Faithan To
MSBA 6341 (002) | *HW1: Algorithmic Fairness in Recidivism Prediction*
Thursday, June 12, 2025

Fix Reality, Not the Machine

What is Truth?

Truth is a reflection of our reality. And when we try to create intelligent machines and empower them with truth, no matter how perfect we make them, they can only mirror the flawed society we live in.

Northpointe has created a machine nurtured by truth - an algorithm named COMPAS that predicts whether criminals will go out and commit more crimes after being released onto our streets. These predictions are meant to guide our justice system in an unbiased way in determining bail, parole, sentencing, and rehabilitation programs.

The information Northpointe has empowered COMPAS with in order to make predictions about recidivism is indeed truth. *Facts about each defendant, such as criminal history, age, gender, and race, are objective metrics that cannot be denied.*

However, our friends at ProPublica (or “FauxPublica”, as I like to call them) believe that *you* should change the truth - that this algorithm “unfairly” judges certain racial groups more harshly than others and that *you* should step on the sacred scales of justice in order to achieve “equality of outcome”.

I don’t think any of us want that.

I believe COMPAS fairly accesses the risk level of defendants based on past behavior and any discrepancies between racial groups calls us to action - not in obscuring the truth like what ProPublica suggests, but rather, in creating a more perfect society such that all Americans have an equal opportunity to succeed.

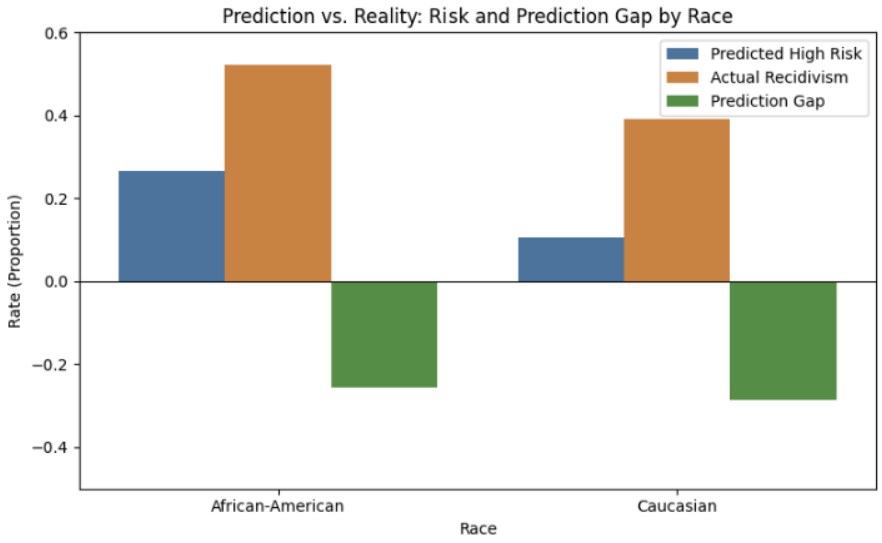
To prove to you COMPAS accurately reflects the truth of our society as it is, I’m going to show you the data. **I analyzed COMPAS’s predictions using 3 different perspectives: scoring disparity, error balance, and predictive parity in order to expose how ridiculous ProPublica’s beliefs are.**

Scoring Disparity Reveals Proportional Discrepancies

We first examine COMPAS’s predictions of risk compared to the actual recidivism rate. African Americans were more likely to be predicted as “High Risk” at 26.6% versus Caucasians at 10.6% - a seemingly concerning 16% discrepancy with statistical significance of p-value < 0.001 (Exhibit 1). However, when we look at the actual recidivism rate of 52.3% for African Americans and 39.1% for Caucasians, the predictions seem much more reasonable (Figure 1).

While COMPAS labels African American defendants as high risk more often, the actual recidivism rate in the dataset is also higher. *The increase in prediction rate is proportional to the increase in actual risk.*

Figure 1



Error Balance Suggests Unfair Treatment

However, ProPublica focused on whether COMPAS's incorrect predictions were distributed evenly across African Americans and Caucasians. *They used recall, which measures how well COMPAS captured reality - that is, whether individuals who actually reoffended (or didn't) were correctly identified.* They found African American defendants were more likely to be misclassified as high risk compared to Caucasian defendants after controlling for similar attributes - both for individuals that actually reoffended and those that did not (Figure 2).

- **Among those who did NOT reoffend,** COMPAS was less accurate in labeling African American defendants as not high risk (0.861) compared to Caucasian defendants (0.952).
- **Among those who DID reoffend,** COMPAS was nearly twice as likely to correctly identify African American defendants as high risk (0.382) compared to Caucasian defendants (0.197).

Figure 2

Recall	**All Defendants**	**African American**	**Caucasian**
True Positive Rate (Reoffended)	0.321	0.382	0.197
True Negative Rate (Did Not Reoffend)	0.903	0.861	0.952

In fact, their regression analysis showed race was a statistically significant attribute in COMPAS's predictions (Exhibit 2, Exhibit 3) and the confusion matrices further illustrate how *COMPAS appears stricter on African American defendants and more lenient on Caucasian defendants (Exhibit 4).*

Predictive Parity Proves Calibrated Algorithm

That being said, the ultimate goal of COMPAS is to accurately predict if a defendant is going to reoffend. *To that end, we must examine whether its predictions are equally accurate across racial groups using precision.* This measures how often COMPAS's predictions were correct such that if it

- **Predicts a defendant to be high risk,** they are equally likely to reoffend and
- **If it predicts low/medium risk,** the groups are (again) equally likely to reoffend.

In Figure 3, although it seems like there are some minor differences in scoring between African American and Caucasian defendants, *these differences are not statistically significant when looking at the regression analysis (Exhibit 5, Exhibit 6), with p-values well above 0.05.*

Figure 3

Precision	All Defendants	African American	Caucasian
Positive Predictive Value (Reoffended)	0.745	0.750	0.726
Negative Predictive Value (Did Not Reoffend)	0.599	0.559	0.649

Therefore, when looking at predictive values by race, we can conclude a fairly similar correct prediction rate for both those labeled high risk (who reoffended) and those labeled low risk (who did not reoffend) between African American and Caucasian defendants. This suggests COMPAS satisfies predictive parity - its risk labels are equally valid across groups.

If ProPublica had their way, they would rig the algorithm so COMPAS would be less biased against African Americans, but as a result have different scoring methods for different races - codifying discriminatory judicial treatment.

We Should Change Reality, Not Hide It

Ultimately, Northpointe has created a machine with biases in how it predicts defendants' recidivism risk scores, but these biases largely reflect our current reality where certain groups end up being more likely to reoffend - a metric COMPAS uses to ensure its scoring is consistent. *This metric is an injustice, not because of its usage in the COMPAS model, but because its predictive ability is so strong in the first place.*

We live in a flawed reality and the truths we feed into our machines such as COMPAS reflect that. Our reality is filled with very uneven access to resources such as school funding, childcare, healthcare, and even intangible things such as mental health support and supportive communities - these inequalities are often strongly correlated with crime and race. *ProPublica wants to "fix" these inequalities post facto, but initiatives like these are lazy and akin to slapping a bandaid on the gaping wound of race-based disenfranchisement.*

Instead we should strive to ensure all Americans across different demographics and socioeconomic statuses have an equal opportunity to succeed from the start so that one day, ***our machines will reveal our progress and a future regression analysis will show COMPAS without these biases.***

Exhibits

Exhibit 1

Model (Base)	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-2.1803	0.118	-18.413	0	-2.412	-1.948
race[T.Caucasian]	-0.5792	0.09	-6.457	0	-0.755	-0.403
age_cat[T.Greater than 45]	-1.6574	0.156	-10.642	0	-1.963	-1.352
age_cat[T.Less than 25]	1.1628	0.091	12.747	0	0.984	1.342
c_charge_degree[T.M]	-0.3943	0.09	-4.396	0	-0.57	-0.218
sex[T.Male]	0.1451	0.107	1.36	0.174	-0.064	0.354
priors_count	0.2046	0.009	23.01	0	0.187	0.222

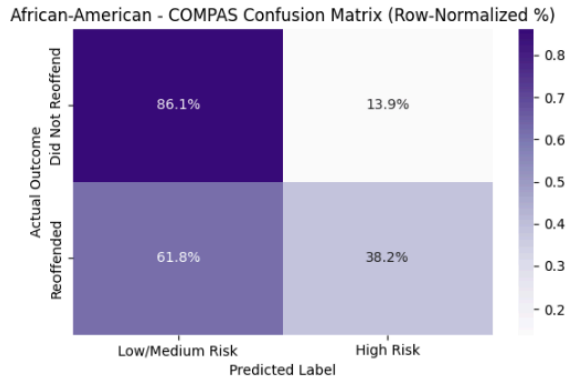
Exhibit 2

Model (False Positives)	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-2.6276	0.189	-13.906	0	-2.998	-2.257
race[T.Caucasian]	-0.7359	0.159	-4.641	0	-1.047	-0.425
age_cat[T.Greater than 45]	-1.4935	0.258	-5.781	0	-2	-0.987
age_cat[T.Less than 25]	1.1654	0.159	7.336	0	0.854	1.477
c_charge_degree[T.M]	-0.2173	0.151	-1.437	0.151	-0.514	0.079
sex[T.Male]	0.0866	0.172	0.504	0.614	-0.25	0.423
priors_count	0.1977	0.016	12.05	0	0.166	0.23

Exhibit 3

Model (False Negatives)	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-1.6149	0.157	-10.268	0	-1.923	-1.307
race[T.Caucasian]	-0.4737	0.111	-4.249	0	-0.692	-0.255
age_cat[T.Greater than 45]	-1.5302	0.194	-7.891	0	-1.91	-1.15
age_cat[T.Less than 25]	0.9982	0.115	8.674	0	0.773	1.224
c_charge_degree[T.M]	-0.443	0.113	-3.921	0	-0.664	-0.222
sex[T.Male]	0.0604	0.141	0.43	0.667	-0.215	0.336
priors_count	0.176	0.011	16.295	0	0.155	0.197

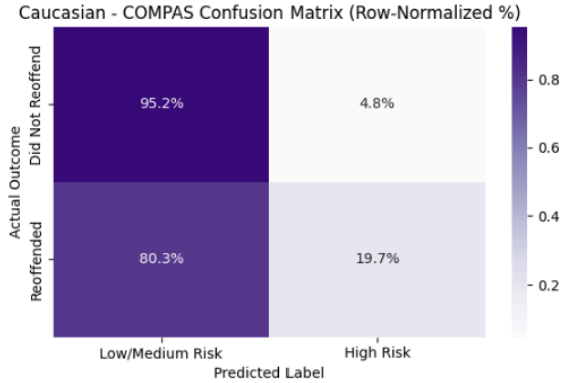
Exhibit 4




```
=== Classification Report ===
                precision    recall  f1-score   support

Did Not Reoffend       0.559     0.861     0.678      1514
      Reoffended       0.750     0.382     0.506      1661

        accuracy                          0.610      3175
       macro avg       0.655     0.621     0.592      3175
    weighted avg       0.659     0.610     0.588      3175
```




```
=== Classification Report ===
                precision    recall  f1-score   support

Did Not Reoffend       0.649     0.952     0.772      1281
      Reoffended       0.726     0.197     0.310       822

        accuracy                          0.657      2103
       macro avg       0.688     0.575     0.541      2103
    weighted avg       0.679     0.657     0.591      2103
```

Exhibit 5

Model (Positive Predictive Value)	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	0.0181	0.226	0.08	0.936	-0.425	0.462
race[T.Caucasian]	0.1254	0.179	0.703	0.482	-0.224	0.475
age_cat[T.Greater than 45]	-0.5702	0.28	-2.039	0.041	-1.118	-0.022
age_cat[T.Less than 25]	0.4814	0.172	2.799	0.005	0.144	0.819
c_charge_degree[T.M]	-0.3777	0.166	-2.273	0.023	-0.703	-0.052
sex[T.Male]	0.4603	0.196	2.343	0.019	0.075	0.845
priors_count	0.1003	0.015	6.612	0	0.071	0.13

Exhibit 6

Model (Negative Predictive Value)	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-0.9764	0.094	-10.371	0	-1.161	-0.792
race[T.Caucasian]	-0.07	0.069	-1.019	0.308	-0.205	0.065
age_cat[T.Greater than 45]	-0.62	0.087	-7.122	0	-0.791	-0.449
age_cat[T.Less than 25]	0.625	0.086	7.274	0	0.457	0.793
c_charge_degree[T.M]	-0.1139	0.069	-1.638	0.101	-0.25	0.022
sex[T.Male]	0.3209	0.084	3.824	0	0.156	0.485
priors_count	0.1561	0.011	14.496	0	0.135	0.177