Ben Puryear

ethics.pdf

DA3 – ethics.pdf

1. The Introduction states, "...attempting to reduce human behavior, performance, and potential to algorithms is no easy job." The parents loved Sarah Wysocki, but the algorithm did not. For this algorithm, Sarah was an "instance" and there were several attributes used to describe her. These attributes were weighted to produce an IMPACT score. What do you think some of the attributes were? Can you think of attributes that could better represent "human behavior, performance, and potential"? Is it possible to fully code a teacher's impact on student learning? What might a feedback loop look like for the IMPACT WMD?

The case of Sarah Wysocki is just one example for how differently computers and humans think. There is a difference between how good a teacher is at teaching their students, and how approachable and caring the teacher is. The parents might praise Wysocki because of how kind she is to the students, but if that comes with the sacrifice of test scores or content covered, the algorithm will produce a lackluster score. Whereas parents of fifth grade students might be willing to sacrifice a slight amount of their child's education in order for them to experience more love and affection in their life, a computer program can not calculate that, and therefore it is left out of the algorithm. All the computer sees is that the students aren't learning as much as other classes, and thus determines that Wysocki is a subpar teacher.

A feedback loop that could result from this could be strictness. If a strict teaching style is being enforced by the IMPACT algorithm because its results in higher scoring students, then this could result in the students themselves being more critical and harsher on others. This could span generationally and evolve into a blunter society that, although successful, is not happy. Although this feedback loop is one that will take a very long time to see results.

2. The Introduction also states, "...data scientists all too often lose sight of the folks on the receiving end of the transaction." What is an example of an automated data-based system where you were on the receiving end of the transaction? Do you think the output from the system used to label you was accurate and fair? Try to think of an example not mentioned in the Introduction.

An example in my life of an automated data-based system could me being recommended music by Spotify during their "Yearly Wrap-Up". Almost every year I have the same 3 most listened to artists, with changing orders, but Spotify's algorithm always ends up saying that I would love *insert EDM artist*. This makes sense because one of my most listened to artists is Porter Robinson, who started off as an EDM DJ, but who's music hasn't really fit into the definition of EDM for the last 10 years or so, with his more modern music being classified as Electronic and/or infused pop (slight difference in name, but the dance aspect of EDM is missing). This is the only example I could think of that was inaccurate (hence why it is a bit of a nitpicky stretch).

3. What else struck you about this introduction?

Most of this introduction made a lot of sense and wasn't terribly surprising. It did get me interested in reading the rest of the book with one of the ending lines "All of the life domains are increasingly controlled by *secret* models wielding arbitrary punishments". I really liked that sentence.