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

Garrett Morrow and Jeffrey Sternberg
POLS 2399 Research Methods
Aeshna Badruzzaman
Spring 2020

Northeastern University
NULab for Texts, Maps, and Networks

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

Mini Get-to-know-you Activity

What are some recent advertisements you have seen pop up on your Snapchat, Instagram, Facebook, Youtube, and other online spaces?

- Which advertisements interested you?
- What are some surprising advertisements?
- Why do you think you received those advertisements?

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

Workshop Agenda

- Objectives and Goals
- Introduce ‘Big Data’ Concepts
- Algorithmic bias and policy implications
- Algorithm activity
- Research ethics

Slides, handouts, and data available at

http://bit.ly/diti-spring2020-badruzzaman2

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

Workshop Objectives

- Understand the ways in which technologies reflect cultural, social, and political biases.
- Understand the ways data is being used in society as well as how algorithms impact and shape our daily lives.
- Explore the ways in which these questions and methods are influencing how 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, including Andrew Yang.

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!

Big Data: What is it and why should we care?

- Big data **sources** include: digitized records, social media/internet activity, and 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 insurance sales decisions.
- Big data can often reproduce results that may harm certain communities.

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

AWARENESS | SCIENCE & TECH | AUG 3, 2019 AT 11:08 AM.

Google's File on You is 10 Times Bigger Than Facebook's — Here's How to View It

Google, Amazon, Apple, and Microsoft are all central players in "surveillance capitalism" and prey on our data.

THE ANTI-MEDIA

Example: If you have **location services** turned on for Google (like if you use Google maps), Google can track your every move. Go to:

https://www.google.com/maps/timeline

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?
  - If I use big data sources in my research, what ethical issues must I think about? Is my big data source representative?

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 guide:

https://hackblossom.org/cybersecurity/

Northeastern University
NULab for Texts, Maps, and Networks

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

Policy Implication Example:

COMPAS Risk Assessment Algorithm

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

Risk Assessment: Algorithmic Bias

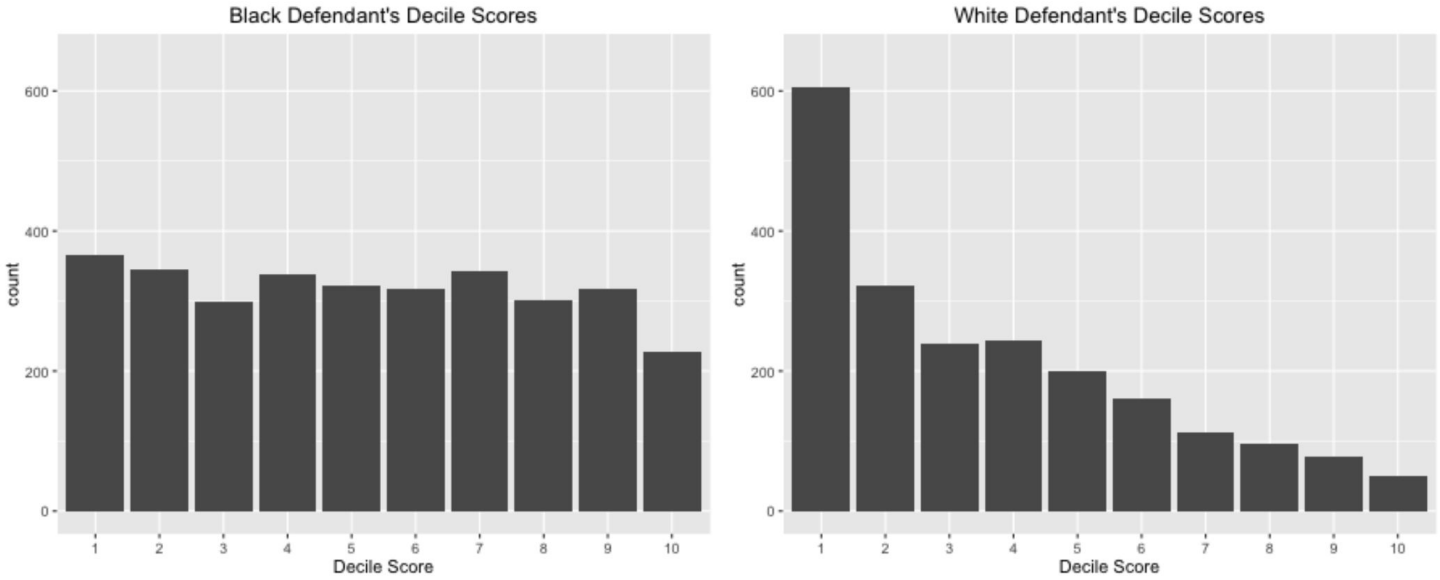
Risk assessment: used to determine the likelihood that someone will reoffend, not appear for trial, etc..

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…

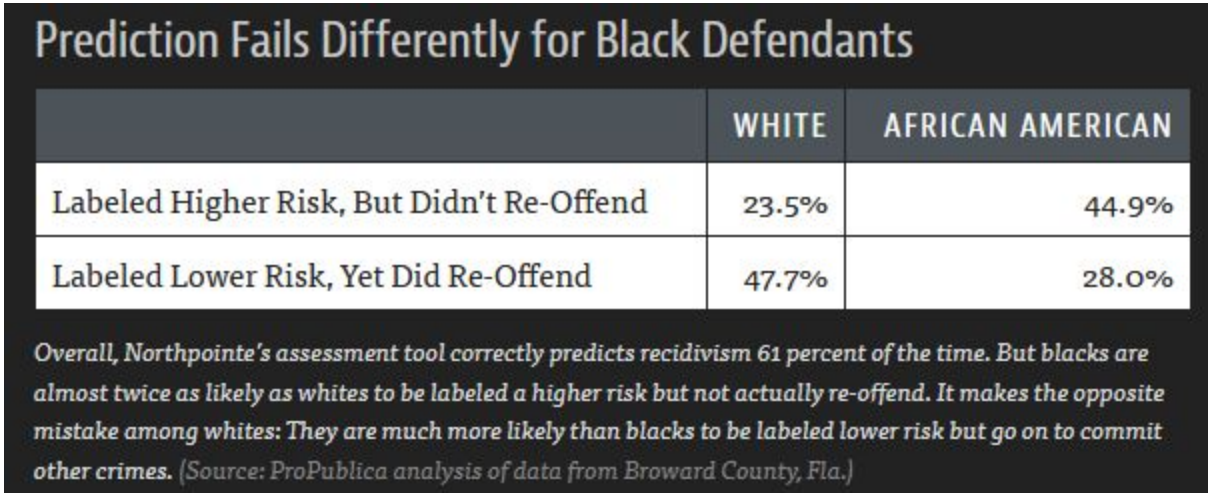


https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm

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

COMPAS Algorithm & ProPublica's Analysis

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



Prediction Fails Differently for Black Defendants

	WHITE	AFRICAN AMERICAN
Labeled Higher Risk, But Didn't Re-Offend	23.5%	44.9%
Labeled Lower Risk, Yet Did Re-Offend	47.7%	28.0%

Overall, Northpointe's assessment tool correctly predicts recidivism 61 percent of the time. But blacks are almost twice as likely as whites to be labeled a higher risk but not actually re-offend. It makes the opposite mistake among whites: They are much more likely than blacks to be labeled lower risk but go on to commit other crimes. (Source: ProPublica analysis of data from Broward County, Fla.)

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!

Algorithms and Bias

Northeastern University
NULab for Texts, Maps, and Networks

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

Activity: Adopt or Not?

Small Group: Find a partner or two! You all work for an adoption agency and have to decide if someone can adopt a dog. On your handouts, please read the four previous adoption applications and decide if the new applicant can adopt or not.

Do you think this new applicant should be allowed to adopt a dog? Why or why not?

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

Class Discussion: Adopt or Not?

- Would you ACCEPT or REJECT their application? Why?
- What questions from the application did you weigh more? Why?
- What might be some implicit biases in this application form, the process, and in your choices?

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

Adopt or Not? Algorithm

Algorithms can “read” through data such as these applications, and help us make decisions. Here are some questions to think about when assessing algorithms:

- Where might you see these algorithms being used to make decisions? Why are they being used? What are they replacing or adding on to?
- What biases may be ingrained in the data collected for the algorithms? What biases may be ingrained in the actual process of using the algorithms?
- In what ways might the algorithms prevent or reinscribe human biases?

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

Want to learn more about accountability and best practices when creating algorithms?

Visit https://www.fatml.org/, or Fairness, Accountability, and Transparency in Machine Learning

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!

Using Big Data in Your Own Research: Opportunities and Vulnerabilities

Opportunities

- Massive & Passive
- “Nowcasting”
- Data on social systems themselves
- Natural and Field Experiments
- Making big data small

Vulnerabilities

- Generalizability
- Too many big data
- Artifacts, reactivity, and drift
- Ideal user assumption

Source: Lazer and Radford, 2017

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

Thank you!

If you have any questions, contact us at:

Garrett Morrow
DITI Research Fellow
morrow.g@husky.neu.edu

Jeffrey Sternberg
DITI Research Fellow
sternberg.je@husky.neu.edu

Slides, handouts, and data available at
http://bit.ly/diti-spring2020-badruzzaman2

Schedule an appointment with us! **http://calendly.com/diti-nu**

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