Does AI see color?
Caucasian vs non-Caucasian AI Fairness Analysis

Luz Patricia Aguirre-Rodriguez | Sergio R. Morales | Dr. Saeid Tizpaz-Niari

The University of Texas at El Paso, Cybersecurity Research Experience For Educators Through Data Science (CREEDS)

ABSTRACT
The present study utilized the Correctional Offender Management Profiling for Alternative-Sanctions (COMPAS), the risk assessment software, to study and compare the correlation between higher risk scores for non-Caucasian defendants compared to Caucasian defendants. Our study found that at first glance the risk assessment was equitable when grouping all non-Caucasian vs Caucasians.

INTRODUCTION

• Northpointe, a for-profit company has developed (COMPAS) Software (Correctional Offender Management Profiling for Alternative Sanctions). Its purpose lays on collecting data for risk assessment scores in various racial categories.
• The study showed that there was a big discrepancy in recidivism scores when comparing Caucasians against African Americans. It proved to be biased.
• Our focus was to validate whether the bias became more proportion when comparing Caucasians as a group and all other races as another holistic group (Non-Caucasians).
• The following graphic shows that when comparing these two groups, the distribution of Decile scores the discrepancy becomes more balanced but not so in the high-risk recidivist score.

RESULTS AND DISCUSSION

As shown in the graphics above, the numbers represent the percentage of recidivism in female and male populations. Female Caucasians vs. female non-Caucasians have more equivalent percentages across the board. Both populations have about the same low recidivism score as well as medium and high.

- Female Caucasians: 62% - low, 28% medium, 13% - High
- Female Non-Caucasian: 58% - low, 28% medium, 14% - High

In contrast, male Caucasians vs. male non-Caucasians score differently across the board. Male Caucasians start with a high score in the low recidivism category of the graph (meaning that 68% of this population are less likely to reoffend against 48% of their non-Caucasian counterparts.)

As we take a look at the middle part of the spectrum, we see similar numbers (21% Male Caucasian compared to 28% Male non-Caucasians.)

Towards the end of the graph, we find that non-Caucasians have a high recidivism score (24%) compared to male Caucasians (11%) meaning that there is a higher possibility of recidivism in the male non-Caucasian population, which shows a great discrepancy in AI fairness scoring.

- Male Caucasians: 68% - low, 21% medium, 11% - high
- Male Non-Caucasians: 48% - low, 28% medium, 24% - high

POTENTIAL LESSON PLAN ELEMENTS

- Students learn about data base files such as Excel (.csv) and Google Slides, Columns, rows, etc. and practice typing information in cells according to topic.
- Students elaborate excel files to keep track of personal grade data, finances, social events, etc.
- Students learn how to use google collab, pandas.
- Students learn how to create graphs and tables to visualize gathered data with pandas.
- Students explore topics on AI language learning models and fairness.
- Students investigate how AI uses data to predict results and tendencies.
- Students create a school wide data gathering project and report on the results using google colab, pandas and graphics.

REFERENCES


ACKNOWLEDGEMENTS

This work is supported by National Science Foundation Award # 2206182