Introduction to Machine Learning for Epi

Assignment 3 due Tuesday February 2, 2021 by 5:29 PM

**Sydney Ng sn2863**

1. Should users of social media, online applications and/or search engines be notified about the potential use of their personal data (even in aggregated form) for public health measures? Why/why not?

Social media, online application, and search engine users should be notified about the potential use of their personal data, no matter in what form for public health measures. In the case of this company’s health information application, the start-up is collecting their medical records and other vital measures that should be private, confidential, and protected by HIPAA. It is unethical to use these data for analysis without some sort of informed consent, because the analyses done using these data is still research with human subjects. The company should at least put in their terms and conditions for the app that the data will be used in such a manner, including how they will store the data and their specific uses.

Regarding the extent to which users should be notified, this can be tricky. Playing the devil’s advocate, I also believe that not everything needs to be fully explained in all of its detail. In the case that we enacted full informed consent, it would be difficult to explain the data analysis methods in full detail to the lay public. Although I do think it is important for users to be notified about how their data will be used, it may be better to communicate the details in a high-level overview manner.

1. What are some of the ethical risks of the proposed big data analytics or algorithms discussed in this case study, specifically focused on outbreak tracking? What might be done to mitigate the risks?

An ethical risk of using the smartphone application and big data/machine learning algorithms on outbreak tracing is that it can leave behind the lower-income individuals and households who do not quite have access to smartphones, watches that can track sleep, and healthcare in general. As a result, the data that the start-up captures will be comprised of those who have good health records and have smartphones and smart watches to track their health information. Also, it is important to consider that people who track their eating habits tend not to be the people who worry about their next meal being out of their control. Food security issues are lower-income issues.

Two things that can be done to mitigate these risks are to get a more representative sample of the entire population and to recognize the biased sample when doing the analyses and creating the algorithms. More specifically, the start-up company can consider offering discounts to residents of certain areas for their app/smartphone services to open access up to lower-income neighborhoods that cannot afford the application otherwise. Furthermore, the start-up’s analysis team needs to be transparent about their methods and data collection so their results do not get misconstrued and misused when tracking and responding to the epidemic.

1. What are some of the ethical risks of the proposed big data analytics or algorithms discussed in this case study, specifically focused on outbreak tracking? What might be done to mitigate the risks?

One opportunity for for-profit companies collecting, managing, and analyzing large health data can lie in the creation of their machine learning algorithms. In particular, when they define “success” in their analyses, they can define and maximize success in a way that is not to advance public health and can cause greater disparity. Pulling from our Weapons of Math Destruction discussion, for-profit companies may define success based on cost over care like in the case of UnitedHealth and the company’s racial bias lawsuit.

A potential policy/action I particularly like is to mitigate this risk is forcing companies to have an outside company audit their algorithms and methods. I am not exactly sure if I fully support forcing companies to publish their methods and results, however, I do believe that there needs to be another eye on the data and analyses by a third party. This is similar to the pharmaceutical drug development process, in which the FDA requires Data and Safety Monitoring Board (DSMB) support to protect patient safety in clinical trial analyses. By having such a policy in place, algorithmic audits as mentioned by Cathy O’Neil will mitigate the further perpetuation of health inequities via race and socioeconomic status.