It's not big data that discriminates – it's the people that use it

22 March 2016, by Reuben Binns, University Of Oxford

Credit: Shutterstock

Data can't be racist or sexist, but the way it is used can help reinforce discrimination. The internet means more data is collected about us than ever before and it is used to make automatic decisions that can hugely affect our lives, from our credit scores to our employment opportunities.

If that data reflects unfair social biases against sensitive attributes, such as our race or gender, the conclusions drawn from that data might also be based on those biases.

But this era of "big data" doesn't need to to entrench inequality in this way. If we build smarter algorithms to analyse our information and ensure we're aware of how discrimination and injustice may be at work, we can actually use big data to counter our human prejudices.

This kind of problem can arise when computer models are used to make predictions in areas such as insurance, financial loans and policing. If members of a certain racial group have historically been more likely to default on their loans, or been more likely to be convicted of a crime, then the model can deem these people more risky. That doesn't necessarily mean that these people actually engage in more criminal behaviour or are worse at managing their money. They may just be disproportionately targeted by police and sub-prime mortgage salesmen.

Excluding sensitive attributes

Data scientist Cathy O'Neil has written about her experience of developing models for homeless services in New York City. The models were used to predict how long homeless clients would be in the system and to match them with appropriate services. She argues that including race in the analysis would have been unethical.

If the data showed white clients were more likely to find a job than black ones, the argument goes, then staff might focus their limited resources on those white clients that would more likely have a positive outcome. While sociological research has unveiled the ways that racial disparities in homelessness and unemployment are the result of unjust discrimination, algorithms can't tell the difference between just and unjust patterns. And so datasets should exclude characteristics that may be used to reinforce the bias, such as race.

But this simple response isn't necessarily the answer. For one thing, machine learning algorithms can often infer sensitive attributes from a combination of other, non-sensitive facts. People of a particular race may be more likely to live in a certain area, for example. So excluding those attributes may not be enough to remove the bias.

But more importantly, if sensitive attributes are included in an analysis, it's not clear that the algorithm would be to blame for any unequal outcomes. It all depends on how the algorithm is used and interpreted in practice. In O'Neil's example, the algorithm simply predicts that homeless black families are less likely to get jobs. It's up to the service providers to work out why that might be and how to respond to it. And that's where the ethical considerations come in.

If service providers were to assume that the higher unemployment rate was due to lack of talent or effort, that would almost certainly be wrong. If they then decide to stop offering job counselling to black homeless families on that basis (as O'Neil suggests they would), that would also be unethical. But neither of these outcomes would be justified or dictated by the algorithm. They are assumptions and choices influenced by human bias and ignorance.

Using big data ethically

An enlightened service provider might, upon seeing the results of the analysis, investigate whether and how racism is a barrier to their black clients getting hired. Equipped with this knowledge they could begin to do something about it. For instance, they could ensure that local employers' hiring practices are fair and provide additional help to those applicants more likely to face discrimination. The moral responsibility lies with those responsible for interpreting and acting on the model, not the model itself.

So the argument that sensitive attributes should be stripped from the datasets we use to train predictive models is too simple. Of course, collecting sensitive data should be carefully regulated because it can easily be misused. But misuse is not inevitable, and in some cases, collecting sensitive attributes could prove absolutely essential in uncovering, predicting, and correcting unjust discrimination. For example, in the case of homeless services discussed above, the city would need to collect data on ethnicity in order to discover potential biases in employment practices.

Computer scientists are just beginning to discover the ways that machine learning can be used to both detect and mitigate the effects of discrimination. Coupled with a strong understanding of the dynamics of discrimination and strong legal and governance frameworks, the next generation of data scientists could avoid past and present injustices. Rather than entrenching inequalities, big data might just help us overcome them.

This article was originally published on The Conversation. Read the original article.

Source: The Conversation

APA citation: It's not big data that discriminates – it's the people that use it (2016, March 22) retrieved 21 June 2021 from https://phys.org/news/2016-03-big-discriminates-people.html