Data Ethics: Understanding Big Data, Algorithmic Bias, and Research Ethics

Cara Marta Messina and
CRIM 3600 Research Methods
Megan Denver
Fall 2019

Northeastern University
NULab for Texts, Maps, and Networks

Feel free to ask questions at any point during the presentation!

Discussion

When you are doing research–whether you are researching for a class, your career, or personal interest–what are some ethical guidelines you follow?

Feel free to ask questions at any point during the presentation!

Workshop Agenda

- Objectives
- Introduce ‘Big Data’ Concepts
- Activity: Animal or Plant?
- Algorithmic bias and the criminal justice system
- Research ethics

Slides, handouts, and data available at **http://bit.ly/33xzyUr**

Feel free to ask questions at any point during the presentation!

Workshop Goals

- Understand the ways in which technologies reflect cultural, social, and political biases.
- Explore the basic process for machine learning algorithms
- Understand the ways data is being used in society as well as how algorithms impact and shape our daily lives and the criminal justice system
- Explore the ways in which these questions and methods are influencing how humanists and social scientists do research and practice their craft

Feel free to ask questions at any point during the presentation!

What is ‘Big Data’?

Big data has been called the ‘new oil’ by some.

Shoshana Zuboff argues that we now live in an era of ‘surveillance capitalism.’

The four components of big data are: **volume**, **variety**, **velocity** and **veracity**

Feel free to ask questions at any point during the presentation!

Why should we care?

- Big data is characterized by its **scale**
- Big data **sources** include: digitized records, social media/internet activity, or sensors from the physical environment.
- Big data is often **privately owned**
  - Example: an insurance company purchasing social media activity from facebook in order to make specific insurance sales decisions.

Feel free to ask questions at any point during the presentation!

Ethical Implications

- Cambridge Analytica Controversy
- Big data also raises questions of autonomy, anonymity, privacy, discrimination, and bias.
- Questions to consider:
  - How are we being represented online?
  - How is our data being used?
  - Who is using it and for what purposes?
  - How might it be used in the future?

Feel free to ask questions at any point during the presentation!

DIY Cybersecurity and Tightening your Privacy

Want to make your life more private? Follow this “DIY Guide to Feminist Cybersecurity”

https://hackblossom.org/cybersecurity/

Northeastern University
NULab for Texts, Maps, and Networks

Feel free to ask questions at any point during the presentation!

Algorithms

Big data relies on the collection of high amounts of information and **algorithms** to parse through, categorize, and "read" that information.

Algorithms are a set of procedures to be followed by certain technologies (computers, cell phones, etc). Algorithms typically rely on data and a set of instructions to "read" that data in some way.

Feel free to ask questions at any point during the presentation!

Activity: Is It A Plant?

Feel free to ask questions at any point during the presentation!

Create Your Classifier

Break into groups of 2-3 people!

- Discuss the characteristics of a plant (TRUE or FALSE)
- On the handout, write down 5-10 criteria for the characteristics of a plant (preferably criteria you should be able to see in a photo). Do not fill out the image boxes yet

Feel free to ask questions at any point during the presentation!

Use Each Other's Classifiers

Pass your classifier to another group. Use these classifiers to determine if the subjects in the upcoming photos are **plants**!

For each image, go through each criteria (5-10) and determine if the criteria is TRUE or FALSE.

Feel free to ask questions at any point during the presentation!

Image 1

Feel free to ask questions at any point during the presentation!

Image 2

Feel free to ask questions at any point during the presentation!

Image 3

Feel free to ask questions at any point during the presentation!

Image 4

Feel free to ask questions at any point during the presentation!

Image 5

Feel free to ask questions at any point during the presentation!

Image 6

Feel free to ask questions at any point during the presentation!

BONUS IMAGE

Feel free to ask questions at any point during the presentation!

Discussion

- What criteria worked well? What did not?
- Were there issues with the complexity of your characteristic list?
- How effective is the TRUE/FALSE system of analysis? What other forms of analysis might you suggest?

Feel free to ask questions at any point during the presentation!

So what do algorithms have to do with the criminal justice system?

Northeastern University
NULab for Texts, Maps, and Networks

Feel free to ask questions at any point during the presentation!

Risk Assessment: Algorithmic Bias

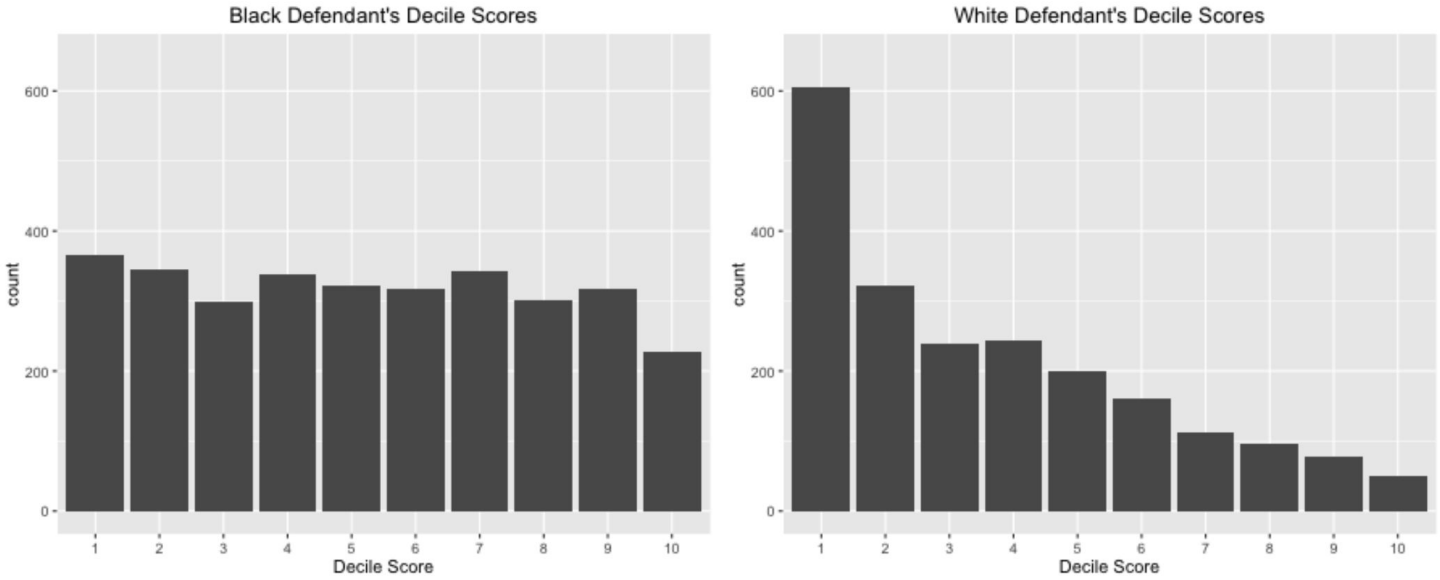
Risk assessment: used to determine the likelihood that someone will reoffend, not appear for trial, etc..

What happens when machine learning algorithms are used to help determine risk assessment?

Feel free to ask questions at any point during the presentation!

COMPAS Algorithm & ProPublica's Analysis

The COMPAS recidivism algorithm does not "see" race. Yet…



https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm

Feel free to ask questions at any point during the presentation!

Algorithmic Bias

Northeastern University
NULab for Texts, Maps, and Networks

Feel free to ask questions at any point during the presentation!

So what can we do?

Northeastern University
NULab for Texts, Maps, and Networks

Feel free to ask questions at any point during the presentation!

Questions Researchers Must Ask

- What **information** is being collected and from where? To whom does this data **belong**?
- How is it being **collected**? Do **participants** know that it is collected, how it will be collected, and how will it be used?
- **How** will the data be analyzed? What **biases** and **ideologies** may be implicit in this analysis?
- Who will this research impact? Who will it **benefit**? Who will it potentially **harm**?

Feel free to ask questions at any point during the presentation!

Thank you!

If you have any questions, contact us at:

Cara Marta Messina
Digital Teaching Integration
Assistant Director
messina.c@husky.neu.edu

DTI Consultant 2
Digital Teaching Integration
Title
Email

Slides, handouts, and data available at **http://bit.ly/33xzyUr**

Feel free to ask questions at any point during the presentation!