Responsible Data Science

SLIDES BY:

JIANNAN WANG

https://www.cs.sfu.ca/~jnwang/

Data scientists have a lot of power

A lot of data

A lot of ML/Stats methods

A lot of data-driven decisions

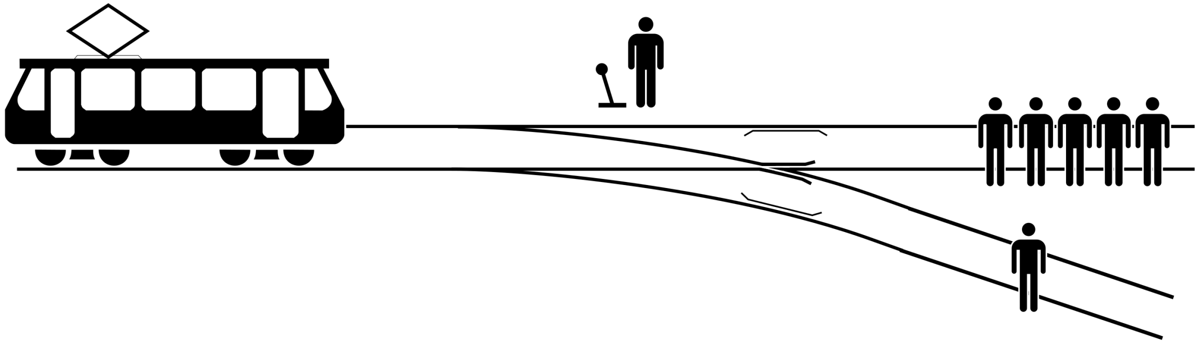
- Whether Tom can get admitted by a university
- Whether Tom can get an offer from a company
- Whether Tom can get a loan from a bank
- Whether Tom can express his option on a website
- Whether Tom can be treated properly in a hospital
- . . .

What is a right decision?

EASY

or

HARD



Women less likely to be shown ads for high-paid jobs on Google, study shows

Automated testing and analysis of company's advertising system reveals male job seekers are shown far more adverts for high-paying executive jobs

"One experiment showed that Google displayed adverts for a career coaching service for "$200k+" executive jobs 1,852 times to the male group and only 318 times to the female group. Another experiment, in July 2014, showed a similar trend but was not statistically significant."

CNBC

MARKETS BUSINESS INVESTING TECH POLITICS CNBC TV WATCHLIS

Amazon scraps a secret A.I. recruiting tool that showed bias against women

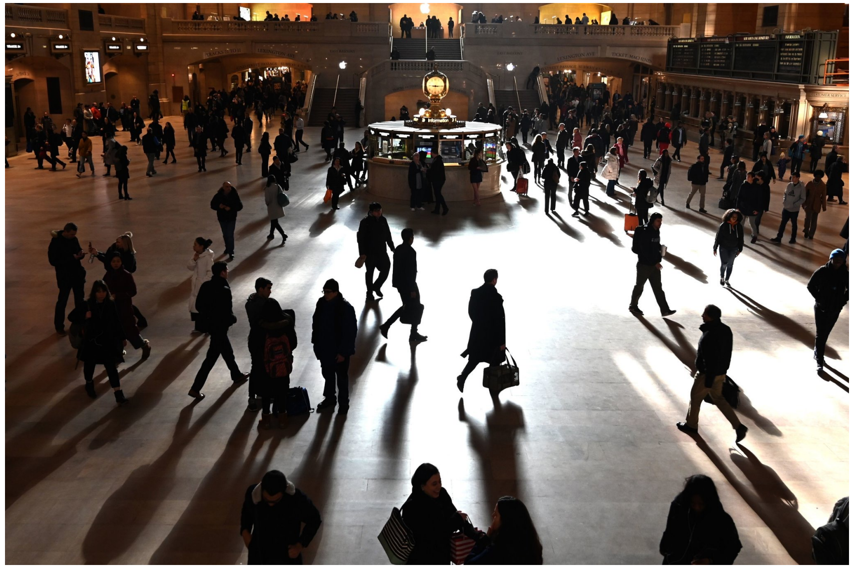
PUBLISHED WED, OCT 10 2018•6:15 AM EDT | UPDATED THU, OCT 11 2018•2:25 PM EDT

- Amazon.com's machine-learning specialists uncovered a big problem: their new recruiting engine did not like women.
- The team had been building computer programs since 2014 to review job applicants' resumes with the aim of mechanizing the search for top talent, five people familiar with the effort told Reuters.
- The company's experimental hiring tool used artificial intelligence to give job candidates scores ranging from one to five stars — much like shoppers rate products on Amazon, some of the people said.

The New York Times

Many Facial-Recognition Systems Are Biased, Says U.S. Study

Algorithms falsely identified African-American and Asian faces 10 to 100 times more than Caucasian faces, researchers for the National Institute of Standards and Technology found.



By Natasha Singer and Cade Metz

Dec. 19, 2019

Data Science Ethics

Informed Consent

Data Ownership

Privacy

Data Validity

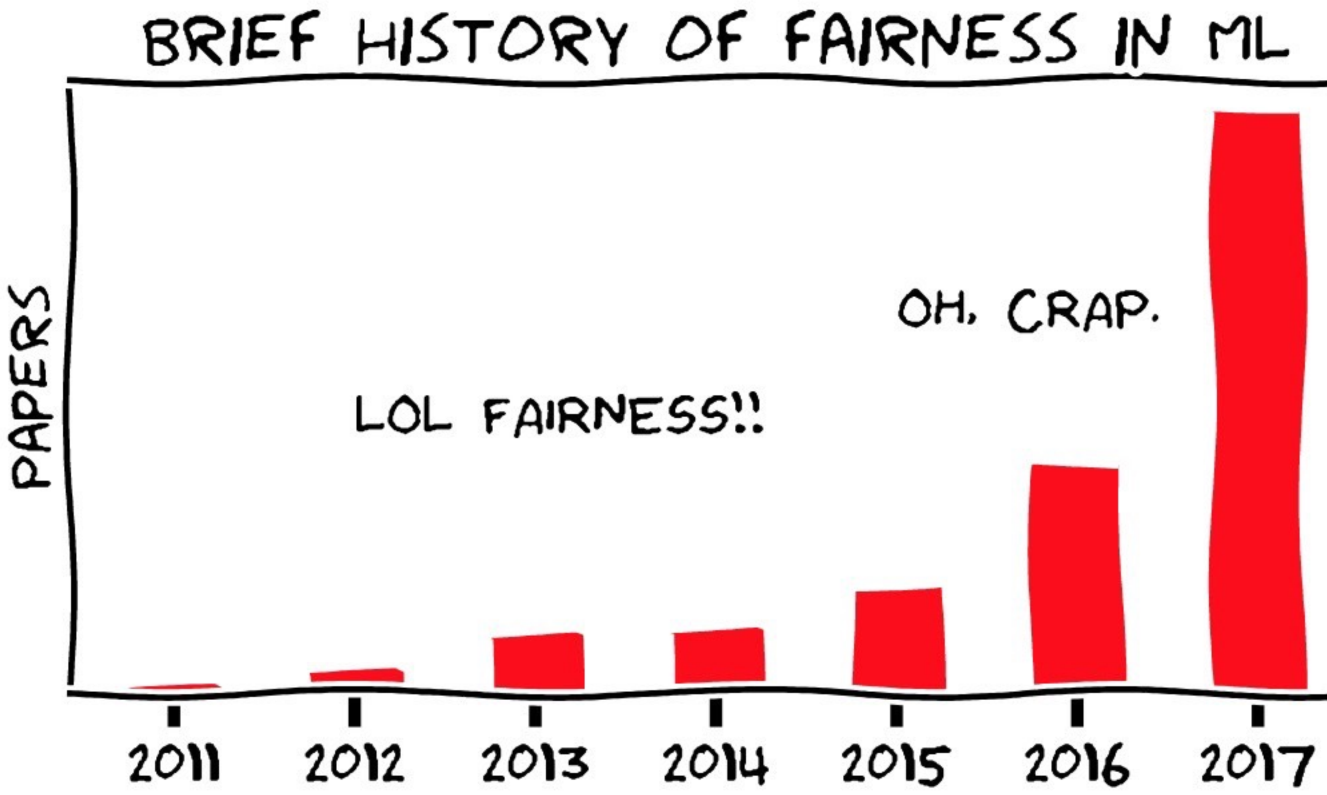
Algorithmic Fairness

https://www.coursera.org/learn/data-science-ethics/

DS-GA 3001.009: Special Topics in Data Science: Responsible Data Science

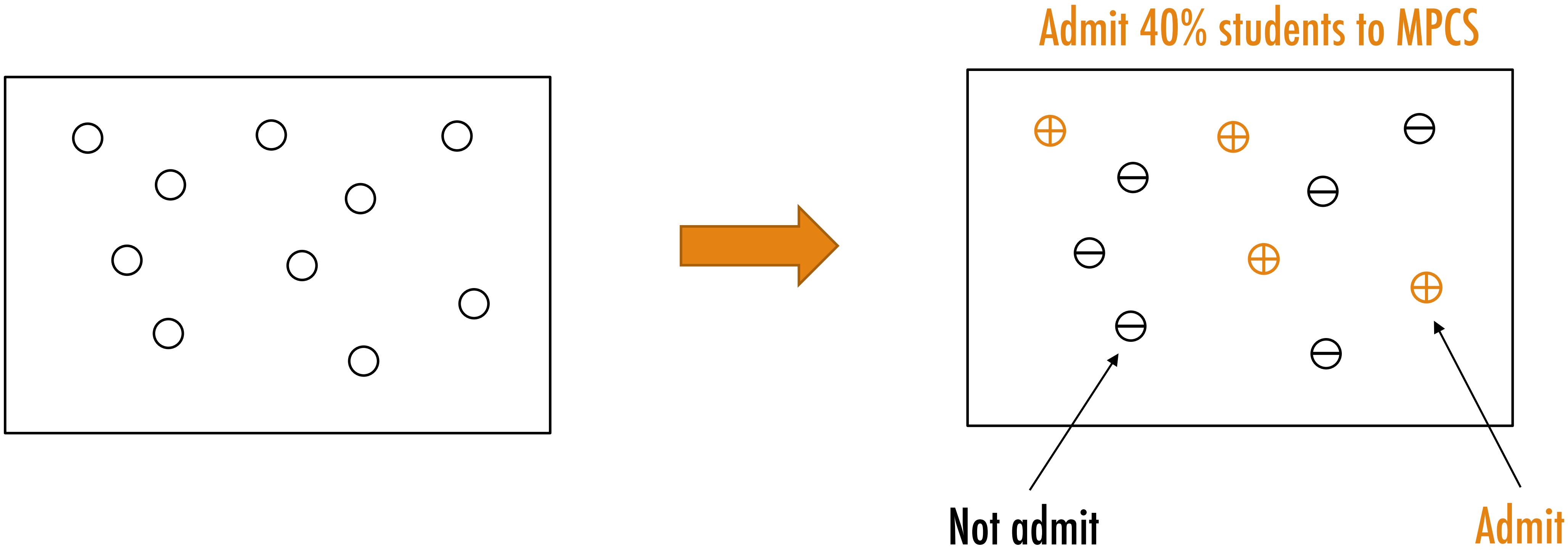
https://dataresponsibly.github.io/courses/spring19/

Fairness in Machine Learning



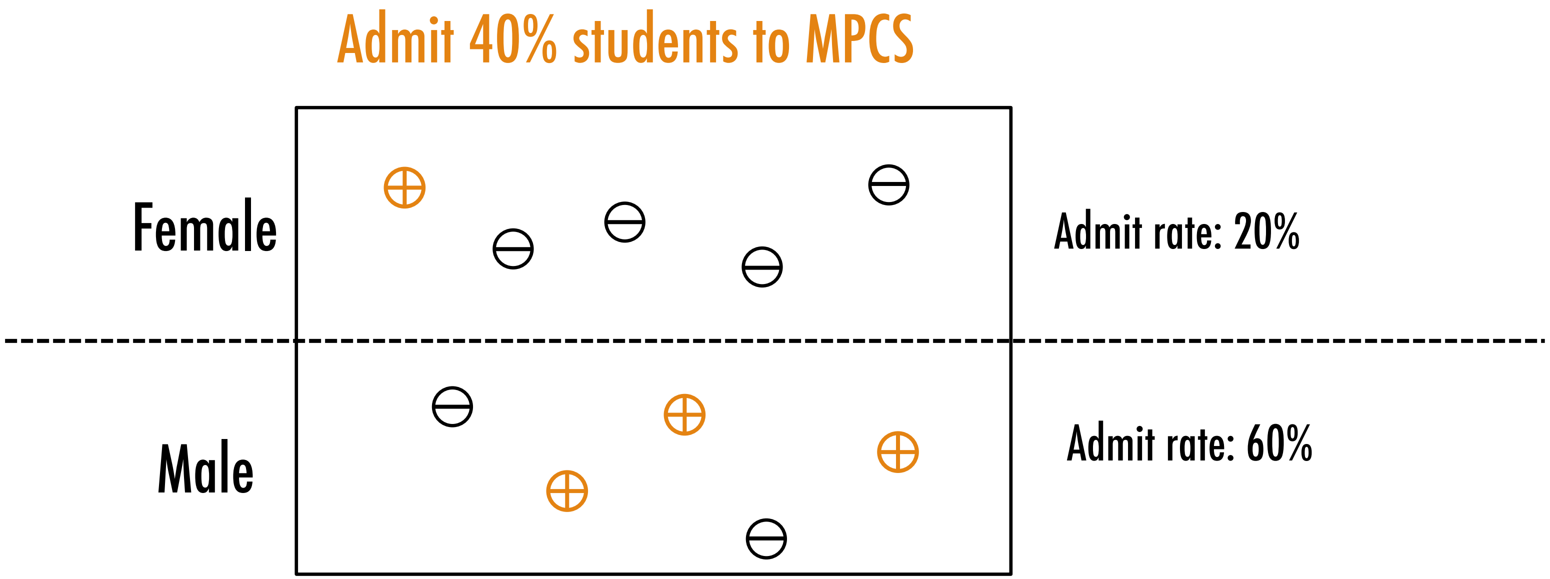
Fairness in Machine Learning

Is my model fair?



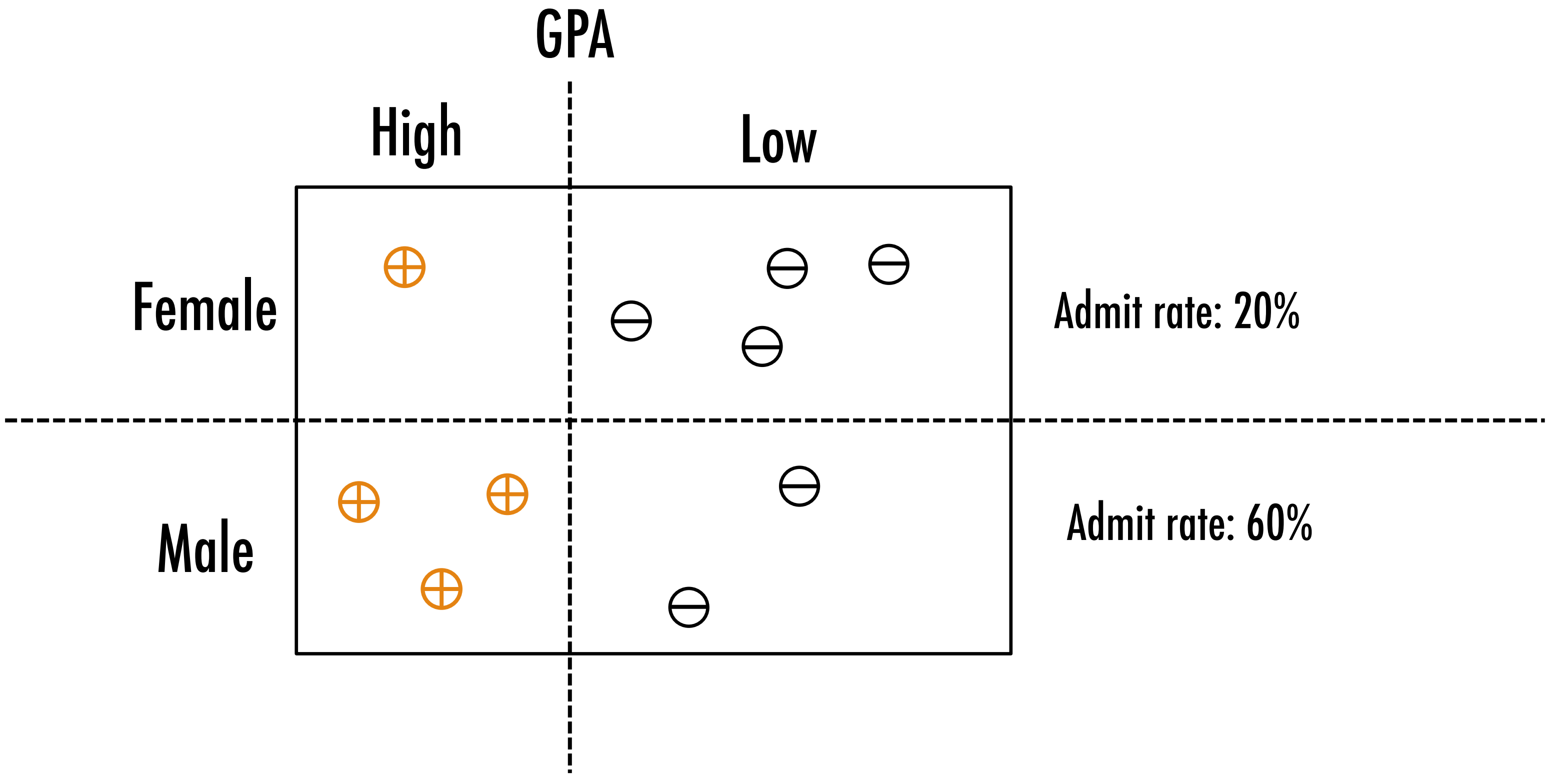
Fairness in Machine Learning

Female and male applicants are treated differently



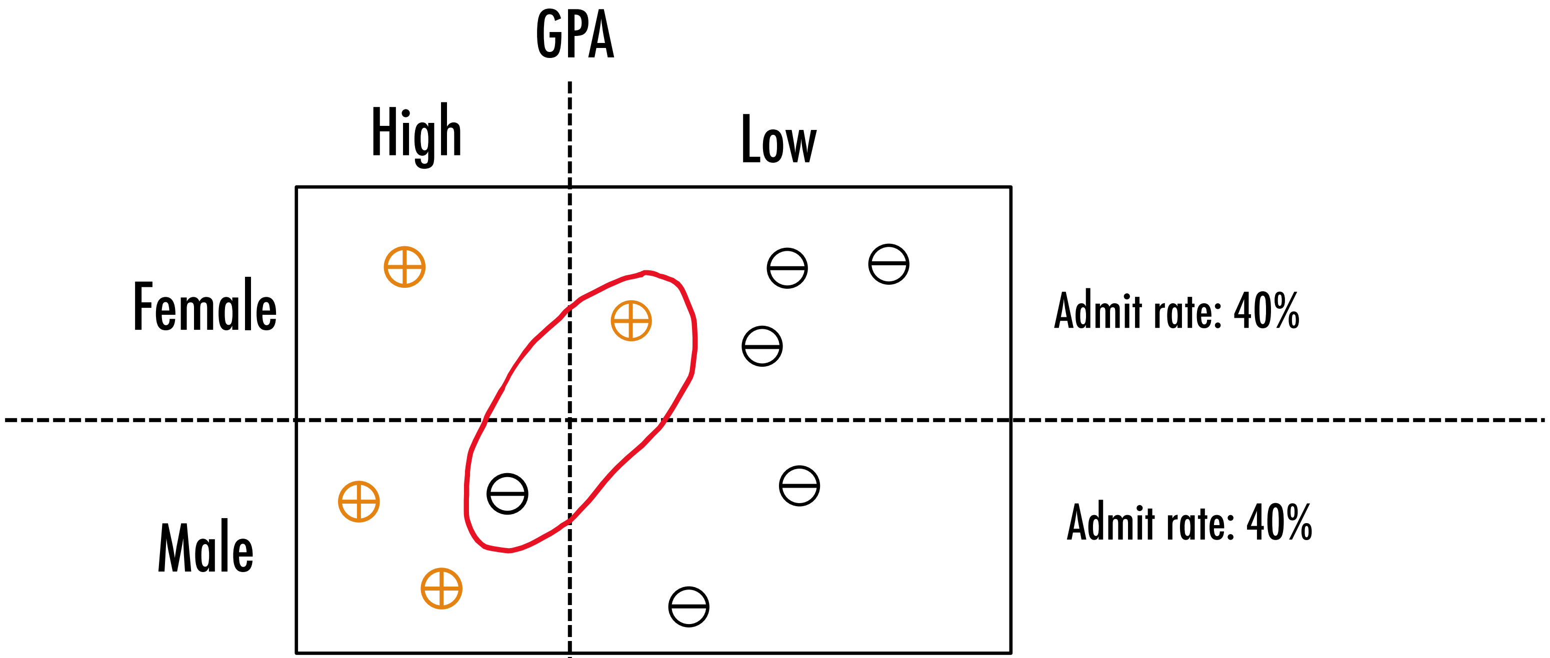
Fairness in Machine Learning

How to make my model fair?



Fairness in Machine Learning

How to make my model fair?



Is it fair?

Two notions of fairness

Equality

Giving everyone the same thing

Equity

Giving everyone access to the same opportunity

Toolkits

Microsoft https://github.com/fairlearn/fairlearn

IBM https://github.com/Trusted-AI/AIF360

Google https://github.com/tensorflow/fairness-indicators

AIF360 https://github.com/Trusted-AI/AIF360

Datasets

Toolbox

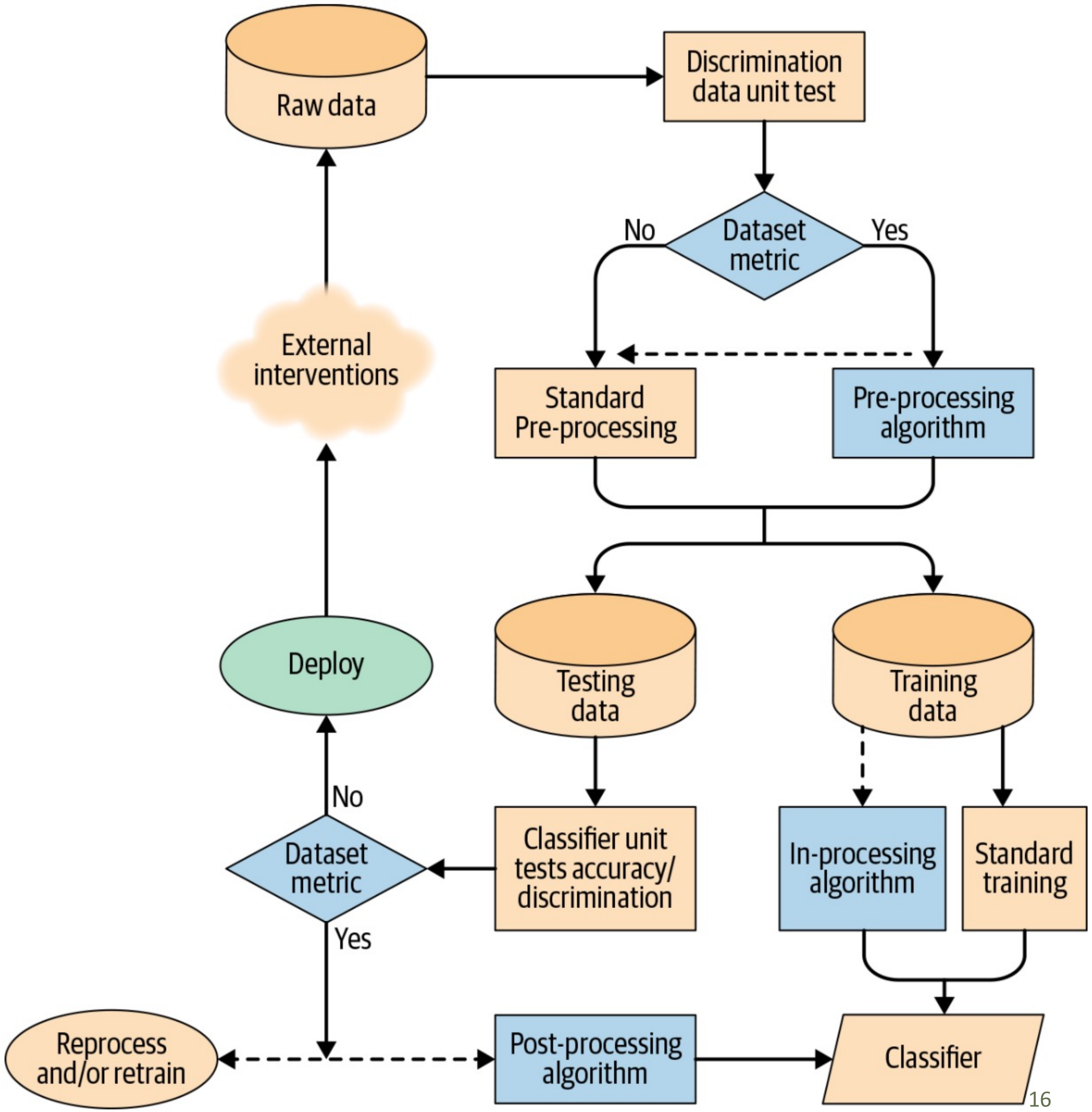
- Fairness metrics (30+)
- Fairness metric explanations
- Bias mitigation algorithms (9+)

Guidance

Industry-specific tutorials

Bias In the Machine Learning Pipeline

AI Fairness by Trisha Mahoney, Kush R. Varshney, and Michael Hind



AIF360 Algorithms

Pre-processing

- Reweighing
- Disparate Impact Remover
- Learning Fair Representations
- Optimized Preprocessing

In-processing

- Calibrated Equality of Odds
- Equality of Odds
- Reject Option Classification

Post-processing

- ART Classifier
- Prejudice Remover
- Post-processing

Reweighting

Modify the weights of different training examples such that

P(admit | Sex = 'Female')

=

P(admit | Sex = 'Male')

Sex	Ethnicity	Highest degree	Job type	Class
M	Native	H. school	Board	+
M	Native	Univ.	Board	+
M	Native	H. school	Board	+
M	Non-nat.	H. school	Healthcare	+
M	Non-nat.	Univ.	Healthcare	–
F	Non-nat.	Univ.	Education	–
F	Native	H. school	Education	–
F	Native	None	Healthcare	+
F	Non-nat.	Univ.	Education	–
F	Native	H. school	Board	+

Reweighting

Algorithm 3: *Reweighing*

Input: $(D, S, Class)$

Output: Classifier learned on reweighed D

1: **for** $s \in$ {F, M} **do**

2: **for** $c \in \{-, +\}$ **do**

3: Let $W(s, c) := \dfrac{|\{X \in D \mid X(S) = s\}| \times |\{X \in D \mid X(Class) = c\}|}{|D| \times |\{X \in D \mid X(Class) = c \text{ and } X(S) = s\}|}$

4: **end for**

5: **end for**

6: $D_W := \{\}$

7: **for** X in D **do**

8: Add $(X, W(X(S), X(Class)))$ to D_W

9: **end for**

10: Train a classifier C on training set D_W, taking onto account the weights

11: **return** Classifier C

F. Kamiran and T. Calders, “Data Preprocessing Techniques for Classification without Discrimination,” Knowledge and Information Systems, 2012 (https://link.springer.com/content/pdf/10.1007%2Fs10115-011-0463-8.pdf)

Reweighting - Example

Sex	Ethnicity	Highest degree	Job type	Cl.	Weight
M	Native	H. school	Board	+	0.75
M	Native	Univ.	Board	+	0.75
M	Native	H. school	Board	+	0.75
M	Non-nat.	H. school	Healthcare	+	0.75
M	Non-nat.	Univ.	Healthcare	–	2
F	Non-nat.	Univ.	Education	–	0.67
F	Native	H. school	Education	–	0.67
F	Native	None	Healthcare	+	1.5
F	Non-nat.	Univ.	Education	–	0.67
F	Native	H. school	Board	+	1.5

$$\frac{5 \times 6}{10 \times 4} = 0.75$$

$$\frac{5 \times 4}{10 \times 1} = 2$$

$$\frac{5 \times 4}{10 \times 3} = 0.67$$

$$\frac{5 \times 6}{10 \times 2} = 1.5$$

Conclusion

Big Picture

- Why responsible data science?
- Data science ethics

Fairness

- Equality vs Equity
- AIF360

Reweighting

or