## Introduction

Risk assessment algorithms and biased results are not pleasant combination. In this paper we revisit the famous ‘Machine Bias’ article published by ProPublica in May 2016 about how the COMPAS systems used for pre-trial risk assessment is bias towards a race. Our work questions and corrects rudimentary assumption ProPublica made which turns out to be an error. We also make an effort to build a fair assessment score for the given data only based on criminal activities, eliminating the race or any racial effect in the outcome.

## Literature review

United States has the world’s highest incarceration rate of 655 per 100,000 of population as per the world prison brief data. To control this rate and inflow of the prisoners, states of New York, Wisconsin, California, Florida and few other jurisdictions uses **COMPAS** (developed and owned by Northpointe), an acronym for *Correctional Offender Management Profiling for Alternative Sanctions* which provides certain risk score for an offender to access its likelihood of being a recidivist. These scores are used by US courts to grant the pre-trial release too.

ProPublica, an American non-profit newsroom based out of New York, presented a study in May 2016 and claimed that COMPAS generated ‘Risk Scores’ are biased against African-American race group. For this study they looked at more than 10,000 criminal defendants in Broward County, Florida, and compared their predicted recidivism rates with the rate that actually occurred over a two-year period. To do so, they also collected data about public incarceration from the ‘Florida Department of Corrections’. By joining these two data sets with a defendant’s first name, last name and Date of birth ProPublica assess about 11000 records.

Their analysis was at fault in terms of sampling error and terminologies as they swapped the cause-effect which inflated the result. The article ‘*False Positives, False Negatives, and False Analyses’ : A Rejoinder to “Machine Bias: There’s Software Used Across the Country to Predict Future Criminals. And It’s Biased Against Blacks.”,* stateshow ProPublica has failed to test the bias as per standard and used faulty statistical references and benchmarks to produce the result.

Moreover recently, ProPublica’s COMPAS data was revisited by Matias Barenstein, published on July 08, 2019. He highlights a classic data sampling mistakes made by ProPublica. In which ProPublica included extra 40% (1000) recidivists in the analysis which inflated the results of racial bias by 24%. In order to check for recidivism, ProPublica should have used a cut off on the COMPAS screening date (at 1st April 2014). This would ensure that these people are observed for next two years, as the database contains data through March 2016.

## Data

We studied and recreated ProPublica’s analysis and based on that study, we defined 3 modes of analysis:

* To study the marginal effects of available attributes in data on the decile score (risk score) generated using COMPAS which is used by Courts to tag an offender with ‘low’, ‘medium’, or ‘high’ risk of recidivism
* To build ‘fair score’ from the available data that only concerns to the criminal history and current crime description of an offender and no other social factors, and compare its variation with original ‘decile score’
* To compare the predictive power and error rate of ‘decile score’ and ‘fair decile score’ for recidivism

## Method

ProPublica have a copy of database they worked with on their GitHub repository. We sourced the database copy and extracted various tables which had information collected from Broward county COMPAS system. Obtained the data used by ProPublica for COMPAS analysis. They had acquired this dataset from Broward County, FL. It includes the pretrial defendant’s data from January 2013 to December 2014. They also got the data from Broward county’s Sherriff’s office to observe these defendants for recidivism till April 2016.

Firstly, we replicated the data pre-processing and model built by ProPublica and got the same results as them. Then we used the data and pre-processed it again. We used basic filter same as ProPublica, included the cut-off for COMPAS screening data at 1st April 2014, did feature engineering and excluded some variables from model based on logic. A data dictionary is provided in annexure.

## Our Analysis

## Decile score predictors

The COMPAS systems provides criminals with an assessment upon completion of which the system generates a risk score called ‘Decile Score’ which is a number from 0 (Low) to 10(High) risk of

We studied the individual and combined effects of attributes on decile score.

1. Decile score as a function of crime and demographic factors

*Decile score =*

1. Decile score as a function of crime and demographic factors with sex and race interaction

*Decile score =*

1. Decile score as function of crime related factors

*Decile score = )*

1. Decile score as a function of just Race and sex interaction term

Decile score =

Decile Score Predictors results:

## Building fair\_score using crime related factors

As per the literature review and research done, our hypothesis for an ideal ‘Fair Risk Score’ for a criminal should only concern his/her past and current criminal records and risk score should be independent of socio-economic and demographic factors.

In the above section where we analyzed various combination of the decile score, we particularly chose one which just related to the offender’s criminal data available with us. i.e. *priors count, juv\_fel\_count, juv\_misd\_count, charge\_degree.*  We computed the Beta coefficients for this

Data Acquisition 🡪 Problem and analysis process

1. Excluded “Two-year recidivism” from the model. It is a feature used by ProPublica which is a future variable. It indicates if the arrested person committed Crime in 2 years after COMPAS screening. It will not be available at the time of screening. So, we are not including the same in our risk prediction model.
2. Excluded “Crime factor” from the model and used another variable which defined the severity of the crime more precisely. Crime factor is a feature used by which has two distinct values F(Felony) and Misdemeanour(M). Instead of that, we have used charge degree feature which includes different levels of felony and misdemeanours, like F1, F2, etc.
3. Created a binary variable to find the defendant’s involvement in substance abuse. Used charge description to find if the defendant was charged for any drug possession or use. We have used this feature in prediction of “risk of recidivism” since studies show that people involved in substance abuse tend to re-offend to either fulfil their drug requirements or under the influence of drugs.

**Results**

**Conclusion**

I Introduction:

Importance, goals, overview, literature review.

M Methods

Scope of your methods (test1, 2, )

R Results

Results for each method

If you were able to achieve the goals?

D Discussion

Contribution to existing research

Significance of each result

C Conclusion

Summary of objectives

Major findings

Implications of finding