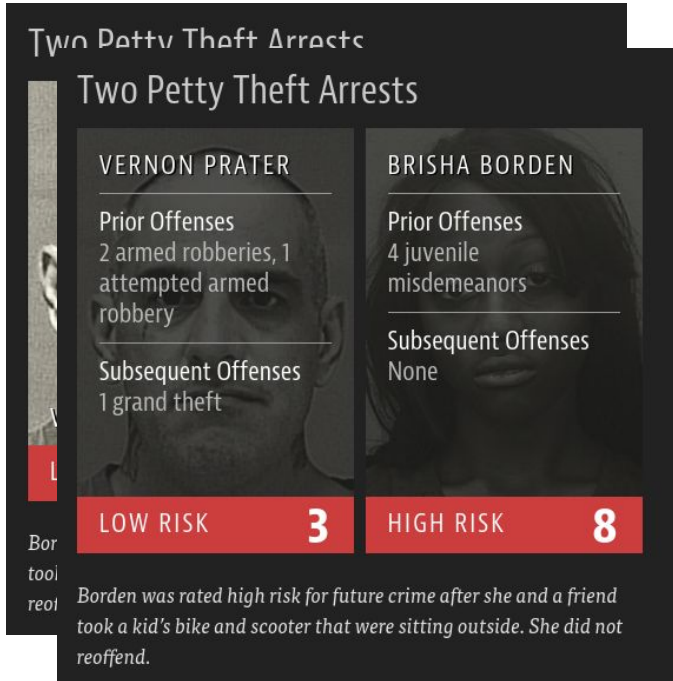
HIGH RISK

8

Bor
too
reot

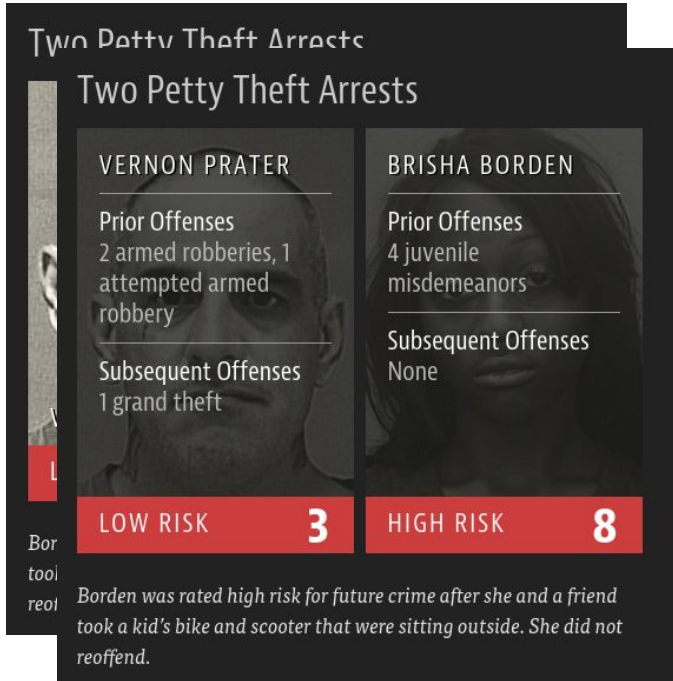
Borden was rated high risk for future crime after she and a friend took a kid's bike and scooter that were sitting outside. She did not reoffend.

More Than Biased Data



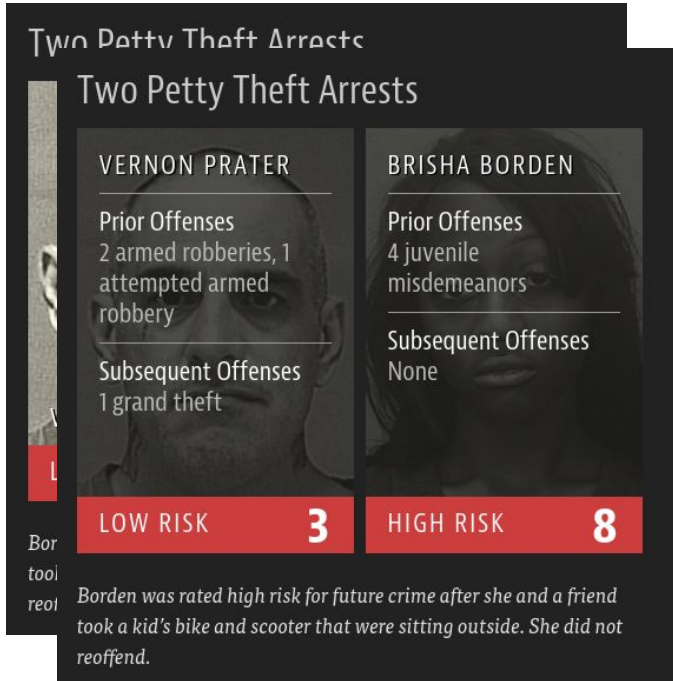
- Equally Calibrated! But...

More Than Biased Data



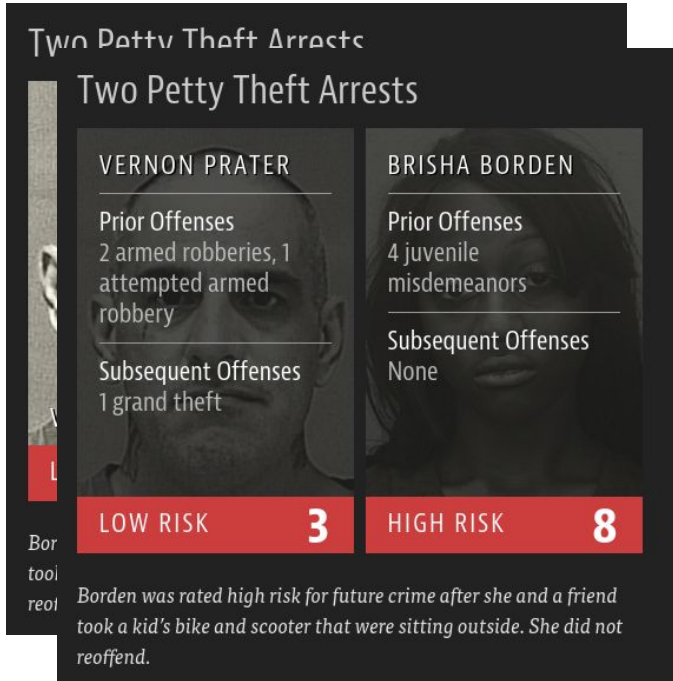
- Equally Calibrated! But...
- False Positive Rate for White Defendants: 23%

More Than Biased Data



- Equally Calibrated! But...
- False Positive Rate for White Defendants: 23%
- False Positive Rate for Black Defendants: 45%.

More Than Biased Data



- Calibration: 50% of the people assigned a risk score of 50% actually recidivate.
- Equal False Positives: The fraction of non-recidivists predicted to be high risk is the same in both groups.
- Equal False Negatives: The fraction of recidivists predicted to be low risk is the same in both groups.

More Than Biased Data



- Calibration: ... a risk

Inherent Trade-Offs in the Fair Determination of Risk Scores

Jon Kleinberg *

Sendhil Mullainathan †

Manish Raghavan ‡

Calibration Negatives: The fraction of recidivists predicted to be low risk is the same in both groups.

Formalizing Fairness

Putting Fair in ML

- Asking algorithms to be fair requires precise definitions of what we mean by fair.
 - Many possible definitions! You have seen three.
 - Two dominant classes of definitions: *Group fairness* and *individual fairness*.
 - Open Question: Is there a universal definition?
 - Can fairness be formalized?
-

Demographic Parity

Formal Setup:

Individuals are represented by triples (x, y, a) : A feature vector x , a target variable y , and a group membership variable a .

A binary classifier C .

A set of groups G .

Prominent notion of group fairness. We'll see this again!

Intuition: A machine-learning predictor should assign the "same" outcomes to each group.

A binary classifier C , for a set of groups G , is said to *satisfy demographic parity* if for every two groups g, g' from G it holds that: $\Pr[C(x) = 1 \mid a = g] = \Pr[C(x') = 1 \mid a = g']$.

Individual Fairness

Formal Setup:

Individuals are represented by 2-tuples (x, y) : A feature vector x , and a target variable y .

A randomized binary classifier C .

A task-specific similarity metric d : $d(x, x')$ is how similar individual x is to x' .

A distance metric D over probability distributions.

Intuition: "Similar" individuals should be treated "similarly."

A randomized binary classifier C is said to *satisfy individual fairness* if for every pair of individuals x, x' it holds that:

$$D(C(x), C(x')) \leq d(x, x').$$

How to Theory

(or How I Theory)



Background and Motivation

Cost of Fairness

The engine of machine learning is mathematical optimization.

Requiring that a machine-learning predictor satisfy a suitably chosen notion of fairness constrains the underlying machine-learning task, which in general imposes a cost to the *optimization objective*.

This should not *necessarily* be understood as reflecting a cost in reality.

Fair Representations: Problem

Solution concept in group fairness.

A data publisher has a dataset of individuals $D = \{(x_i, y_i, z_i)\}$. and wishes to release data to a data consumer for machine learning. The publisher wants to make sure that the data consumer will be fair in the sense of satisfying demographic parity.

What can the publisher do?

Formal Setup:

Individuals are represented by triples (x, y, z) : A feature vector x , a target variable y , and a group membership variable z .

A binary classifier C .

A set of groups G .

Fair Representations: Solution

Formal Setup:

Individuals are represented by triples (x, y, z) : A feature vector x , a target variable y , and a group membership variable a .

Fair-representation space Z .

Transformation r that maps feature vectors in X to points in Z .

An adversarial data consumer will exploit demographic information in the data to discriminate: Release a *transformed* dataset D' that removes the demographic information!

Z is said to be a *fair representation* under transformation r if for every point z in Z , and group g in G , it holds that:

$$\Pr[a = g \mid z] = \Pr[a = g]$$

We call this condition *demographic secrecy*.

Observations

Demographic secrecy seems like a pretty strong property.

Data sale, resale, and reuse is important. Companies collect, buy and sell tons of data every day.

Is there something here?

The Weeds of Theoretical Research

Ruining the Punchline

- We discovered a whole new *cost of demographic secrecy* that is distinct from and in addition to the cost of fairness.
 - The cost of demographic secrecy only occurs when data are *reused*.
-

Goal

See if there is a there, there. And if there is, find out if there is anything interesting about it.

The Method of the Madness

Build a model for the phenomenon you are studying.

Analyze the model for interesting consequences.

Modeling Fair Representations

Formal Setup:

Individuals are represented by triples (x, y, z) : A feature vector x , a target variable y , and a group membership variable a .

Fair-representation space Z .

Transformation r that maps feature vectors in X to points in Z .

There is a large literature on fair representations which focuses on *learning the transformation*.

How do you build a model that applies to the entire literature?

At least two moves: 1) Generalize and abstract. 2) Focus on commonalities.

Focus on the Commonalities

Fair Representations Formal Setup:
Individuals are represented by triples (x, y, a) : A feature vector x , a target variable y , and a group membership variable a .

Fair-representation space Z .

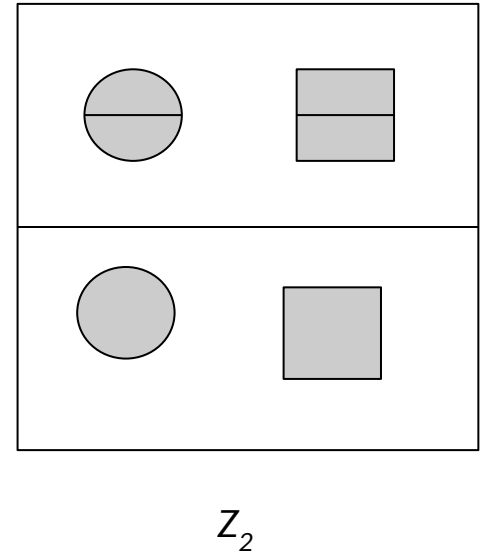
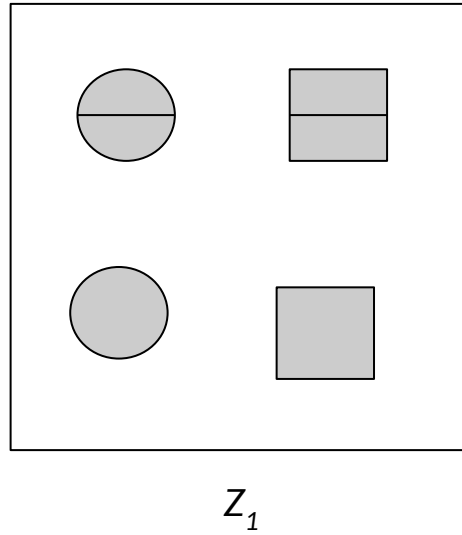
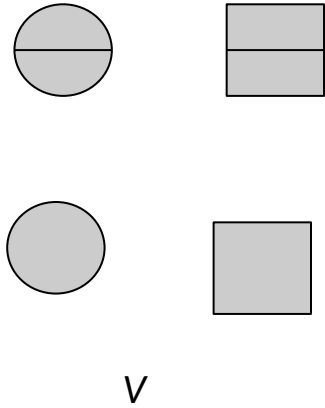
Transformation r that maps feature vectors in X to points in Z .

Project Formal Setup:
Each individual is represented by an element v of a finite set V , has group $\gamma(v)$, given by group membership function γ , class $f(v)$ given by a binary class-membership function f .

A representation is a partition Z of V .

A transformation r is a function that maps individuals in V to parts in a partition Z .

Model Example



I'm Not Crazy! Check Consistency

What is demographic secrecy in our model: Does it make sense? One piece of notation, for any set of individuals S , denote by S_g the set of all individuals in S belonging to group g .

A representation (i.e. partition) Z satisfies demographic parity if for every z in Z , and every group g in G it holds that:

$$|z_g| / |z| = |V_g| / |V|$$

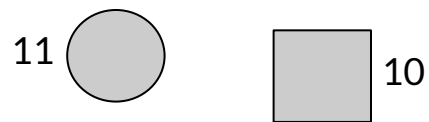
First Result: No Reuse, No Cost.

Think about the classifier C that achieves the maximum accuracy possible on predicting f *while* also satisfying demographic parity.

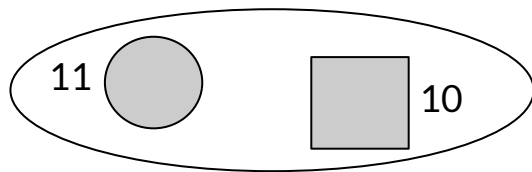
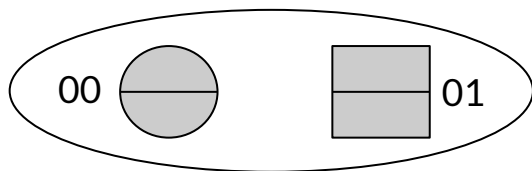
Eureka: C partitions V !

In theory, ignoring all other concerns, a data publisher can find a demographically secret representation that has a cost equal to the cost of fairness.

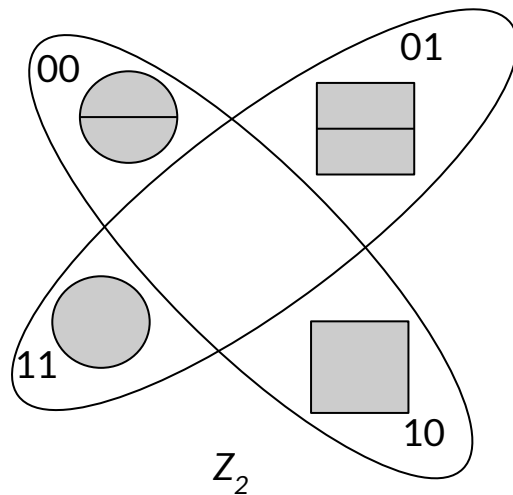
Second Result: Reuse is Costly



V



Z_1



Z_2

Questions for Me?

And!

A Question for you :)

What is the greater responsibility?

Contact: roland@cs.columbia.edu
