

Lab 15 Preparation

Your Name

November 11, 2021

About this upcoming lab

Choose one of the two algorithmic applications below that you are most interested in thinking about and discussing in class with a small group of your peers. Videos and/or readings are provided for each topic, as well as the questions you'll consider with your peers in class. **You should complete this preparation ahead of class on Thursday!** I strongly encourage you to take notes on the questions as you watch/read/listen, so you can refer to them during your in-class discussion (you could take notes here, for example!).

Option 1: [Predictive Policing and Predicting Recidivism](#)

Option 2: [Predicting Financial Risk](#)

The resources and questions provided are designed to increase awareness around the complexity of the ethical considerations in data science, and around algorithms in particular.

Please complete this very brief [Google form](#) to specify your choice.

Option 1 **Predictive Policing and Predicting Recidivism**

Watch the following videos:

The videos below contain content related to policing in the US and focus on racial bias present in algorithms.

- [Joy Buolamwini: How I'm Fighting Bias in Algorithms](#) (up to: 7:00 minutes)
- [CNBC: How AI Could Reinforce Biases in the Criminal Justice System](#) (~ 8:30 minutes)
- [Henry Farid: The Danger of Predictive Algorithms in Criminal Justice](#) (up to 16:10 minutes)

- 1.1 In Buolamwini's facial recognition example, there seems to be a straightforward solution: expand the training set to include diverse individuals. What factors do you think have prevented this "straightforward solution" to being implemented more readily?
- 1.2 Do you think there is an analogous solution to the bias present in the predictive policing and recidivism algorithms? Why or why not?
- 1.3 In the CNBC video, McDonald states that "[Data and algorithms] are always going to be a more effective, more fair, and more just means of directing patrol officers than relying on their intuition or their hunches". In what ways is using data and algorithms in this context more effective and more fair? In what ways is using data and algorithms in this context less effective and less fair? (Consider how you are defining effective and fair.)
- 1.4 In Farid's TedX talk, Farid argues that biased decisions based on computer algorithms are worse than biased human decisions. Why? What are his justifications for that claim?
- 1.5 Both humans and algorithms are flawed. Which do you think is "worse" in this context? What about if the algorithm was going to determine whether or not you get into graduate school? Or whether or not you get the next job or summer internship that you apply to?

Option 2 **Predicting Financial Risk**

Proceed in this order:

1. Read [McKinsey & Company: Derisking Machine Learning and Artificial Intelligence](#) (~ 8:00 minutes)
 - Note that there may be a number of terms you haven't heard before (Shapley values, gradient boosting, random forest) or don't understand in the McKinsey article. The details of, for instance, a random forest model, are not important for purposes of this discussion—rather, focus on the introduction and the section around bias. (See questions below.)
 2. Read [Wall Street Journal: What Your Face May Tell Lenders About Whether You're Credit Worthy](#) (~4:00 minutes)
 3. Watch [Joy Buolamwini: How I'm Fighting Bias in Algorithms](#) (up to: 7:00 minutes)
 4. Listen to [Cathy O'Neil: Do Algorithms Perpetuate Human Bias?](#) (12:00 minute listen)
-
- 2.1 In the McKinsey article, interpretability and bias are two elements added to the framework for validating machine-learning-model risk. Consider the four definitions they offer for what constitutes fairness: demographic blindness, demographic parity, equal opportunity, and equal odds. Can you think of any other ways to measure fairness? Do you think it is possible for an algorithm to satisfy all notions of fairness at the same time?
 - 2.2 In the WSJ article, the facial recognition technology developed by Ping An is used to gauge customers' health, which in turn can affect pricing of their insurance policy (e.g., discounts on the monthly premium if $BMI < 30$). In what ways is using facial recognition data and algorithms like this more effective (e.g., consider potential advantages to the consumer and the company) and more fair? In what ways is using data and facial recognition algorithms like this less effective and less fair? (Consider how you are defining effective and fair.)
 - 2.3 In Buolamwini's facial recognition example, there seems to be a straightforward solution: expand the training set to include diverse individuals. What factors do you think have prevented this "straightforward solution" to being implemented more readily?
 - 2.4 Do you think "expanding the training set" would solve any potential fairness issues in the context of Ping An's use of facial recognition? What about if the algorithm was going to determine whether or not you get into graduate school? Or whether or not you get the next job or summer internship that you apply to?
 - 2.5 The McKinsey article notes, "Machine learning models typically act on vastly larger data sets, including unstructured data such as natural language, images, and speech. The algorithms are typically far more complex than their statistical counterparts . . .". It is the complexity of these algorithms that Cathy O'Neil argues makes them so dangerous. Why? What are her justifications for that claim? Do you agree or disagree? Please explain.