Review Algorithms as discrimination detectors

1[ga] Guerrero, Abraham, 2[oc] Orozco, Carlos, 3[rd] Sanchez, Diego, 4[mb] Moreno, Braian,
dept.Universidad Nacional de Colombia, Bogotá, Colombia

***Abstract*—In this document, we explain the article named algorithms as discriminations detector by Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan and Cass R. Sustein.**

***Index Terms*—component, formatting, style, styling, insert**

I. Introduction

The use of algorithms for the development of classification models is based on certain classifications defined and organized by analysts who categorize that information. To some extent, we can say that classification algorithms learn from bias. In the United States, the probability for a person is very low; in fact, there are strong studies in economics on the discrimination that people face in areas such as the labor market. One of these studies is "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination," conducted by Marianne Bertrand and Sendhil Mullainathan. In this study, fictitious resumes with names considered 'ethnically white' (such as Emily and Greg) and names considered 'ethnically African American' (such as Lakisha and Jamal) were sent to employers, and significant differences were found in the rates of interview callbacks.

II. Resume

"Discrimination in a Low-Wage Labor Market: A Field Experiment" by Devah Pager is another relevant study. In this research, fictitious job applications with names associated with different ethnic groups were sent to employers, and clear evidence of racial discrimination was found in the responses and hiring rates. In the Colombian context, there is a significant paper on how people discriminate in classification. An example in Bogota is 'Do firms redline workers?' written by Ana Maria Diaz and Luz Magdalena Salas. The results showed that employers statistically discriminate (practice redlining) based on commuting time to work. Specifically, living half an hour away from the vacancy reduces the rate of interview callbacks by 14 percent, holding residence attributes constant. No evidence was found that employers respond to neighborhood effects. This demonstrates that in some cases, labor discrimination is based not on experience but on certain characteristics of the evaluated individual.

This argument is key in the paper: humans discriminate, and this implies that algorithms potentially can discriminate as well, mainly because 'algorithms are built by humans. They are trained on data generated by humans. Humans discriminate, so the algorithms they build can also discriminate.' The paper raises two key points. If the system mirrors how humans classify, the problem of discrimination will be biased. On the other hand, the 'black box' with which humans classify things is to some extent simple, but the designed machines are not. That's why, through an appropriate legal framework, algorithms can detect and prevent discrimination.

According to the paper, our psychological system makes decisions in two ways: some based on strict and advanced reasoning, and others based on a more systematic aspect. Our minds naturally categorize to think, differentiating between internal and external groups. 'Our automatic systems pay close attention to others' age, race, and sex.' In many cases, these biases cannot be visualized. For example, if we focus solely on quantitative data, it's important to consider aspects such as academic qualifications, past work data, among others. However, all these measures in the evaluation do not define gender, such as academic qualifications. As the paper mentions, 'The challenges in using statistical evidence to show intentional discrimination, small sample sizes, unclear objectives, and the general opacity of human cognition combine to create a fog of ambiguity.'

Cases like credit, school admissions, or job opportunities are clear instances where algorithmic use is potentially present. What we currently know about candidates allows us to predict something about them. The training of these algorithms is based on inputs such as grades, income, and the number of reports in a credit bureau. The authors mention that discrimination is an 'ambiguous' issue since some regulations have ambiguous topics. In some countries, it is clear that discrimination based on factors like race, religion, sex, and age is prohibited. In some cases, there is also unequal treatment towards a protected group, and in some instances, there isn't a powerful reason.

It's worth mentioning that the principle of algorithms is to classify and, to some extent, discriminate between two subjects. There are two types of algorithms: classifiers, which have defined parameters to group between categories, and training algorithms, which, based on past experiences, define a technique and predictors for the algorithm. According to the authors, there are three types of bias: bias in the choice of input variables, bias in the choice of outcome measure, and bias in the construction of the training procedure. The paper manages to evaluate the distance of bias by proposing a metric of distance between the existing bias and the bias of a function. This use of algorithms would allow us to answer whether a woman would have been hired if she were a man.

Finally, there are key points like the development of the initial classification. It is important to evaluate it and ensure that it doesn't involve biases. It may happen that in some additional cases, the risk comes from classifications matching managerial decisions. In other cases, decisions made by individuals may have less bias than those made by a machine, as other human sentiments can be less interesting.

REFERENCES

[1] Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. American economic review, 94(4), 991-1013.

[2] Diaz, A. M., & Salas, L. M. (2020). Do firms redline workers?. Regional Science and Urban Economics, 83, 103541.

[3] Pager, D., Bonikowski, B., & Western, B. (2009). Discrimination in a low-wage labor market: A field experiment. American sociological review, 74(5), 777-799.

[4] Kleinberg, J., Ludwig, J., Mullainathan, S., & Sunstein, C. R. (2020). Algorithms as discrimination detectors. Proceedings of the National Academy of Sciences, 117(48), 30096-30100.