Cumulative Disadvantage

To: Head of the Admissions Committee, XYZ University

From: V.P. Data Science Ethics, XYZ University

RE: Ethics and the What-If Tool

We are tasked with the unenviable responsibility of picking a very limited number of applicants for enrolment to the prestigious XYZ University from amongst thousands of applicants each year. I emphasize the word "responsibility" as we, the gatekeepers to the heart of the education world, have a personal responsibility to ensure that we conduct our due diligence to do everything we can in order to pick the most deserving of the applicants who are most likely to succeed in the academic program in a fair and ethical manner. As you know, we currently have a machine learning model which helps us to accomplish this using various data points voluntarily provided by the applicants. An internal ethics review of it is pending. Google's What-If Tool (WIT), while initially promising for a number of reasons listed below, also lacks in certain critical areas. Based on my analysis of this tool, I do not recommend employing the use of it in our internal ethics review.

Pros:

1. Easy to Use

   The WIT is catered towards individuals who do not require any prior machine learning experience in order to use it and interpret its findings. Some basic statistical knowledge is sufficient to make full use of its capabilities. The functionalities of the tool are fairly intuitive and the interface is straightforward, such that the members of the Admissions Committee can effectively explore the machine learning model's recommendations during application reviews without any hitches.

2. Transparency and Accountability

   The ease of use of the WIT would allow for greater transparency and accountability not only within the University regarding the internal ethics review, but also for external entities such as auditors, potential clients who want to make use of our model for their universities, and the general public (Wallach, 2014). Our willingness to be open and transparent in this regard would garner trust and goodwill among the public, adding to and acting as a testament for the excellent reputation of the University.

3. Data Slicing

   The WIT includes functionalities for basic data analysis like normality of skews, distributions, and exploratory visualizations which help expose patterns in our student data. It is well-known that trends that hold true for the majority might not hold for the minority (Wallach, 2014), which is why the slicing feature is useful to explore trends in the potential student population, including minority subgroups.

4. Error Correction

The tool has various features to enable error correction and adjustments for equality in subsets. The results of these adjustments are made immediately apparent and can be readily checked and changed as needed. Since some of the core values of the University are inclusion and diversity to offer equal opportunities for everyone to succeed, this might be a good functionality to employ while picking deserving applicants for enrolment in a fair and ethical manner. However, one should be cautious while using this feature as it could lead to either a very conservative or a very tolerant model, rendering the point of its development moot.

Cons:

1. Minimal Ethics Check

Using the WIT as the sole Holy Grail for ethical review would be a blunder. Implementing this tool internally without a clear policy regarding safeguards around its use is a slippery slope as it could easily replace other systems and act as a "catch-all", creating the illusion of a complete ethical analysis (Sandvig, 2020a), often leading to decisions which would be against the core values of the University. The use of the WIT alone does not qualify as an adequate ethics review of the applicant selection model.

2. Missing Data Quality Checks

This is one of the critical areas where I found the WIT to be seriously lacking. The WIT does not help identify proxies or substitutes in the data (for example, a zip code provided by the applicant in the Address field could be an indicator of their income or race). These proxies are unintended, extraneous factors which could affect the model's recommendations. Further, the WIT does not take into account how the data was collected and processed, the quality of the resulting dataset, and what kind of features are included in the dataset, which provides low confidence on its findings.

3. Cumulative Disadvantage

A deep analysis of the workings of the applicant selection model is required as it is important that the data (and the recommendations of the model) are representative of people from diverse walks of life, ensuring that they all have equal opportunities to be admitted into the University akin to the core values of the University. It is easier to glean patterns representative of the majority rather than the more subtle ones of the minority (Wallach, 2014), as a consequence of which the members of the Admissions Committee, who may be unfamiliar about the nuances of the subset groups, incorrectly use the slicing functionality in the WIT, leading to skewed results which do not favour the admission of minority applicants simply because there isn't enough data on them. This could add to a growing heap of historical disadvantages for that minority group or "cumulative disadvantage" (Sandvig, 2020b), which goes against the ethical principles of the University. We have less historic data from minority groups because other disadvantages such as lower income, lack of academic and professional growth opportunities, difficulty in affording quality education, tutors, or academic resources due to economic factors, lack of participation in extracurricular activities due to other commitments representative of socio-economic status

leading to limited holistic development, etc. have an impact on data collection and could lead to erroneous misrepresentations of the minority group, affecting their chances of admission and adding to their cumulative disadvantages.

Keeping these grave disadvantages in mind, I do not recommend using the WIT for our internal ethics review of the applicant selection model. By using the WIT indiscriminately, the sense that ethics were examined could lead us to ignore other important parts of the process, leaving us vulnerable to serious ethical lapses. A better move would be the appointment of an Ethical Review Committee who would act as advisors to the Admissions Committee for applicant selection, among other things. The Ethical Review Committee should consist of a diverse, interdisciplinary team of data scientists with experience in social analytics or sociology. A diverse team could provide a more nuanced examination of the data and the model than the What-If Tool is capable of producing (Wallach, 2014). This team would be able to identify trends in minority groups and to be cautious about potential proxies. They would be mindful of the way they collect and process the data, keeping the pitfalls discussed above in mind. This would allow for a more accurate generalization of the findings of our model, accounting for equal representation of minority groups in our student body to ensure that we continue to uphold the ethical principles and values this University has continued to thrive on.

References

Citron, D.K. & Pasquale, F. (2014) The scored society: Due process for automated predictions. Washington Law Review, 89(1).

Sandvig, C.. (2020a, March 15). Assignment introduction The What-If Tool [Video]. Coursera. https://www.coursera.org/learn/siads503/lecture/17SI3/assignment-introduction-the-what-if-tool

Sandvig, C. (2020b, March 15). Cumulative disadvantage and protected classes [Video]. Coursera. https://www.coursera.org/learn/siads503/lecture/5yC8Q/cumulative-disadvantage-and-protected-classes

Wallach, H. (2014, Dec. 14). Big Data, machine learning, and the social sciences: Fairness, accountability, and transparency. Medium.com.