CSE 4/587
Data Intensive Computing

Dr. Eric Mikida
epmikida@buffalo.edu
208 Capen Hall

Day 25
Bias in Data

Announcements and Feedback

- Course Evaluations (and potential bonus points)

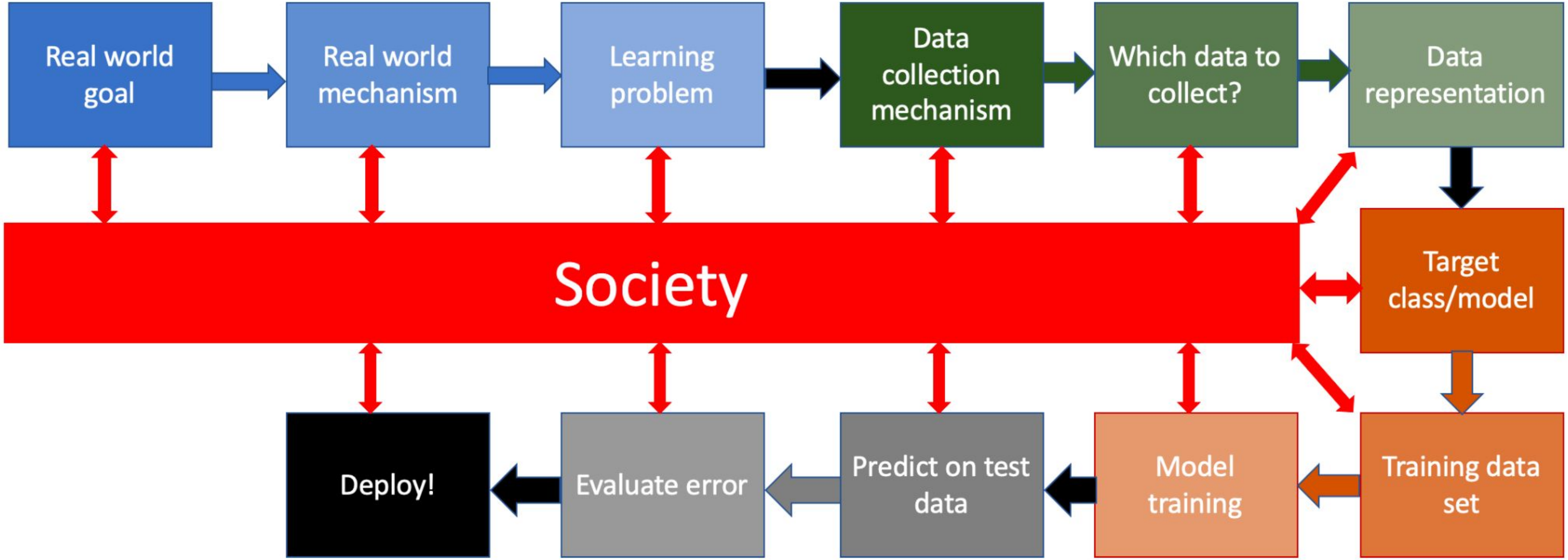
Our Products and Society

So you've just made your data product and want to release it into *society*

What factors should you be considering as you plan to have your product interact with society for the first time?

…wait…is this even the first time it's interacted with society??

Machine Learning Pipeline (CSE 440/441)



http://www-student.cse.buffalo.edu/~atri/ml-and-soc/support/notes/pipeline/index.html#pipeline-soc

Algorithms (and data) are BIASED

An experiment shows that Google Translate systematically changes the gender of translations when they do not fit with stereotypes. It is all because of English, Google says

https://algorithmwatch.org/en/google-translate-gender-bias/

Algorithms (and data) are BIASED

Facial-recognition systems are more likely either to misidentify or fail to identify African Americans than other races, errors that could result in innocent citizens being marked as suspects in crimes.

http://news.mit.edu/2018/study-finds-gender-skin-type-bias-artificial-intelligence-systems-0212

And though this technology is being rolled out by law enforcement across the country, little is being done to explore—or correct—for the bias.

https://www.theatlantic.com/technology/archive/2016/04/the-underlying-bias-of-facial-recognition-systems/476991/

Our Data Products have REAL Impacts

In 2016, ProPublica published an article titled Machine Bias, which studied a software called COMPAS that was used to predict recidivism. It showed a bias against black defendants when compared to white defendants.

https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Our Data Products have REAL Impacts

Driverless cars:
Who should die in a crash?

https://www.bbc.com/news/technology-45991093

Biases in Data

Data can and will be biased!

Some questions to consider:

- What data have you collected?
- What was the method of collecting the data?
- Who/What is represented by your data? (and who/what is not?)
- What is the target population (and does it match the who/what)?
- How are you evaluating your models/algorithms?

Types of Bias

1. How far is the distribution of values in our data set from that of a truly random dataset? (this bias is necessary for our algorithms to work at all in the first place)
2. Statistical Bias - basically how well do the outcomes of our models reflect the actual distributions of our target variables.
3. **The colloquial use of the term bias - is our model "fair" or not?**
   a. **Where do these biases come from?**

Sources of Bias in the Colloquial Sense

1. Historical bias
2. Representation bias
3. Measurement bias
4. Aggregation bias
5. Evaluation bias
6. Deployment bias

Historical Biases

- These are the biases that already exist in society (independently from our data science/ML pipeline)
  - Even if we could perfectly sample the data, the bias would still exist!
- Examples in the US:
  - Redlining (https://youtu.be/O5FBJyqfoLM)
  - Hiring practices (https://www.survivalofthebestfit.com/)

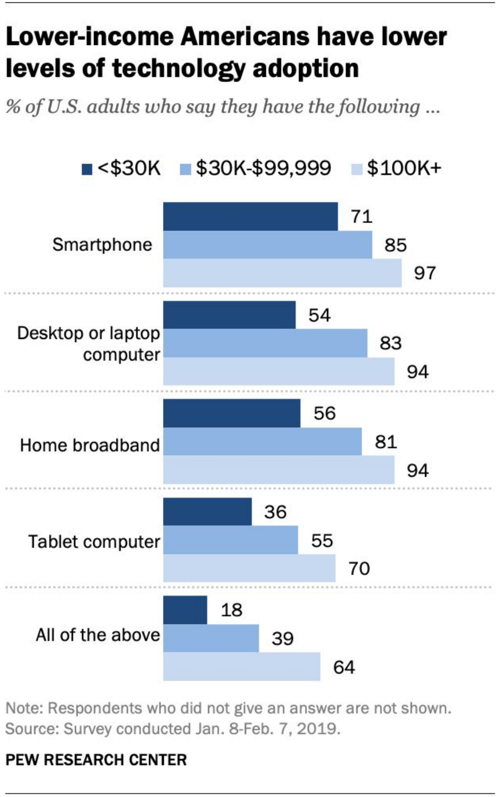
Representation Bias

- These forms of bias occur when certain groups of people might be under-represented in the dataset you are using to train your models
- What might cause this?
    - Data collection excludes part of the population
    - Training data population does not match deployed population
- Examples
    - Collecting from smartphone users (http://www.streetbump.org/)
    - Tweets pre-, post- and during hurricane Sandy
    - Voter survey data given at the polls
    - Facial-recognition software

Representation Bias

Examples

- Collecting from smartphone users (http://www.streetbump.org/)
- Tweets pre-, post- and during hurricane Sandy
- Voter survey data given at the polls
- Facial-recognition datasets (http://www.image-net.org/ 1%-2% Chinese and Indian faces, 36% of global population)



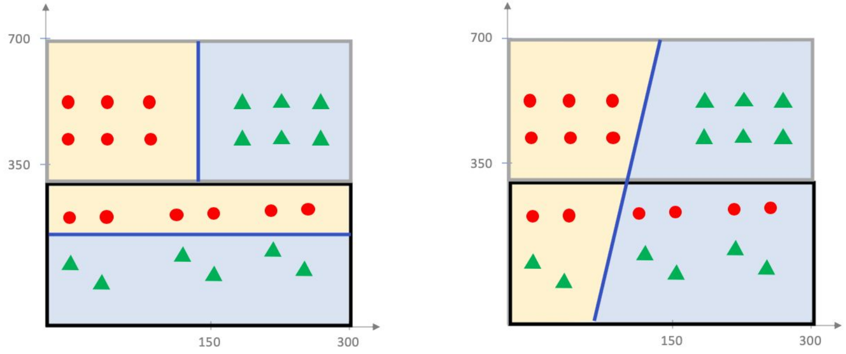
Measurement Bias

- These forms of bias occur when the values in the data do not exactly match up to what we actually wanted to measure
- Often, measuring the actual underlying values is difficult/infeasible/expensive so we measure a "proxy" instead
  - Using GPA as a proxy for "success"/"intelligence"
  - Treating re-arrest as a proxy for re-offending
  - Diagnosed with a certain condition as a proxy for having the condition
  - Predictive Policing (https://youtu.be/ZMsSc_utZ40)

Aggregation Bias

- These are biases that can occur when we attempt to use one model across multiple groups, leading to sub-optimal prediction for certain groups

For the dataset to the right, a decision tree could perfectly group the two types of data points, but a simple linear model would fail to adequately classify at least one of the groups



Evaluation Bias

- These are biases that are introduced by how we evaluate our models
    - This can occur both via our choice of test dataset, and evaluation metric
- Imagine a population with two groups, where group A constitutes 95% of the population and group B constitutes 5% (ie Finland)
    - A system that perfectly classifies population A, and mis-classifies population B will have a 95% accuracy…
    - We need to be careful in how we choose our evaluation metrics
    - See also: http://gendershades.org/

Deployment Bias

- These are biases that occur when assumptions made about society are not true - the product is not used "as intended" when deployed into the real world
  - See https://en.wikipedia.org/wiki/Tay_(bot)

How can we combat these biases?

First and foremost, we must be aware that our data is biased…

Sometimes this is, in some sense, the best we can do

- For historical biases, we cannot simply make them go away. We will probably have to explicitly handle them outside of our models in some way - ie diversity initiatives in conjunction with automated hiring

How can we combat these biases?

- For representation biases, we **must** understand the nature of our data
  - Where did it come from?
  - How was it collected?
  - What does it represent?
  - Do we have any control over the collection process?

How can we combat these biases?

Some types of bias MUST be handled in the data science pipeline, because their introduction into our models is solely due to our own methods

- For measurement biases, make sure the variables you are measure are actually the variables you are trying to measure. If not, at least pick variables as close to the target as possible, and be cognizant of the biases you may be introducing.

How can we combat these biases?

Some types of bias MUST be handled in the data science pipeline, because their introduction into our models is solely due to our own methods

- For aggregation biases, make sure the models you are picking are appropriate, and if necessary pick different models for different groups represented in the dataset.

How can we combat these biases?

Some types of bias MUST be handled in the data science pipeline, because their introduction into our models is solely due to our own methods

- For evaluation biases, make sure your test dataset is independent from your training dataset, and be careful how you evaluate the models. If necessary, evaluate based on multiple different metrics.

How can we combat these biases?

Deployment biases by their very definition cannot be handled just within our data science pipeline, but the following paper has some notes about common "traps" to avoid:

> Andrew D. Selbst, danah boyd, Sorelle Friedler, Suresh Venkatasubramanian and Janet Vertesi, Fairness and Abstraction in Sociotechnical Systems . In FAT*2019.

Reference

http://www-student.cse.buffalo.edu/~atri/ml-and-soc/support/notes/society/index.html