Ethics in Data Science

UNIVERSITY OF
SAN FRANCISCO

Abbie M. Popa

BSDS 100 - Intro to Data Science with R

Outline

- Why Ethics?
- Ethics Frameworks
- Ethics Examples

Part I: Why Ethics

Why Ethics?

- Consent and Ownership
- Biased Algorithms

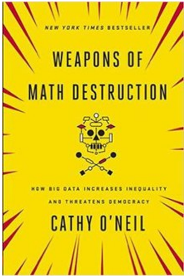
Consent and Ownership

- A fake example... imagine Google makes an algorithm to better recognize faces, they sell it to the police to use as a tool to help them find individuals with outstanding warrants. Google receives $250 million per year for this software
    - As someone who uses Google, can they use photos I stored in google photos without my consent?
    - Do I have the right to ask them to remove my photos from their training set?
    - If they use my photos to train their algorithm, should I receive a share of the profits?
- Does the answer change if the algorithm will directly benefit or harm me?

Biased Algorithms

- A fake example... imagine I own a company and need to hire a software engineer. I know that I may have an implicit bias toward hiring certain individuals so I decide to screen resumes using an algorithm. To train my algorithm I give it the resumes of the last 100 software engineers who applied to the company the last time we had an opening, labelled with whether we made them an offer or not.
- Why might this produce bias?
  - The training data will already contain my implicit biases, i.e., if I am more likely to hire men, the training set will contain more "hired" men than "hired" women
- But wait, if humans are already biased, why might this be worse?

Biased Algos Versus Biased Humans


NEW YORK TIMES BESTSELLER
WEAPONS OF MATH DESTRUCTION
HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY
CATHY O'NEIL

In *Weapons of Math Destruction* Cathy O'neill observes algorithms have a particular tendency to reinforce and increase bias due to three factors:

1. Opacity
2. Lack of regulation and difficulty of contesting
3. Scalability

Initial Thoughts

- Do you feel it's important for individuals to consent to how their data are used?
- Are you concerned about biased ML algorithms?
- Why?

Part II: Ethics Frameworks

Possible Ethical Frameworks

- Based on Cathy O'Neill's Book
- Data for Democracy Guidelines
- GDPR
- Others?

From WMD

- **Opacity**: we must understand why the algorithm is doing what it's doing, no black boxes
- **Lack of regulation and difficulty of contesting**: We must be able to "argue" with a decision made about us, such as to not hire
- **Scalability**: Algorithms have the potential to affect many more people than human decisions (automation), we must ensure we aren't causing a dangerous cycle
  - From WMD: "If a poor student can't get a loan because a lending model deems him too risky (by virtue of his zip code), he's then cut off from the kind of education that could pull him out of poverty, and a vicious spiral ensues."

From Data For Democracy

- For more see http://ethicspledge.wpengine.com/
- **Fairness**: understand, mitigate, and communicate the presence of bias
- **Openness**: transparent practices, community engagement, responsible communication
- **Reliability**: understand what is in the data, where it came from, and how it was created
- **Trust**: maximize informed participation
- **Social Benefit**: consider the impact on individuals, communities, and world-at-large

GDPR

- European Union: General Data Protection Regulation (GDPR)
  - **No Automated Decision Making**: There must be a human in the loop
  - **Right to Explainability**: Many ML models are so complex it's difficult, if not impossible, to explain them. Under GDPR the individuals involved in the data have a right to understand the models.
  - **Legal Basis for Processing Data**:
    - "legitimate interest" e.g., fraud prevention
    - Explicit consent (that's why you got all those e-mails last spring)
- California also passed it's own digital privacy law, similar but not as strict (users must ask for information from companies or to be removed)

Some Edge Cases

- Even if you don't include something, say, race, in your model, other elements like zip code or first name may be indicators of that factor. Is that okay?
- What about manipulations? Do we need to get consent if we show one group of users one version of our website and a different group a different version and measure their behavior?
- Control of algorithms, if I own a self-driving car can I edit its algorithms?

Thoughts on Frameworks

- What are the differences or similarities between these frameworks?
- Do you prefer one?
- Is there anything that you feel is missing or you would add?
- Do you have an alternate framework I missed?

Part III: Ethics Examples

Examples of Data Ethics

- Bail/Parole/Sentencing Algorithms
- Facebook Experiments on People
- Automatic Soap Dispensers

Bail/Rehabiltiation/Sentencing Algorithms

- Increasingly, court systems are relying on algorithms to make recommendations regarding bail, rehabilitation, or sentencing
- Beginning in October of 2019, California will eliminate cash bail. Instead, an algorithm will recommend whether the accused should be released or remain in jail
- In several states, length of sentence may be recommended based on output of an algorithm
- In Rhode Island, the same algorithm is used to determine which of the convicted should receive certain rehabilitative services
- Further Reading:

  http://www.abajournal.com/magazine/article/algorithm_bail_sentencing_parole/

Bail/Rehab/Sentencing Algorithms

- This algorithm does not directly inquire about race, but certain factors (such as social connections) are correlated with race
- This algorithm does include gender
- The exact algorithm is proprietary and unshared
- What do you think, is this an ethical algorithm?
- Does it matter whether the algorithm is being used to determine bail, sentencing, or rehabilitative services?

Facebook Experiments on People

- Facebook consistently tinkers with what users see, the format of facebook, what's on top in their newsfeed, etc. Some users will be assigned to one group, while others are assigned to a different group
- Facebook then measures behavior in both groups to see if the group assignment affected them

Facebook: Political Mobilization

- In 2012 Facebook showed users a message that their friends voted, a reminder to vote, or nothing, and measured effect on voting behavior (confirmed with public voting records). They found an increase of 340,000 votes, for context, Clinton needed to switch 53,650 votes to win the 2016 election.
- Further Reading (Political Mobilization):

  https://github.com/abbiepopa/BSDS100/blob/master/reading/FB-Bond.pdf

Facebook: Emotional Contagion

- Facebook showed some users more "sad" statuses and some users more "happy" statuses in their newsfeed. They found those exposed to sad statuses were more likely to post sad statuses and vice versa
- Further Reading (Emotional Contagion):

  https://github.com/abbiepopa/BSDS100/blob/master/reading/FB-Kramer.pdf

Facebook Experiments on People

- If I wanted to try these things, I would have to get approval from an Institutional Review Board and get consent from my participants, should facebook be held to the same standard?
- Does it matter whether the experiment has the potential to have widespread effects (e.g., influencing an election)?

Automatic Soap Dispensers

- Automatic soap dispensers are trained to detect a hand beneath them and release soap
- Video shows a soap dispenser working well for those with light skin but not those with dark skin
- May seem trivial when it's a soap dispenser, but what about ML algorithms trained to detect skin cancer?
- Further Reading: https://bit.ly/2DeMDJe

Ethics Lab

- In groups, select your favorite ethics framework and example
- Does the example you chose violate any of the standards in the framework you chose?
- Are there any other reasons it is or is not ethical?
- What are the potential consequences (for individuals or society)?
- What could be done to mitigate the problem? Are there reasons a company might not want to use the mitigation strategy you propose?
- Save your answers in a word document, one file per group, with all group members names at the top of the document. Upload to Canvas.

Data for Good

- To end on a more upbeat note... some are trying to use data for good
- Applying big data techniques to solve social problems
- Example, fighting illegal deforestation:

  https://www.blog.google/technology/ai/fight-against-illegal-deforestation-tensorflow/
- Getting data tools to non-profits or communities with low access