Class Concepts: Cumulative Disadvantage and Disproportionate Data Availability for Minority Groups

MEMORANDUM

TO: Joe Admissions
                    Director of Admissions, College University

FROM: Adam Bakopolus
                    Information Technology Director, College University

DATE: November 7th, 2021

RE: Google What-If Tool

        As the Admissions team discusses rolling out a new assessment tool that assigns a "readiness" score for higher learning, I'm writing to ensure that you have accounted for all ethical concerns. My current understanding from reading this tool's documentation is that the "readiness" score assesses not only a student's grades and test scores but also the prestige and rigor associated with their high school, a student family's household income, and a student's typical weekly diet, among other categories. From this, I have initial concerns that the tool may inappropriately place too much weight on factors outside of a student's control. As a result, this tool may not account for cumulative disadvantage, or "the ways in which historical disadvantages cumulate over time and across categories of experience" and unnecessarily bar admission regardless of a student's competency due to uncontrollable factors that should remain out of consideration.[1] To ensure that this tool does not unfairly score minority populations, I recommend your team consider the Google What-If Tool (WIT) to evaluate this new system prior to rollout.

        At a high-level, WIT assesses a tool's fairness and would have many advantages that could be leveraged by you and your team.[2] For example, the "Features" tab of WIT provides demographic information that quickly and easily informs your applicant pool. The percentage of White, Hispanic, etc. applicants can quickly be discerned and would highlight an important point around the "readiness" model. While the data used to generate the model is not available to us, it is a fair assumption to hold that minority groups had less available data considered in the model generation compared to the majority white population.[3] This again ties in with the notion of cumulative disadvantage, as minority groups would unfairly be compared to majority white populations who likely had far more resources and opportunities available to them and would, as a result, more often be deemed as "not ready" through no fault of their own. As discussions around this model's rollout with your team continues, I urge you to review the demographics of your applicant pool and ensure that appropriate measures are taken to remove bias.

        To that end, WIT's "Performance & Fairness" tab would provide the means of assessing the model's fairness across different slices of the applicant pool. For example, if you were to slice the pool by White and African American students and by SAT scores and saw that White students with scores in the 2100 – 2400 range

[1] Sandvig, Christian. "Cumulative Disadvantage and Protected Classes." *Coursera: SIADS 503: Data Science Ethics*, Oct. 2021, https://www.coursera.org/learn/siads503/lecture/5yC8Q/cumulative-disadvantage-and-protected-classes.
[2] *What-If Tool*, https://pair-code.github.io/what-if-tool/.
[3] Wallach, Hanna. "Big Data, Machine Learning, and the Social Sciences: Fairness, Accountability, and Transparency." *Medium*, Medium, 23 Dec. 2014, https://hannawallach.medium.com/big-data-machine-learning-and-the-social-sciences-927a8e20460d.

had a "readiness" score of 88 and African American students had a score of 67, this would be a clear sign that factors consistent with cumulative disadvantage, such as school prestige or family income, and outside of a student's control, are at play. WIT would offer the ability to slice across many demographic factors and would allow your team to assess the level of fairness in the model and that minority groups aren't disproportionately negatively scored. With these capabilities, I again strongly recommend your team consider the WIT tool.

Despite my recommendation, I do understand that there is a notable con to consider, as well, in implementing WIT to evaluate the scoring model. WIT does require an underlying understanding of the ethical concerns and questions that exist within data, and this knowledge, understandably, may not be currently present within the admissions team. A data scientist familiar with these types of ethical concerns would be a valuable resource to pull into the model implementation meetings moving forward to ensure that any existing fairness concerns in the model are identified and either updated or removed. However, overall, I still recommend strongly to consider using WIT, as admission to university is a life-altering event, and we must be as diligent as possible in ensuring that individuals are not negatively impacted during this process by factors that were and continue to be outside of their control.

References:

Sandvig, Christian. "Cumulative Disadvantage and Protected Classes." *Coursera: SIADS 503: Data Science Ethics*, Oct. 2021, https://www.coursera.org/learn/siads503/lecture/5yC8Q/cumulative-disadvantage-and-protected-classes.

What-If Tool, https://pair-code.github.io/what-if-tool/.

Wallach, Hanna. "Big Data, Machine Learning, and the Social Sciences: Fairness, Accountability, and Transparency." *Medium*, Medium, 23 Dec. 2014, https://hannawallach.medium.com/big-data-machine-learning-and-the-social-sciences-927a8e20460d.