

LEGALST 123 Research Proposal

Violet Yao, Zekai Fan

2/22/2019

An article published on ProPublica discussed the software used to predict future criminals and bias against blacks [3]. The article analyzes how statistical risk assessments, which are used in sentencing and bail hearings, are racially biased. An open-source project on GitHub re-analyzed this study and claimed to have found many inconsistencies between the dataset and ProPublica's study [1]. Its basic method is to build a logistic regression model with the score as outcome and race and few other demographic variables as the independent variables. These studies prompt our research project. However, they mainly focus on racial bias of COMPAS. We are interested in how other factors, such as age, gender, marital status, and sentence type, might also result in biased COMPAS scores.

We plan to use the same database used by ProPublica [4]. Aside from racial information, "compas.db" contains 60844 cases, each with DateOfBirth, Sex_Code, MaritalStatus, etc. Since we are not familiar with database operating system, we are learning to use Navicat to read data and may export it as csv files for processing. We found a paper with nice definition of recidivism and factors that might affect recidivist decisions [2].

References

- [1] Annie. Partial Re-Analysis of ProPublica's Analysis, 2017.
- [2] United States Sentencing Commission. Measuring Recidivism: The Criminal History Computation of The Federal Sentencing Guidelines. 2004.
- [3] Angwin et al. Machine Bias. 2016.
- [4] ProPublica. Data and analysis for 'Machine Bias', 2017.