Ethics 2

Assignment:

1. Read the dating casus
   Breeze, Algoritme Discriminatie
2. Formulate your first impression on the ethical issues and how they came about
3. Draw a DAG (or multiple DAGs) for this dilemma
4. After this exercise (nr.3), are there aspects that you missed in your first impression?
5. What recommendation would you give to a data scientist who gets an assignment like this?

First Impression on Ethical Consequences

The Breeze application provides matches based on input variables provided by users and the “like” behaviour of all users. These input variables and this “like” behaviour can already introduce loads of bias to the model, considering not all potential population groups are equally represented and some groups will form the minority, leading up to less seen profiles based on ethnicity.

This minority population experiences this bias in the algorithm, by getting a lot less matches and as a result they leave the Breeze platform, introducing even more bias because the already slim minority population is getting even smaller.

To circumvent this, legal bodies and Breeze itself are in dispute on how to solve this issue, first raising the idea of introducing a man made bias in favour of darker skinned individuals, however that in itself is also discrimination. The legal body mentions that this approach is suitable in this case because it does not harm the dominant population and is a way of compensating the minority population.

Even though consensus was made on solving the algorithm in this way, it will be difficult for Breeze to do this considering ethnical background is special personal information and using that in any way would be against GDPR legislation.

Breeze Ethical DAG

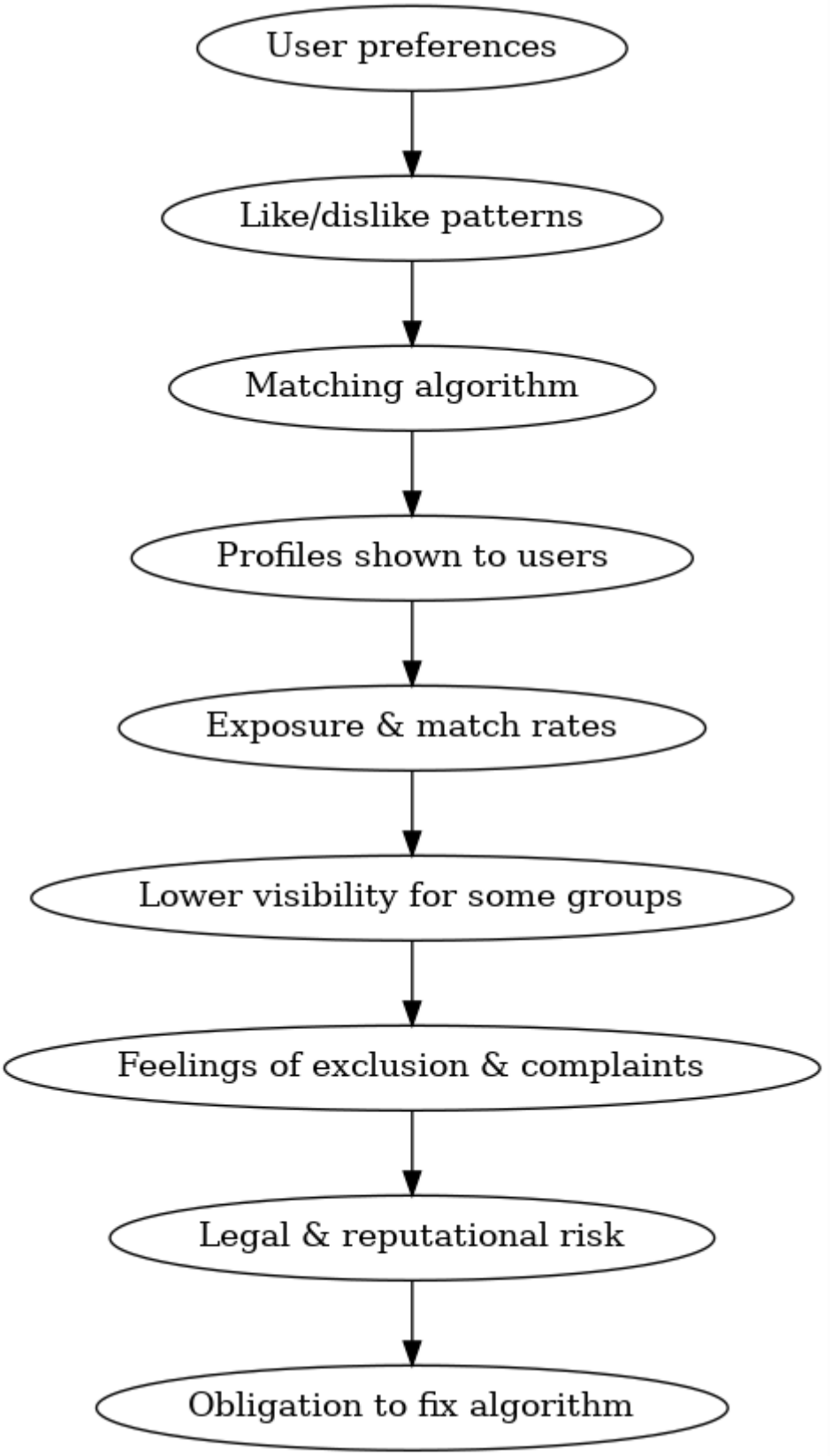

Figure 1 Breeze Ethical DAG

Missed Aspects

I have come to realise that because profiles of minority populations are shown less, the likelihood of them getting likes is also limited… introducing even more bias. So this feature of likes introduces loads of bias that will exponentially grow over time.

Recommendations

To provide recommendations we first need to understand the hiccups in the current setup, see figure 2.

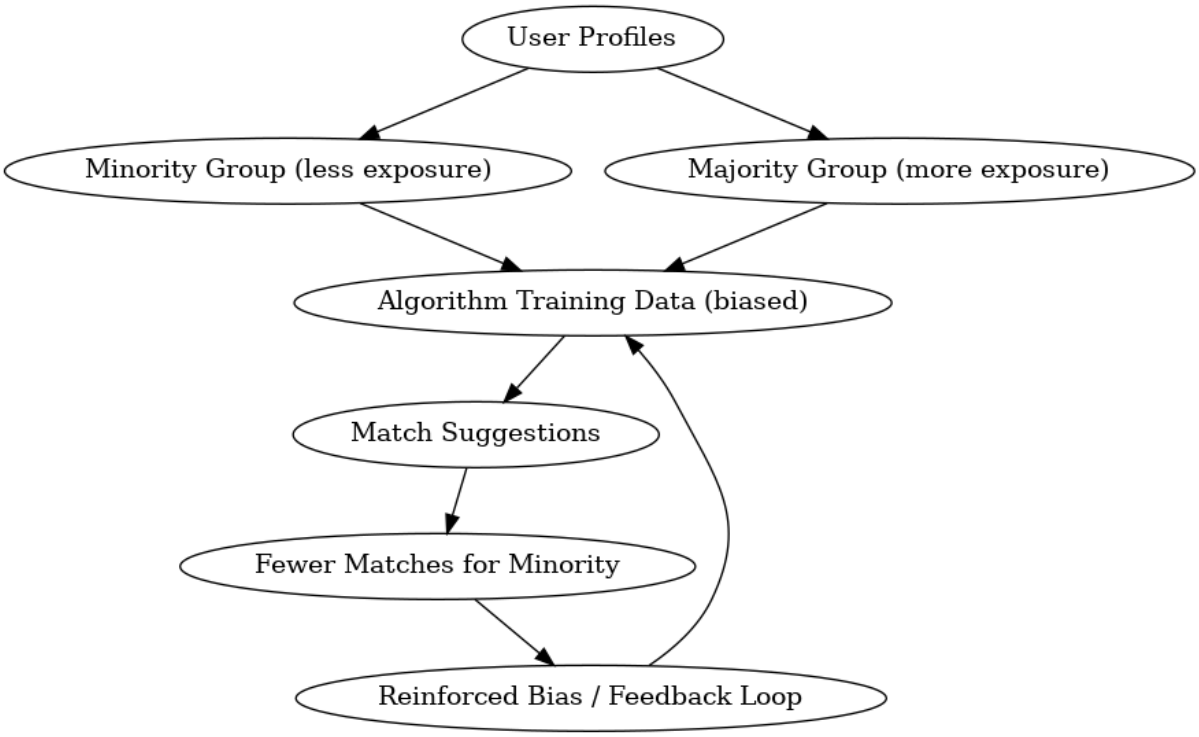

Figure 2 Breeze Reinforced Bias

This situation within the Breeze application is a classic example of class imbalance / population imbalance introducing bias to the algorithm. There are several techniques to solve this; SMOTE, a technique to upscale or downscale the number of instances for a certain class. So in this scenario we could either take a small portion of the Majority Group together with the full population of the Minority group for training our algorithm. Or we could duplicate instances or fabricate new instances for our Minority group to match the number of instances of our Majority group to even it out this way. (Galar, Fernández, Barrenechea, & Sola, 2012)

Another technique we could introduce is using Boosting techniques such as ADA boost and gradient boost which focus on misclassified instances, naturally emphasizing harder-to-classify minority class samples. (Galar, Fernández, Barrenechea, & Sola, 2012)