Problem 3 (Ethics Assignment, 15pts)

Read the article "Amazon Doesnât Consider the Race of Its Customers. Should It?". Please write no more than 1 concise paragraph each in response to the below reflection questions. We do not expect you to do any outside research, though we encourage you to connect to lecture materials and the reading for the module where relevant.

1. Some people think that Amazonâs process for determining which neighborhoods would receive same-day delivery was wrongfully discriminatory, but others disagree. Based on our definitions and discussions from lecture, do you believe that Amazon's same-day delivery process was wrongfully discriminatory? Explain your reasoning.

2. Basing decisions about how to treat others on social group membership often strikes us as being wrongfully discriminatory. For example, most people would say that refusing to hire someone because they are a woman is wrongful discrimination, at least under normal circumstances.

   However, there are some cases in which some people argue that social group membership *should* be taken into consideration when deciding how to treat others. The title of the article poses the question: Do you think that should Amazon consider the race of its customers in its same-day delivery processes? If so, then how?

3. There are many different technical definitions of fairness in machine learning. In this problem, we'll introduce you to the intuition behind two common definitions and invite you to reflect on their limitations.

   Say that Amazon decides to develop a new algorithm to decide its same-day delivery coverage areas. Given your machine learning expertise, Amazon hires you to help them.

   Assume for simplification that Amazon's same-day delivery coverage algorithm f takes as input x, features about an individual Amazon user's account, and outputs binary class label y for whether or not same-day delivery will be offered to user x. User x also has (often unobserved) sensitive attributes a. In this example, we will assume a is a binary label so that $a = 1$ if the user is not white, 0 otherwise.

   One technical notion of algorithmic fairness is called "group fairness"[a]. An algorithm satisfies group fairness with respect to protected attribute a if it assigns the same proportion of positive labels to the group of white and the group of non-white Amazon users. In other words, if 50% of white users have access to same-day shipping, then 50% of non-white users should have access to same-day shipping too.

   What are some limitations or potential issues that may arise with enforcing this definition of fairness in practice? Are there ways that a classifier that satisfies group fairness may still result in discriminatory outcomes?

4. Another technical notion of algorithmic fairness is called "individual fairness"[b]. An algorithm satisfies individual fairness if for all pairs of users x_1 and x_2 that are similar *without taking into consideration their race* a, the algorithm will assign similar probabilities for the attaining the positive label (roughly, $x_1 \sim x_2 \Rightarrow p(x_1) \sim p(x_2)$)[c]. In other words, if two individuals have almost-identical user profiles, then they should both be eligible for same-day shipping, even if one is white and the other is non-white.

   What are some limitations or potential issues that may arise with enforcing this definition of fairness in practice? Are there ways that a classifier that satisfies individual fairness may still result in discriminatory outcomes?

[a] Group fairness is also sometimes referred to as "independence" or "demographic parity". https://fairmlbook.org/classification.html

[b] https://arxiv.org/pdf/1104.3913.pdf

[c] This is an intuitive description of individual fairness (likewise with group fairness above) rather than a precise formalization.

Solution

1. Amazon pledges that it did not intentionally discriminate based on race, it is empirically provable in the article that the results of its same-day delivery process did lead to discriminatory impacts on the basis of a customer's race, therefore, while it's not possible to prove disparate treatment for Amazon's procedure, it is possible to analyze disparate impact. As discussed in the model reading and the lecture materials, disparate treatment requires evidence of intent to discriminate, whereas disparate impact only requires showing different outcomes for members of different protected classes. Typically, this is subject to some guideline such as the four-fifths rule mentioned in the reading. Amazon responded to discrimination concerns with a business necessity argument, saying that it was too expensive to deliver to areas that it left out, which tended to be majority black or hispanic neighborhoods. Under this framework, there are two plausible ways to prove that Amazon's same-day delivery process was wrongfully discriminatory. First, the target variable by definition is segregated by neighborhood, and as the reading points out "due to housing segregation, neighborhood is a good proxy for race and can be used to redline... without reference to race." It can be assumed that this is common knowledge, as the history of housing segregation is very prominent in our culture, so it would be unreasonable to believe that Amazon had no idea that its procedure would discriminate by race through a proxy of zip codes. The results in the article clearly show that Amazon did not take sufficient measures to prevent this racial discrimination, as it surely would not be a significant cost to extend the delivery region to the predominantly black or hispanic neighborhoods that neighbor the existing delivery zone. The Boston Roxbury example is a clear example. This leads to the second main way to prove Amazon's same-day delivery process was wrongfully discriminatory, by showing that there is an achievable less discriminatory alternative, which is identified in the reading as a sufficient criteria to prove disparate impact. Even if it was marginally more costly to deliver to Roxbury or the Bronx neighborhoods, given Amazon's deep pockets, the obvious alternative is that Amazon should have taken a look at their delivery zone and simply extended the zone by hand to cover the predominantly black and hispanic neighborhoods. This is clearly achievable, as Amazon has now already done this in response to the article, but also a verifiably less discriminatory alternative as it would represent more black and hispanic customers in Amazon's neighborhood. Therefore, Amazon's same-day delivery process clearly satisfies the conditions for disparate impact and is therefore wrongfully discriminatory.

2. First, it's important to note that race is a protected class, so it's not legally permissible for Amazon to use race as an explicit feature variable in its model. However, as mentioned previously, the reading shows that the use of zip code data is a proxy for race that contributes to Amazon's model discriminating on the basis of race. Given the response in part 1 that Amazon is wrongfully discriminating, Amazon should consider the racial composition of the neighborhoods that it is including in its same-day delivery process. It could use this data as a way to validate that it is not under-representing protected classes. For example, Atlanta is an extreme example where white residents are more than twice as likely to live in the same-day delivery region than black residents. Black Atlanta residents are significantly less likely to have access to same-day deliver than white Atlanta residents. Amazon could use race in two ways to reduce its discriminatory effect. First, it could use racial composition of zip codes as a means of validating its model. Amazon should choose the model that has the best performance subject to meeting some threshold such that black and white residents will have approximately equal probabilities of living in the same-day delivery zones. A second method makes use of the fact that many of the feature variables Amazon uses are likely correlated to race, making them proxies as discussed in the reading. The reading mentions that it is oftentimes difficult to determine which variables are proxies for race. Amazon could use race data of its customers to evaluate which of its feature variables are most correlated to race (or the greatest proxies for race) and then either omit these variables or regularize them so that they have less impact on the model, reducing the racial discrimination implicit in Amazon's model.

3. One primary issue in achieving this definition of fairness in practice is that it could still result in unfairness at the individual level. The idea of intersectionality plays a key role here. For example, even

if white and non-white users had the same outcomes, it is likely that the non-white users who have positive outcomes are predominantly male and wealthy or of an elite social class. The classifier may find other ways of discriminating against other protected classes, such as gender or social status, in order to achieve the definition of group fairness and still have optimal performance. We can consider a city that is fairly diverse in terms of zip codes, such that there is little racial housing segregation. Even if the classifier has achieved group fairness, it's likely that the model has only selected the wealthiest neighborhoods (suppose these wealthy neighborhoods contain equal compositions of black and white residents), in which case the model would be discriminating against people of a lower social class.

4. A primary limitation of this definition of individual fairness is the possibility of proxy effects inherent in the feature variables for x_1 and x_2. As the reading discusses, even if race is not included as a feature variable, race can still have an impact on the model if race is highly correlated with some of the other feature variables utilized in the model. This implies that even if the model achieves individual fairness, it may not achieve group fairness. For example, in Amazon's model, as discussed in the article, a key feature variable is the amount of Prime Customers living in a Zip Code. The article found that many of black residents in the example cities lived in zip codes that did not have many Amazon Prime customers, leading to those zip codes' exclusion from the same-day delivery service. Even though within a zip code, white and black residents both have equal opportunity for same-day delivery service and thus satisfies the definition individual fairness, we expect that those zip codes have a disproportionately high concentration of white residents and low concentration of black residents. This leads to cases such as Atlanta, where entire zip codes that are composed of predominantly black residents are excluded, which leads to the discriminatory outcomes seen in the article when we consider the entire Atlanta metropolitan area and the notion of group fairness where black residents on average are significantly less likely to receive same-day delivery than white residents.