

Comparison of Scores

- How did my new score compare to the original?

There is an obvious change in the ordering. For instance, Rasheem scored 10 in the original list but ranks second-lowest risk by my new score (7.2). Charlotte stays the highest by the new score (49.8).

- Patterns or outliers?

With the new_decil equation, people with high jail/custody days moved up in ranking. Rasheem is the most significant example, with a very high original decile, but low new score.

- What surprised me?

I did not expect how big the ranking would change though the change of rules, especially seeing how people get moved from bottom to top of the ranking.

Influence of Variables

- Your merge sort algorithm used variables such as a number of total prior offenses, days in jail, etc. How do you think these variables might correlate with broader systemic inequalities?

Prior behaviors, jail days and custody days are not simply a reflection of the history for people. They also reflect policing, charging and bail practices. Those factors differ across communities and locations, so it's very likely that these features carry systemic bias.

- Which variables do you think are the most “objective”, and which may be influenced by bias?

To begin with, they all seemed to be influenced by bias. It seems like that counts of days and number of behaviors are “objective”, but those were also decided outside of that person’s control. For example, people of different races might be treated very differently and therefore affect their records.

Racial Biases in COMPAS

- You may have realized that the original decile scores and new decile scores are tied to race, even though race was not explicitly included in the formula. How is it possible for race to affect the algorithmic outcomes even when it’s not an input variable like P, J, and C?

Because certain parties treat people differently according to their race with or without awareness, this could result in two possible major influences. On one hand, people might face heavier policing that pushes their higher rate of P, J and C to live. On the other hand, the actual decision on their records might differ from races.

- What does this reveal about how data reflects social structures?

Data absolutely cannot reflect everything that exists in our social structures. It is important that people realize this under-represented inequality caused and do not blindly believe in data to be completely fair.

Civic Impact

- What responsibilities do computer scientists and data professionals have when building tools that affect real people's lives?

As people responsible for writing algorithms that process systems, computer scientists should be aware of the potential inequality resulting from either writing the logic of a program, or training algorithms based on any datasets.

- How might algorithmic tools like implementing merge sort be improved to minimize harm for this dataset, if possible at all?

Consult experts on every involved stakeholders and construct fairness tests accordingly. Develop multiple rules of sorting and compare them as we did in this assignment. Report any uncertainty or trends-off made clearly, and have different parties review them.

- How does this assignment influence/change your view of the role of computer science in civic life?

It clearly demonstrates how algorithm bias could be huge on individuals. Therefore, I realize the importance of thinking beyond the algorithm to account for real life scenarios.