Week 3 Assignment

Question 1

I think that the users of these applications and platforms should definitely be notified that their personal data will be used for public health measures. Notifying users would show the most integrity, regardless of whether that will dissuade some users from sharing their data or not. The users have a right to know what their data will be used for, and it is up to them to decide whether they would like to contribute or not. Even if the data is aggregated or de-identified, I think that users should at least know and agree to such terms and conditions. Only when this relationship between the user and the provider is equal (i.e., there is symmetry of information), then the user can respect the provider’s needs for data and the provider can respect the user’s needs for privacy.

Question 2

There are a few possible ethical issues that may entail using big data analytics for outbreak tracking using personal smartphone data. Firstly, there is always the bias in data analytics of any kind – even big data from smartphone tracking – that is just inherent in the way we design algorithms. For statisticians to solve problems, they must make certain decisions and assumptions, which I assume would be especially critical for this case. Here, Zhang has a task to relate eating, sleeping, social and other habits to disease outbreak, and those relationships are not very clear cut. In this case, he will have to make some decisions that may be biased – for example, maybe he has to determine whether a person is starting to feel ill (during an incubation period). There would have to be assumptions for things such as a normal person’s habits, which may not be generalizable to all people. Maybe an incorrect prediction of being exposed leads to unintended anxiety, or on the other hand, an incorrect prediction of no exposure leads to a false sense of security.

It would be very hard to completely eliminate this bias, since algorithms will always have some assumptions. But to mitigate these risks as best as possible, one option would be to avoid giving prescriptive advice, but rather recommend health-promoting behavior. For example, instead of evaluating whether someone is likely to be ill or not, the algorithm may instead remind the person to visit the doctor. In this way, the biases in the data will essentially be diluted since the algorithm will not make hard predictions – and the purpose will be more user-centered, not “self-serving”.

Question 3

When for-profit companies start managing the health data of large populations, we need to start being a little bit skeptical. Since companies’ jobs are to generate money, there is definitely a need to look closer at their methods and whether they are fair. One problem is the lack of transparency that these companies enjoy, typically since their algorithms are kept secret for business purposes. In this way, they are essentially free to introduce their own biases, both for generating revenue through advertising or even pushing political agendas, which may sometimes even include targeting of specific groups and demographics. This ties in with the lack of regulation. Because major tech companies typically have lots of money and influence in the form of user support, it is hard to enforce certain rules that prevent them from misusing the data. However, one policy that could be introduced to mitigate this would be to notify users about advertising policies (such as “Why am I seeing this ad?”), and it must come with a breakdown of how the user’s data contributed to the targeted ad. I think that this is a fair solution because users should know how they are being influenced by their own consumption, and they should have the right to know or review their own consumption habits. At the same time, the companies have the freedom to advertise, as they need revenue to operate, and all they need is to be accountable with the data they use to target the ads.