Data Science For Social Good

Summer Fellowship

Center for Data Science & Public Policy

THE UNIVERSITY OF CHICAGO

Fair and Equitable Decision Making with AI

Pedro Saleiro

DSPT

DATA SCIENCE PORTUGAL

DSPT #55

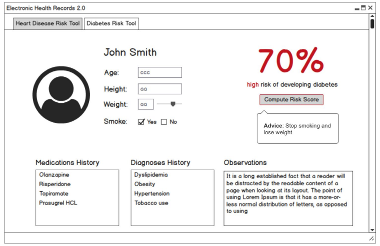
60+ Projects

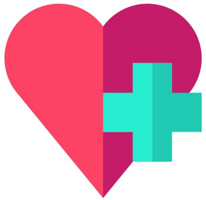
IDHS Illinois Department of Human Services
FEEDING AMERICA
Department of Labor United States of America
VANCOUVER PUBLIC SCHOOLS VPS
NOBLE
WAKE COUNTY PUBLIC SCHOOL SYSTEM
Perspectives Charter Schools
KIPP New Jersey
KIPP Chicago COLLEGE PREP PUBLIC SCHOOLS
INFONAVIT Hogares con valor
cclba
city of CINCINNATI PERFORMANCE & DATA ANALYTICS
Montgomery County Public Schools MCPS
CHICAGO PUBLIC SCHOOLS CPS
POLICE CHARLOTTE-MECKLENBURG
THE WHITE HOUSE WASHINGTON
CHICAGO ALLIANCE TO END HOMELESSNESS
INSTITUTE FOR HOUSING STUDIES
ChapinHall at the University of Chicago
INSTITUTO DO EMPREGO E FORMAÇÃO PROFISSIONAL
City of Memphis
WORLD BANK GROUP
CHICAGO HARRIS PUBLIC POLICY | THE UNIVERSITY OF CHICAGO
City of Chicago Incorporated 4th March 1837
cta
Police City of Chicago
Seal of Cook County Illinois
BERKELEY LAB Lawrence Berkeley National Laboratory
San Francisco Police
METRO POLICE NASHVILLE DAVIDSON COUNTY
Get Covered Illinois The Official Health Marketplace
AGENTIS energy
AUSTRALIAN CONSERVATION FOUNDATION
THE CASE FOUNDATION
EPIC | ENERGY POLICY INSTITUTE AT THE UNIVERSITY OF CHICAGO
Pecan Street Inc.
Arlington Public Schools
SUNLIGHT FOUNDATION
Health Leads
MÉXICO PRESIDENCIA DE LA REPÚBLICA
معهد قطر لبحوث الحوسبة
Qatar Computing Research Institute
City of Milwaukee
Wiki Energy
mesa PUBLIC SCHOOLS
HEALTHY CHICAGO CHICAGO DEPARTMENT OF PUBLIC HEALTH
Ushahidi
careerbuilder
NorthShore University HealthSystem
ENROLL AMERICA
Nurse-Family Partnership Helping First-Time Parents Succeed
SKILLS FOR CHICAGOLAND'S FUTURE
DIVVY
CONSERVATION INTERNATIONAL
CDOT CHICAGO DEPARTMENT OF TRANSPORTATION
MuskingumValley Educational Service Center
JOHNSON COUNTY KANSAS
TULSA PUBLIC SCHOOLS
NEW YORK STATE OF OPPORTUNITY Department of Environmental Conservation
SEDESOL SECRETARÍA DE DESARROLLO SOCIAL
AllianceChicago Innovating for better health
spire
DigitalGlobe
WORLD ECONOMIC FORUM
planet.
New Vision for the Ocean
Syracuse New York
City of Rotterdam
SANERGY
Rijkswaterstaat Ministerie van Infrastructuur en Milieu

AllianceChicago

Innovating for better health

Predict risk of Type II Diabetes → **Personalize** Screening Decisions → **Connect** to interventions and services → **Prevent** diabetes and improve health





Privacy &
Data Ownership

Trust &
Transparency

Ethical
Issues in DS

Accountability

Bias &
Fairness

Gender and racial bias found in Amazon's facial recognition technology (again)

Research shows that Amazon's tech has a harder time identifying gender in darker-skinned and female faces

By James Vincent | Jan 25, 2019, 9:45am EST

Amazon used AI to promote diversity. Too bad it's plagued with gender bias.

Share on Facebook Share on Twitter +

WHAT'S THIS?

Facebook accused of job ad gender discrimination

19 September 2018

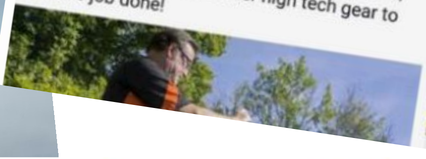
Share

Enhanced Roofing & Remodeling
Sponsored ·

LIKE PAGE

We are looking to add a roofing tech/estimator to our team. If you are looking for a rewarding career in the roofing industry than we are the company for you.

We offer competitive salaries and full benefits package with nice perks like a personal truck, laptop, cell phone and other high tech gear to get the job done!



Suggested Post

JK Moving Services
Sponsored ·

LIKE PAGE

OTR, Regional, and Local: our drivers enjoy no forced dispatch & new trailers. Owner Operators wanted, too!

Google is working to remove gender bias in its translations

Sources of Bias

Sample

Label

Machine Learning Pipeline

Application

Sample Bias

What is the relevant population for the project and how might some individuals be incorrectly excluded or included from the data available for modeling?

Are their underlying systemic biases involved in defining that population in general?

Data quality might not be uniform across groups.

Assuming a fixed feature space, more data (points) to train a classifier results in less test errors while less data leads to worse predictions.

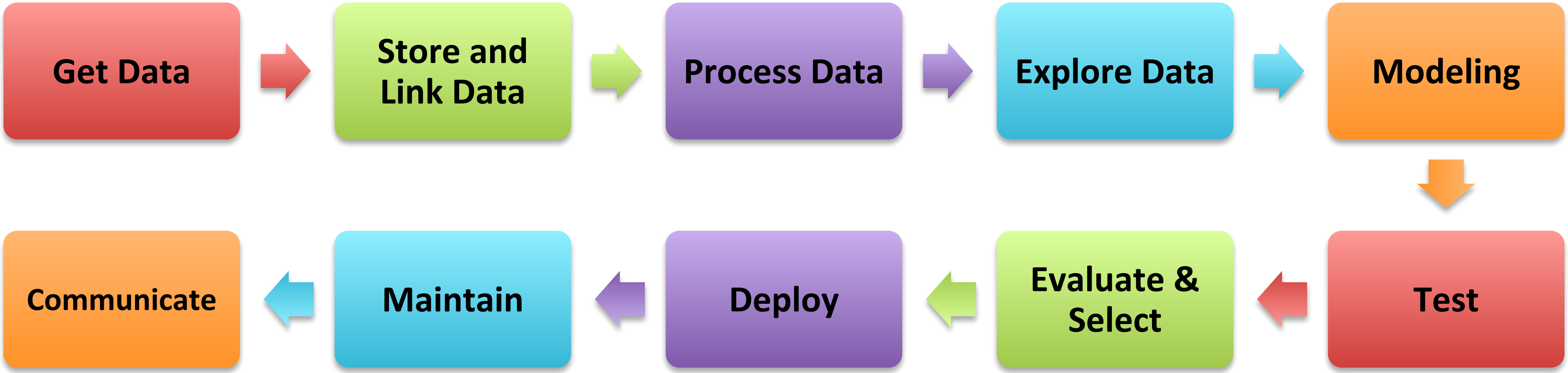
Label Bias

The way the target variable is defined and each data point is labeled might represent disparities between groups.

Differential measurement accuracy across groups (labeling quality).

A variable can be positively correlated with target variable within the majority group but negatively on other groups.

Bias can be introduced in every step of the ML pipeline



Application Bias

Heterogeneity in the effectiveness of an intervention across groups.

Discriminatory ‘overrides’ by the actors conducting the interventions.

“There’s software used across the country to predict future criminals. And it’s biased against blacks.”

ProPublica, 2016

My fairness definition or yours?

There is no universally-accepted definition of what it means for a decision-making model to be fair.

Punitive Example

A model being used to make bail determinations
(keeping people in jail)

Different people might consider it "fair" if:

It makes mistakes about denying bail to an equal number of white and black individuals.

Equal count of False Positives

$$P(\text{wrongly jailed, group } i) = C \quad \forall\, i$$

Different people might consider it “fair” if:

The chances that a given black or white person will be wrongly denied bail is equal, regardless of race.

Equal Group Size-Adjusted False Positives

$$P(\text{wrongly jailed} \mid \text{group } i) = C \quad \forall\, i$$

Different people might consider it “fair” if:

Among the jailed population, the probability of having been wrongly denied bail is independent of race.

Equal False Discovery Rate

$$P(\text{wrongly jailed} \mid \text{jailed, group } i) = C \quad \forall\, i$$

Different people might consider it "fair" if:

For people who should be released, the chances that a given black or white person will be denied bail is equal

Equal False Positive Rate

$$P(\text{wrongly jailed} \mid \text{innocent, group } i) = C \quad \forall\, i$$

Assistive Example

A model being used to subsidy diabetes screening and access to preventive care.

Different people might consider it "fair" if:

It makes mistakes about denying a subsidy to an equal number of women and men.

Equal count of False Negatives

$$P(\text{missed by benefit, group } i) = C \quad \forall\, i$$

Different people might consider it “fair” if:

The chances that a given woman or man will be wrongly missed by a subsidy is equal, regardless of sex.

Equal Group Size-Adjusted False Negatives

$$P(\text{missed by benefit} \mid \text{group } i) = C \quad \forall\, i$$

Different people might consider it "fair" if:

Among the non screened population, the probability of having been wrongly denied subsidy is independent of sex.

Equal False Omission Rate

$$P(\text{missed by program} \mid \text{no subsidy, group } i) = C \quad \forall\, i$$

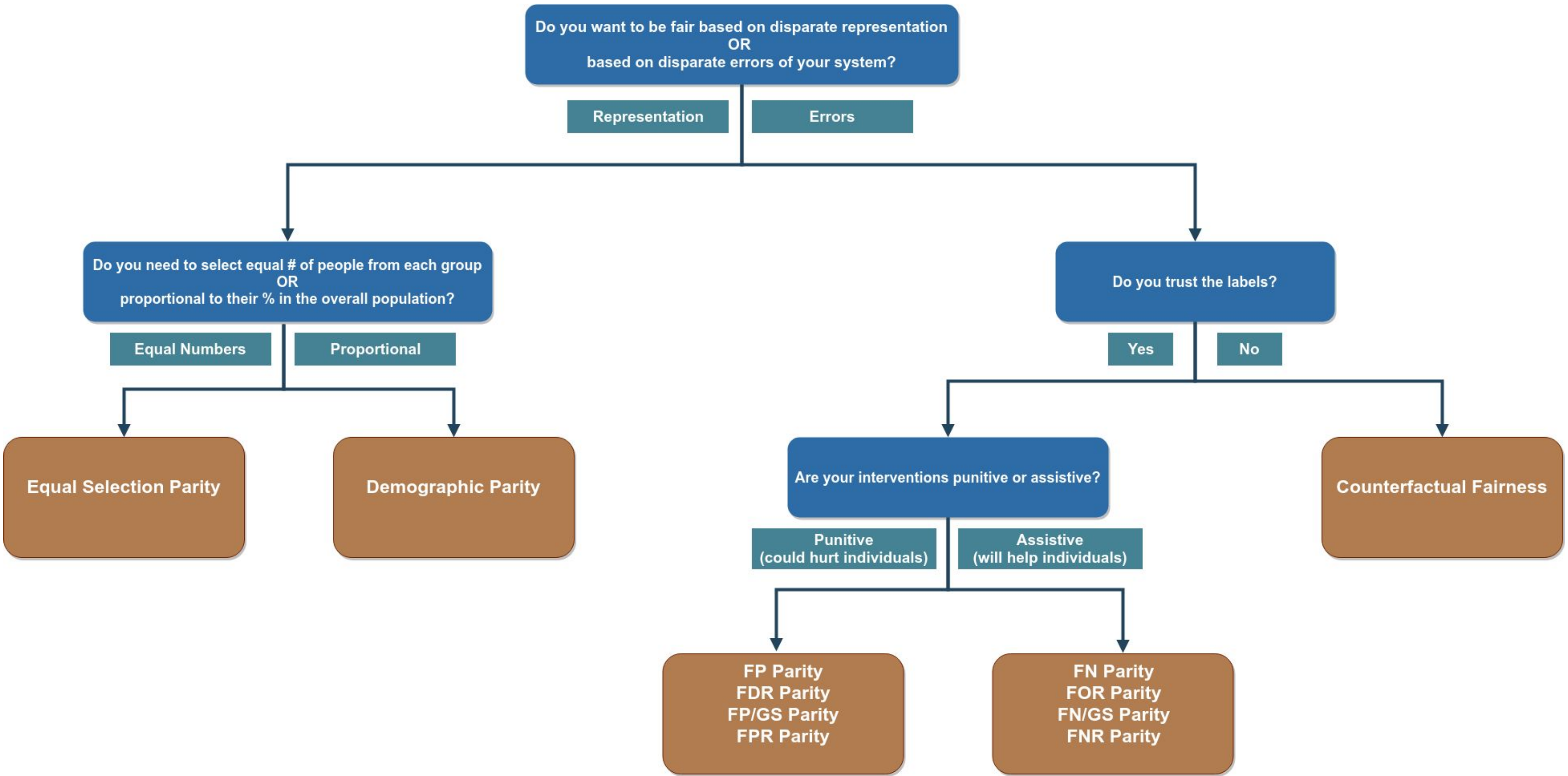
Different people might consider it “fair” if:

For people who need assistance, the chances that a given woman or man will not get a subsidy is equal

Equal False Negative Rate

$$P(\text{missed by subsidy} \mid \text{need assistance, group } i) = C \quad \forall\, i$$

FAIRNESS TREE



No “Supervised Fairness through Unawareness”

Shall I use race or gender in my models? Remove protected attributes?

Well, other features subsume the protected attributes.

Example: Easy to predict gender based on Facebook likes.

No “Supervised Fairness through Demographic Parity”

Example: Accept the same % of women and men for the job as the % of women and men in the population (city? country? candidates pool?)

Decision to be independent from the protected attribute?

Does not ensures “supervised fairness”, as it is possible to have different false positive/negative parities across groups.

Cripples the overall utility metric (e.g. A correlated with Y)

Fairness tradeoffs

If the base rate (prevalence) is different between groups and the classifier is non-trivial (Recall>0) and imperfect (FPR > 0). Then, either:

Precision Parity Fails (same as FDR parity)

FPR and Recall (same as FNR) will be disparate (no equalized odds)

[Kleinberg16, Chouldechova17]

Aequitas

Bias & Fairness Audit

University of Chicago

Pedro Saleiro

Aequitas - Bias and Fairness Audit Toolkit

How can you use Aequitas?

Web Audit Tool

Try our Audit Tool to generate a Bias Report

1. Upload Data (or use pre-loaded sample data)
2. Configure (bias metrics of interest and reference groups)
3. Generate the Bias Report

Try it out! ›

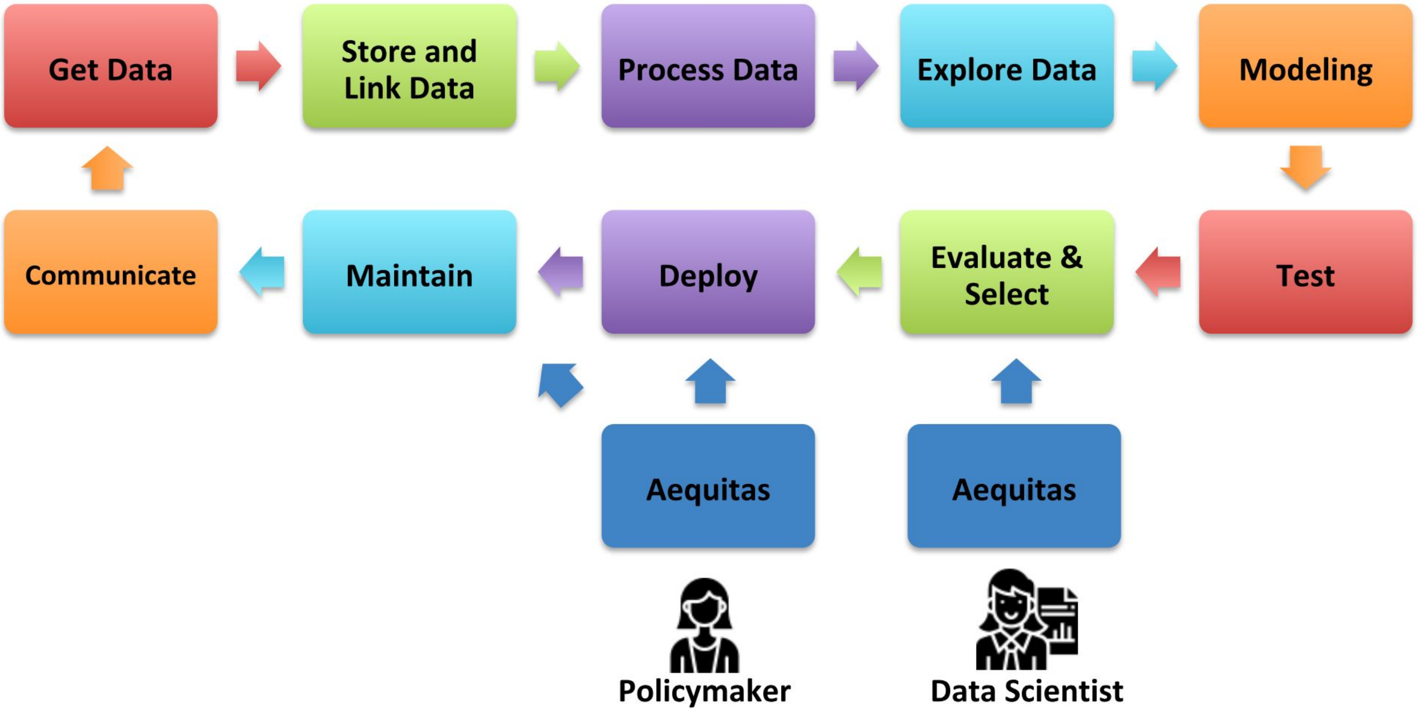
Python Library

Use our python code library to generate bias and fairness metrics on your data and predictions.

Python Code ›

Command Line Tool

Use our command line tool to generate a report using your own data and predictions.


Get Data
Store and Link Data
Process Data
Explore Data
Modeling
Communicate
Maintain
Deploy
Evaluate & Select
Test
Aequitas
Aequitas
Policymaker
Data Scientist

Type II Diabetes
Risk Score
Predictions + Labels

Race
Sex
Age

Disproportionally
Missing People
Across Groups

Upload Data → Select Protected Groups → Select Fairness Metrics → The Bias Report

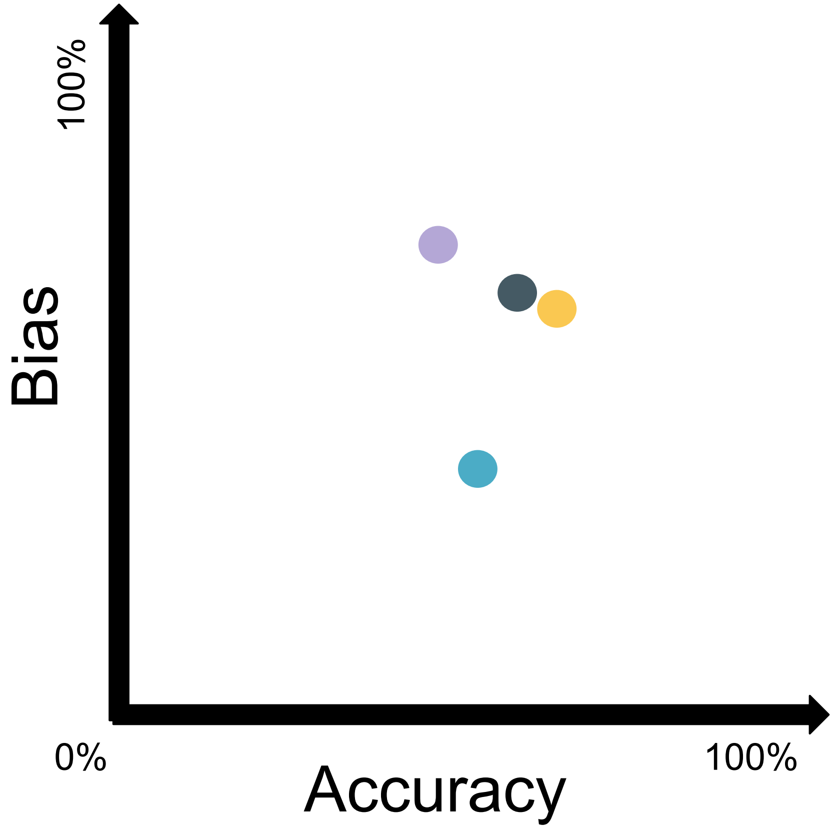
What do you need to define to accept/reject a model based on audit results?

Protected Groups (e.g. age > 70, sex=female)

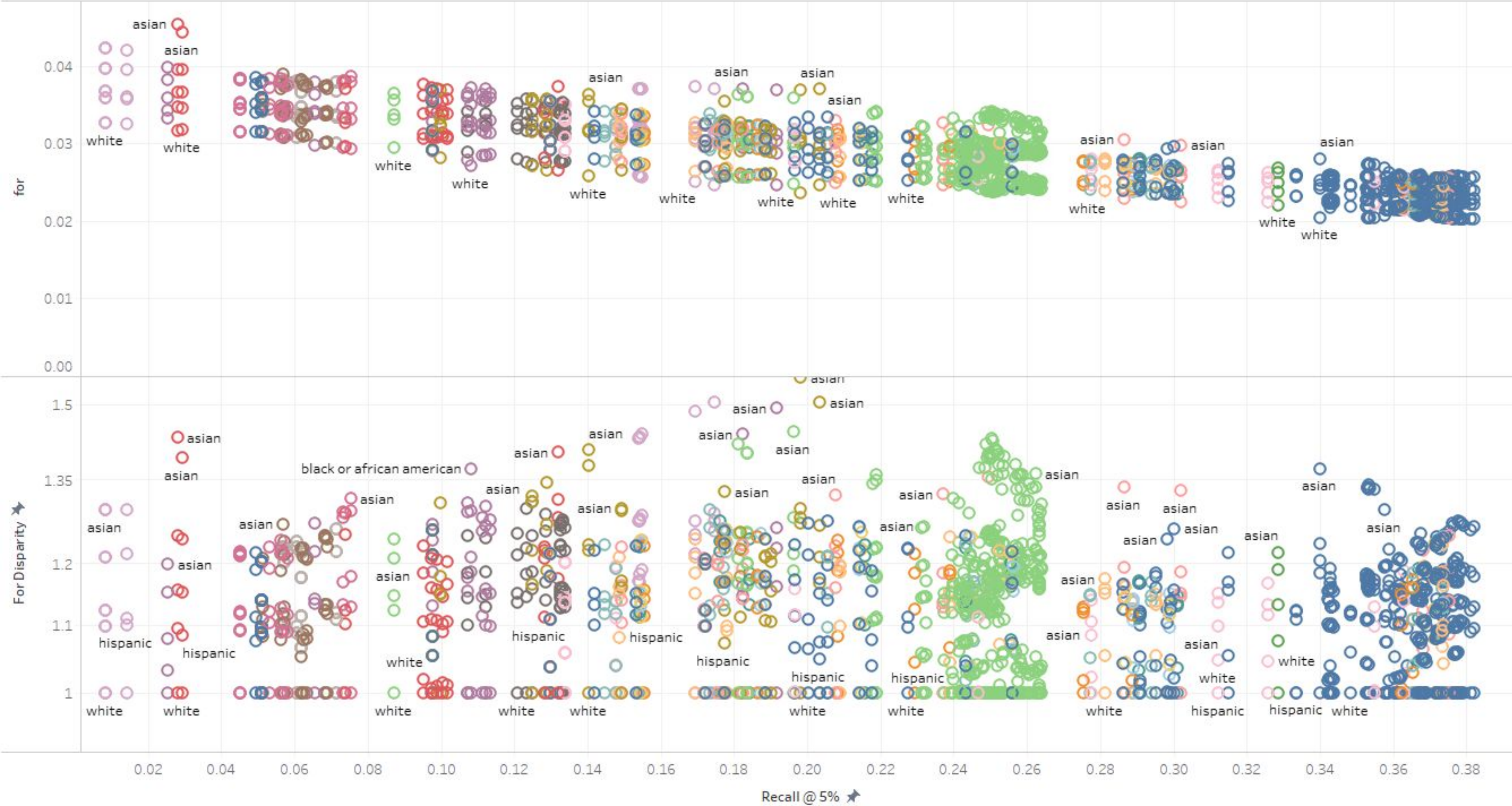
Metrics/Concerns (e.g. disparate # of FP)

Fairness Criteria (e.g. model fails audit if at least one metric for at least one protected group has a disparity > x)

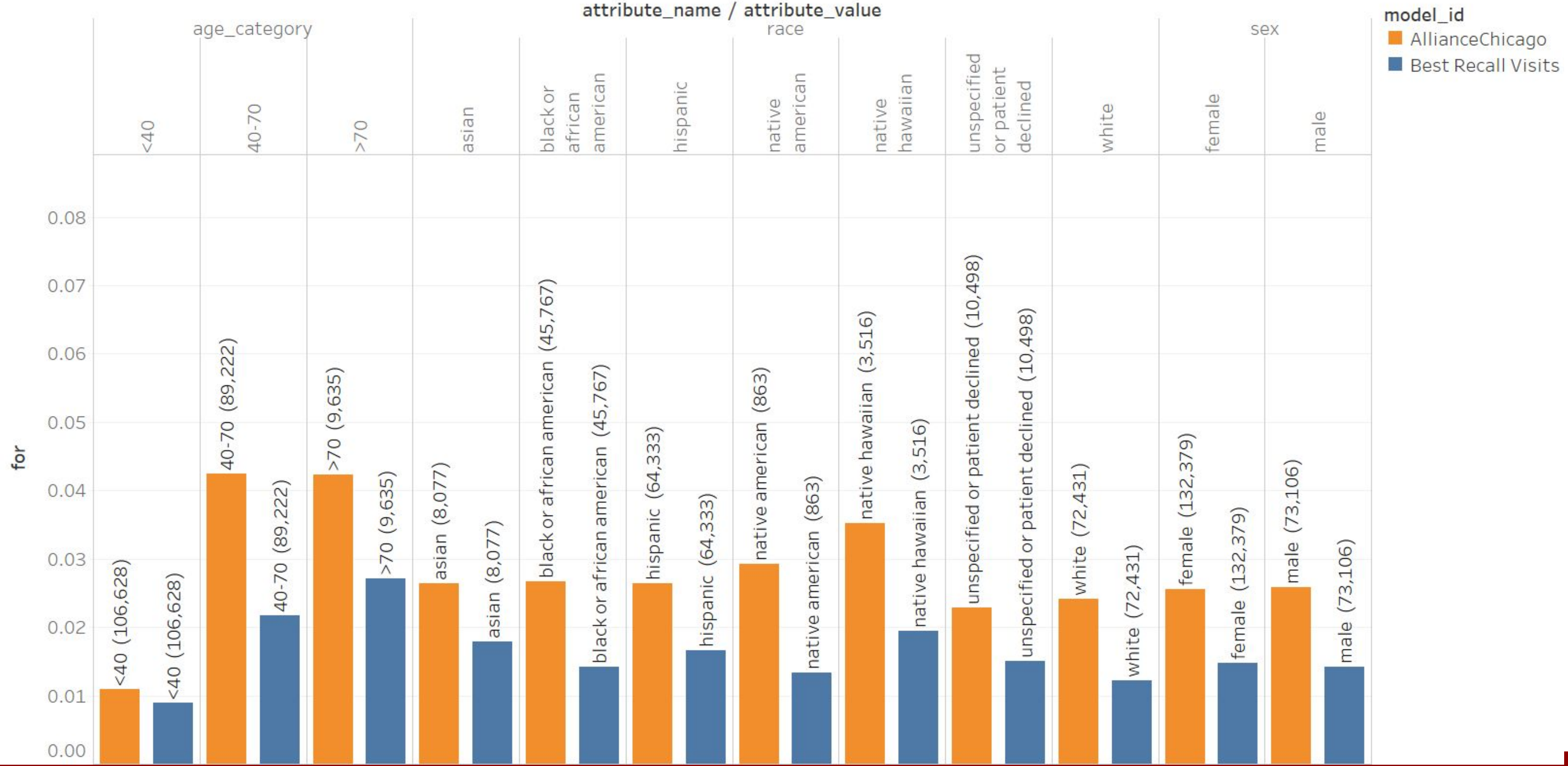

0%
100%
Accuracy


100%
Bias
0%
Accuracy
100%

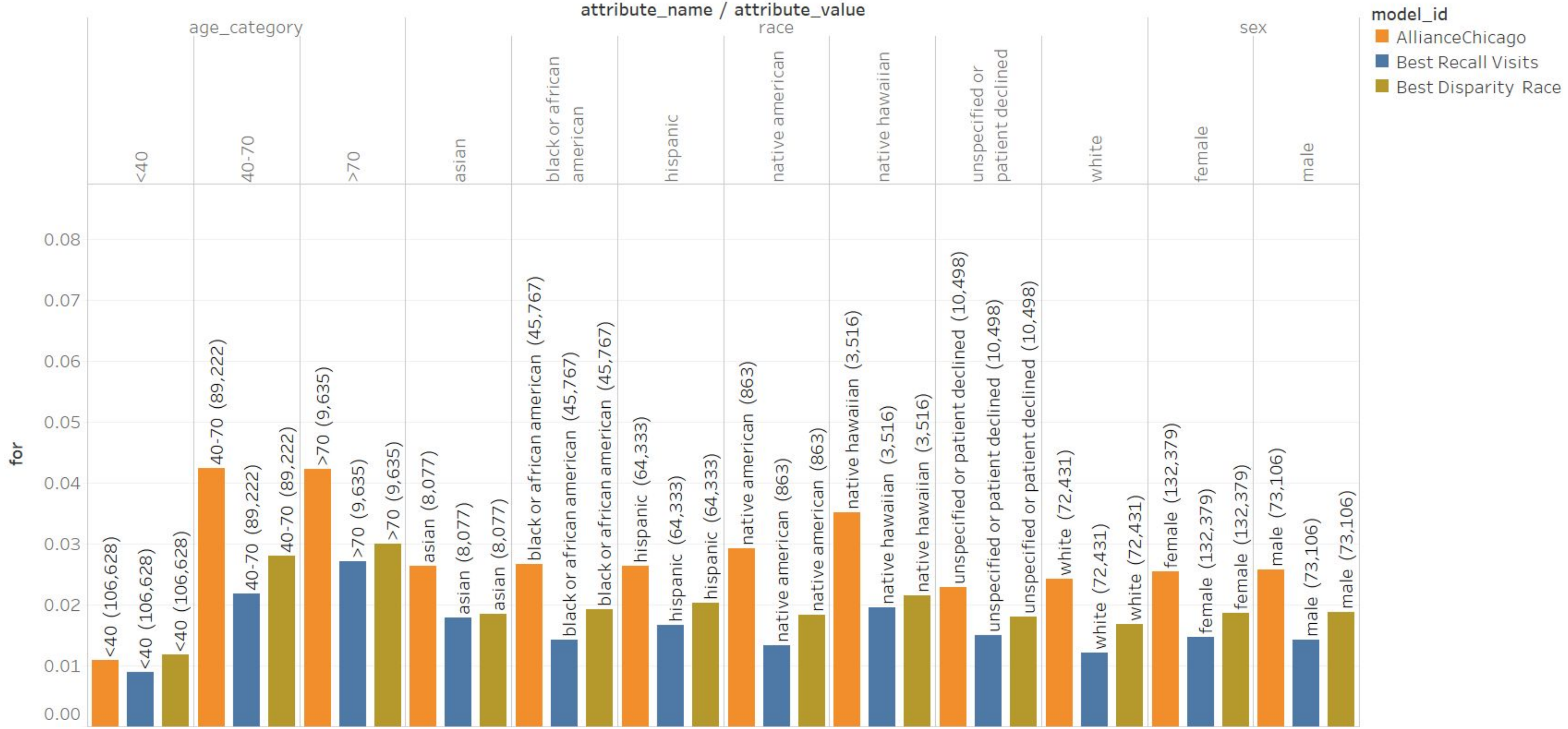
Model by feature group



Alliance vs Best Recall Visits



Best Disparity Race



Model Selection Recipe

1. Audit current practice/baseline

2. Compare the following models:
   a. Best performance (e.g. highest precision@k)
   b. Best for each bias metric of interest (avg or min-max)
   c. Best for each protected group on each bias metric of interest

3. Establish priorities
   a. Maximize overall performance
   b. Minimize error metrics of interest for protected groups
   c. Minimize disparities/differences of protected groups vs non-protected

Takeaway Message

The use of ML & DS in real-world policy issues represents an opportunity to increase effectiveness and efficiency of services and programs.

When scoping DS projects it is fundamental to define the fairness goals and the protected groups based on the planned actions/interventions.

Audit for bias and fairness is a first step to create awareness and make more informed decisions around developing and deploying ML models that can affect people lives.

Data is just an artifact, it reflects societal behaviors and its biases. If we cannot achieve our fairness goals using ML, we might need to tackle the upstream problem.

http://aequitas.dssg.io