Lab 4
Sensitive Attributes and Data Bias

SEN163B – Q3, 22/23 - Week 4
Responsible Data Analytics

By Nadia Metoui

TA Anagha Magadi Rajeev

TU Delft

Learning Objectives

At the end of this lab, you will be able to

- Use data analytics tools to identify Sensitive attributes and proxies
- Use data analytics tools to identify Representation Bias
- Use data analytics tools to identify Historical Bias
- Understanding, Analyse and Discuss
  - Historical Bias,
  - Representation Bias
  - Measurement Bias

Technical(ish) Definitions

- Protected characteristics: *UK Equality Act 2010*
  - *Age,*
  - *Disability,*
  - *Gender reassignment,*
  - *Marriage and civil partnership,*
  - *Pregnancy and maternity,*
  - *Race,*
  - *religion or belief,*
  - *sex(gender), and sexual orientation.*

Technical(ish) Definitions

- Proxies of Protected characteristics

A proxy is a seemingly “non-sensitive” or “non-protected” attribute that can be used to infer a “sensitive” or “protected” attribute.

Inference can be of “logical” or “common sense” nature

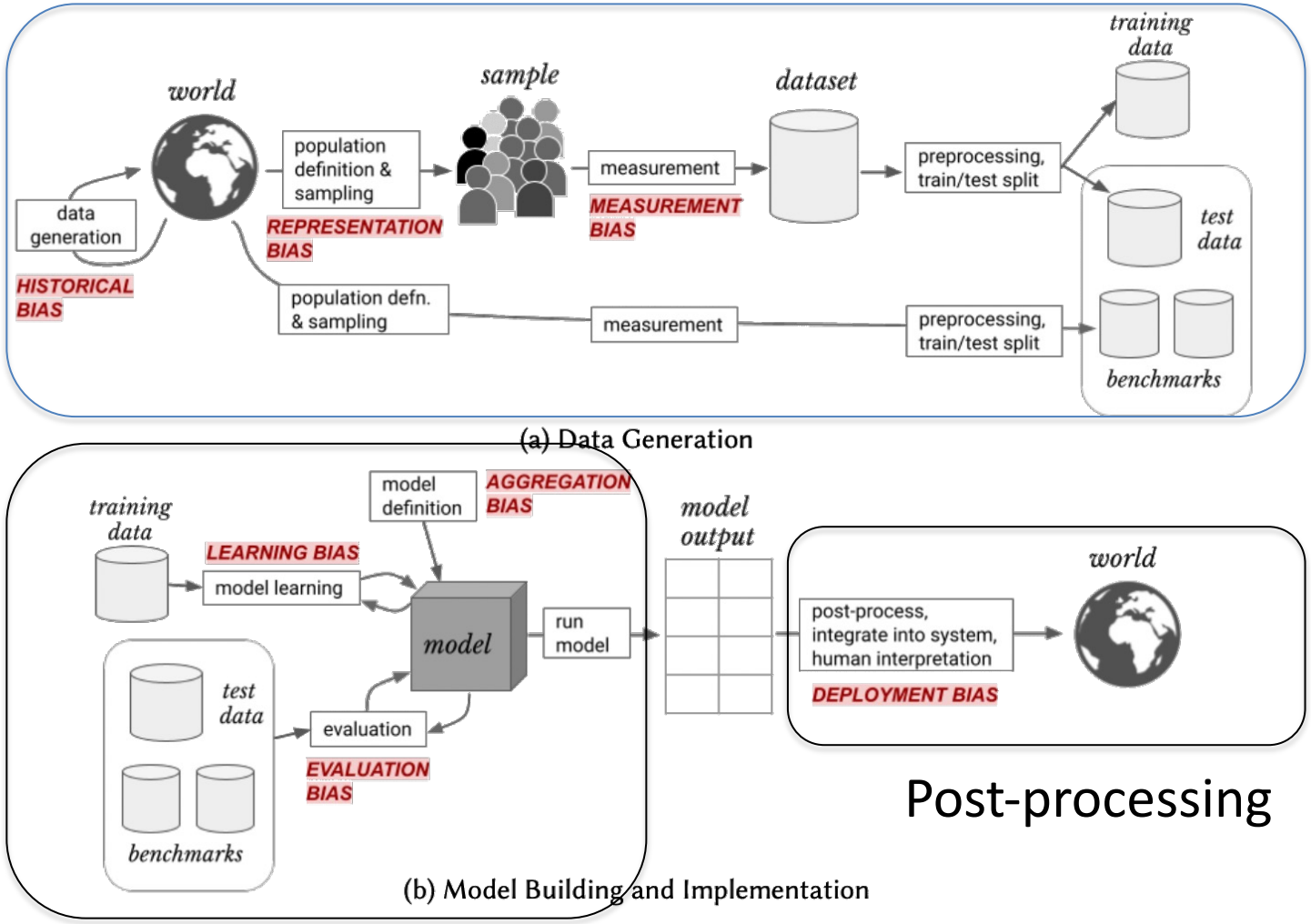
- E.g., data about pregnancies can reveal that a person’s sex

Inference can be of a “statistical” nature: strong correlation in data

- E.g., in the US zip code often reveal the income and ethnicity

Bias in the ML Pipeline (Possible Mapping)

Pre-processing

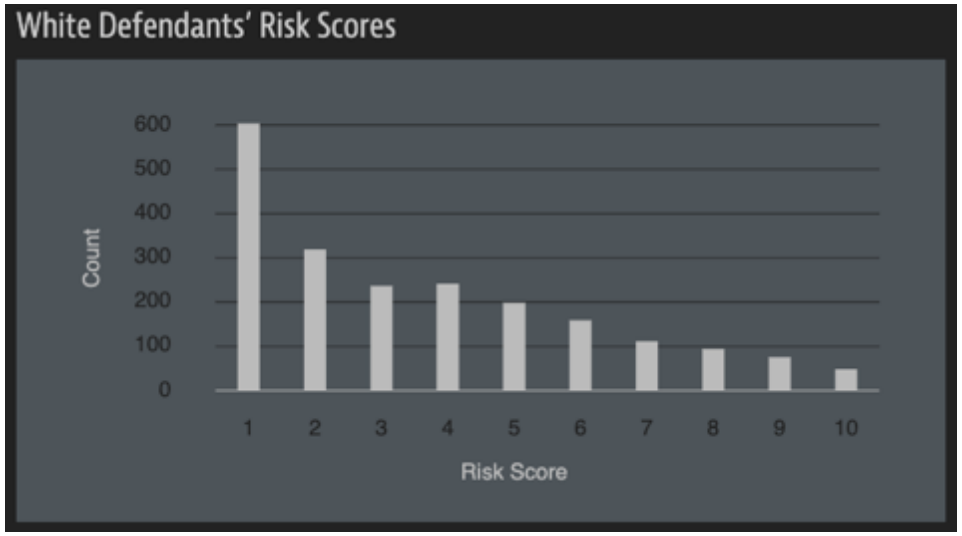


In-processing

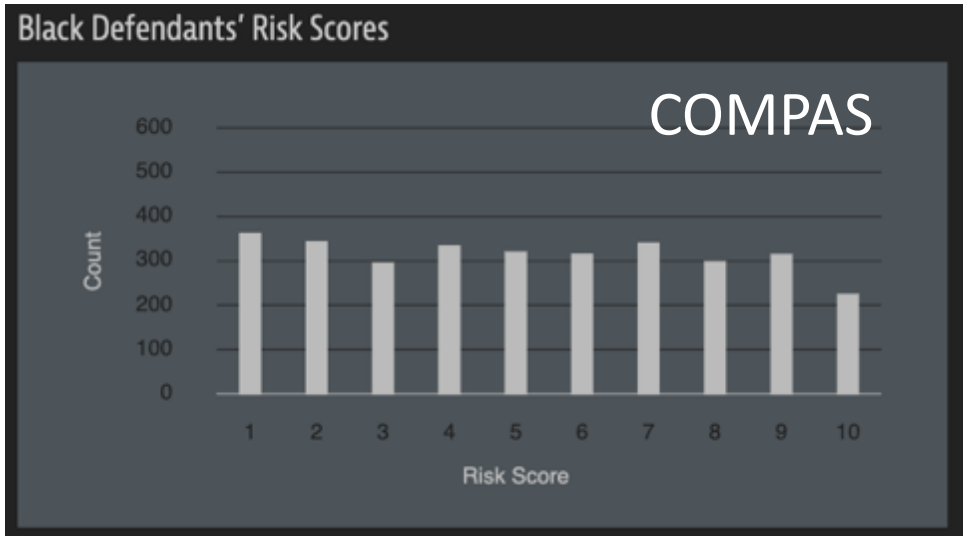
Post-processing

Historical Bias

Historical Biases: occurs when and ML model reproduces or reinforces a harmful stereotypes already existing in the world.



Scores for white defendants were skewed toward lower-risk categories.

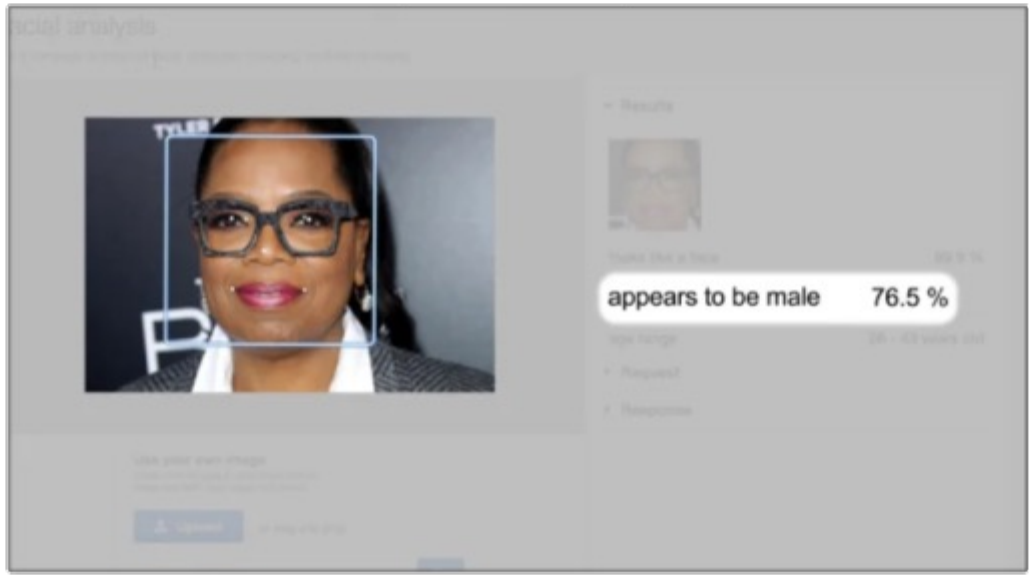


Scores for black defendants were not

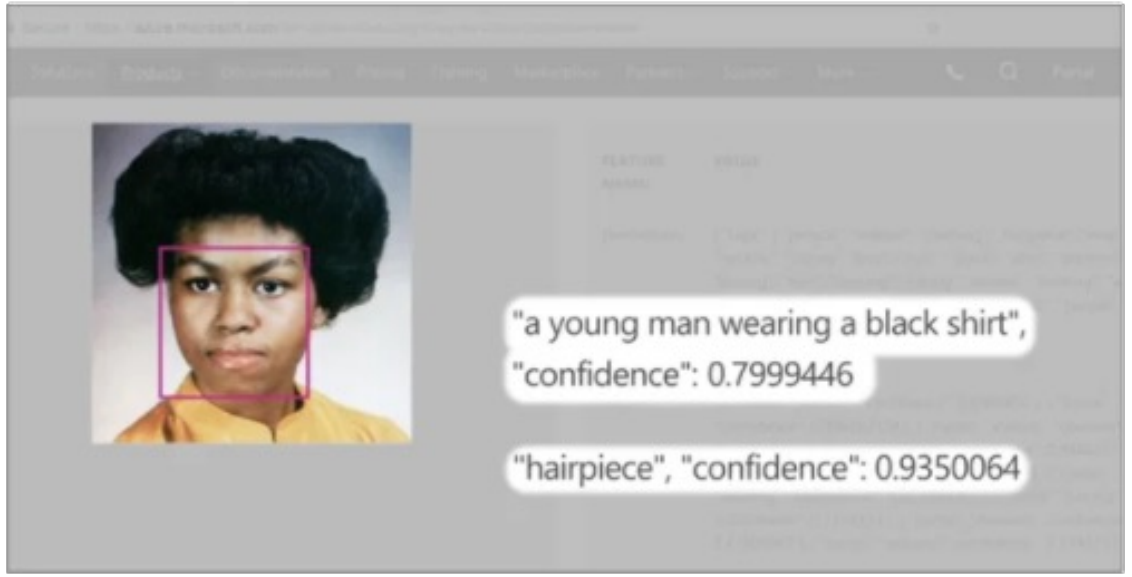
Representation Bias

Representation Biases: occurs when the data is not representative of the population the model is be developed for, or specific categories are largely under-represented.

Oprah Winfrey



Michelle Obama



TU Delft

amazon

Microsoft

Measurement Bias

Measurement Biases: occur when choosing, collecting, or computing the inadequate features and labels to use in a prediction problem.

Features and labels are the concrete measurements chosen to approximate some construct (an idea or concept) that is not directly encoded or observable.

Example from COMPAS: number of “arrest” is used measure “crime” or “arrest”

Lab

During the next lab, we will

- Part I. Understanding the Use Case
- Part II Protected attributes, proxies
- Part III: Representation Disparities, Skew and Bias
- Part IV. Discuss Measurement Bias.

Project (Part II)

After the lecture

- Reflection on the privacy impacts of your data analytics project (300 words)
- Group Positionality and Reflexivity Statement (300 words)

After the lab

- Reflect on Historical Bias (max 500 words)
- Analyse Representation Bias (Code + max 500 words)
- Reflect on Measurement Bias (max 500 words)

TU Delft