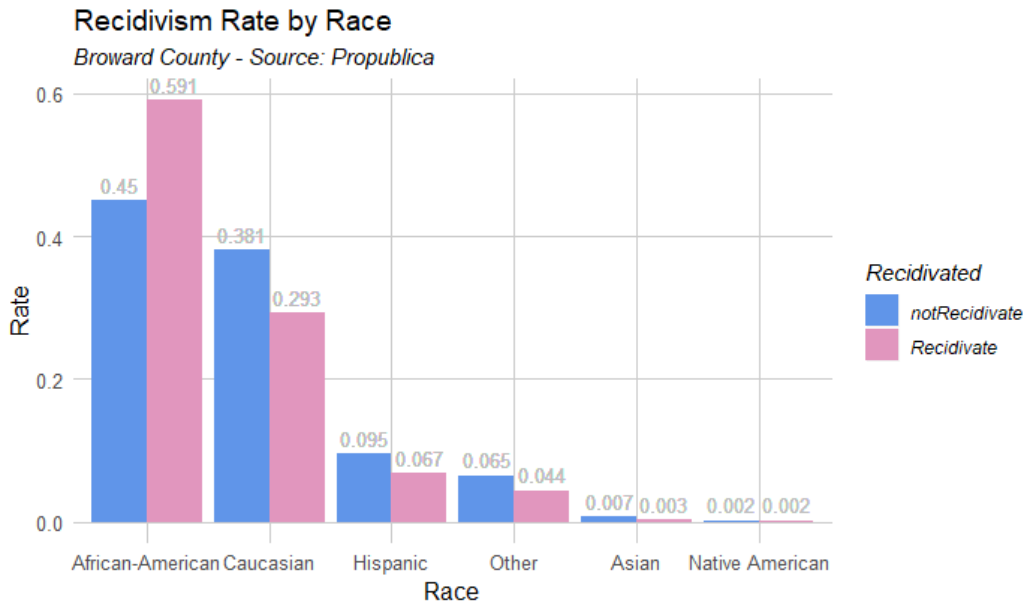
Predicting Recidivism Evaluation
Hanyong Xu
Oct. 31st, 2019

Context

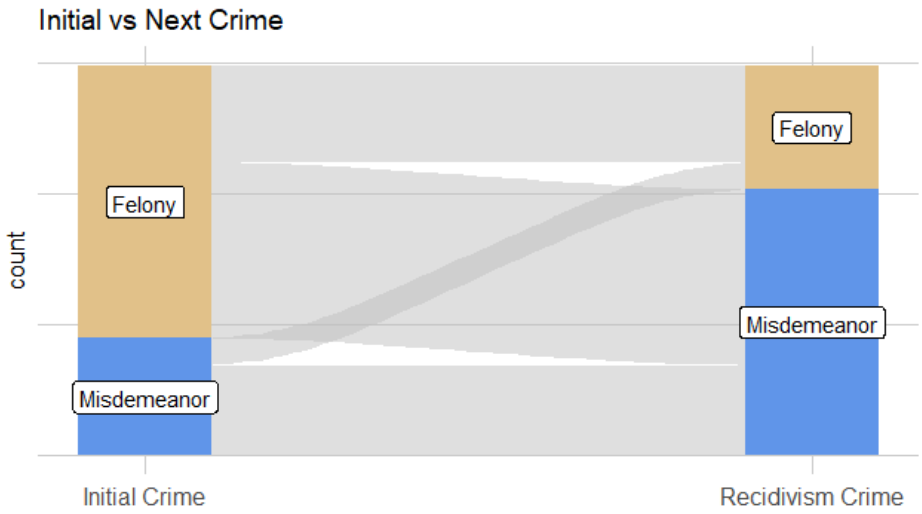
In 2016, ProPublica analyzed the COMPAS algorithm for predicting recidivism and found out its underlying unfairness due to disparate error rates among different racial groups.[1] A similar algorithm that the city might adopt for evaluating ex-offenders to job training programs is discussed in this memo. The recommendation is to not use the model since its results produce varying error rates for different demographics, and it cannot cope with the selection bias due to reasons like discriminatory over-policing.

Evaluation

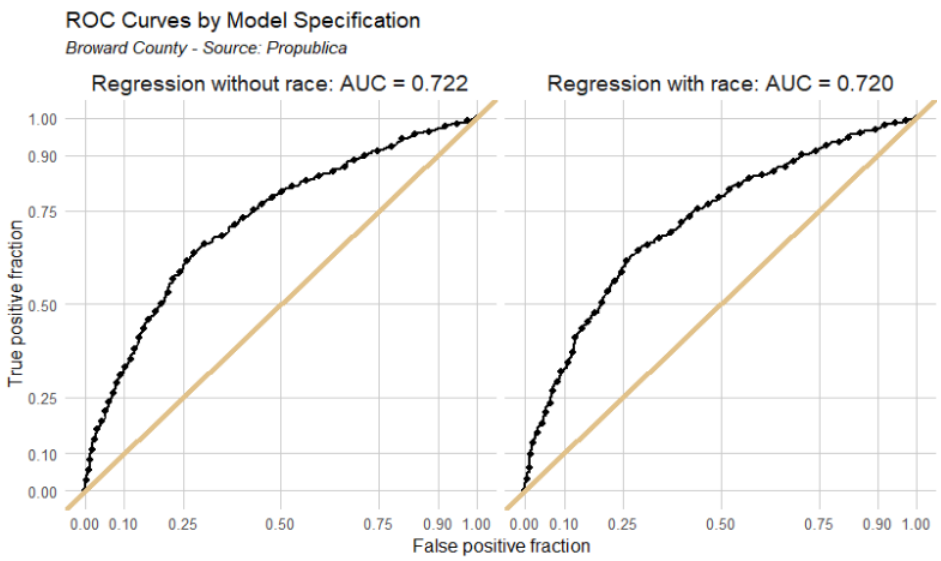
An exploratory analysis is conducted to understand the trend of the dataset. The dataset used is the same as the one used by ProPublica, consisting 6,163 inmates in Broward County. By plotting recidivism rate by race, one can clearly see that African-Americans have much higher recidivism rate (59.1%) than any other racial groups. The second-high recidivism rate is from Caucasians (29.3%). The recidivism rate across groups is disproportionate to the overall racial population, which is potentially due to two reasons, one is that African-Americans are more likely to recidivate, and two is that they are more likely to be arrested due to discrimination in policing. A Sankey diagram is also plotted, which reflects that a larger proportion of the recidivists tend to conduct misdemeanor than felonies.



[1] www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing



Logistic regression is used for prediction, and its accuracy with and without the racial factor is analyzed by comparing the AUCs of the respective ROC curves. Without going into technical details, the Area Under the ROC Curve, or AUC, captures the accuracy of the logistic model. A larger AUC is better, as long as it's not too close to 1 to avoid overfitting. The ROCs of regressions with and without the racial factor are plotted. Including the racial factor in modeling, though controversial, might be able to adjust the underlying bias due to race difference. Thus, one should reconsider the rule of forbidding racial factor in criminal justice predictive modeling. Here, the overall AUC's with and without race are 0.720 and 0.722. This means the model without racial factor is slightly more accurate, but overall, the outputs are similar. Therefore, race does not have a strong predictive power in an overall sense. A similar check could be done for any new algorithm to see if race can make the model fit better. A more in-depth check of whether race predicts accurately across groups are done in the next section. The accuracy is also examined by looking at the AUCs of the three major racial groups, Caucasians, African Americans, and Hispanics. Within each model, the AUCs show that the models predict the recidivism of Hispanics the best, African American the second, and Caucasians the third. Both models are almost equally accurate across racial groups.



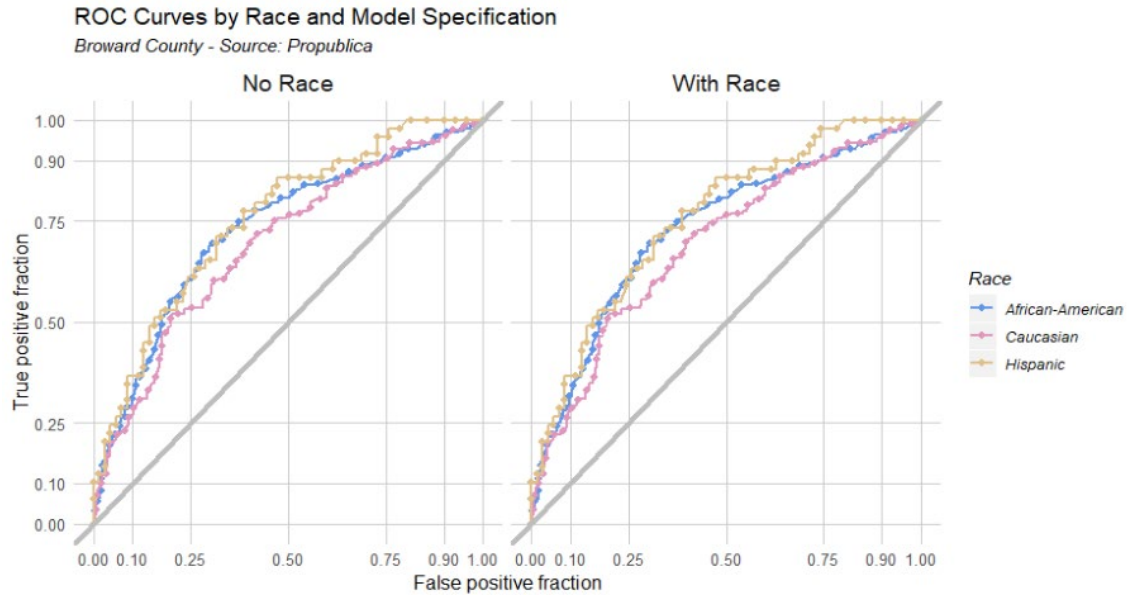


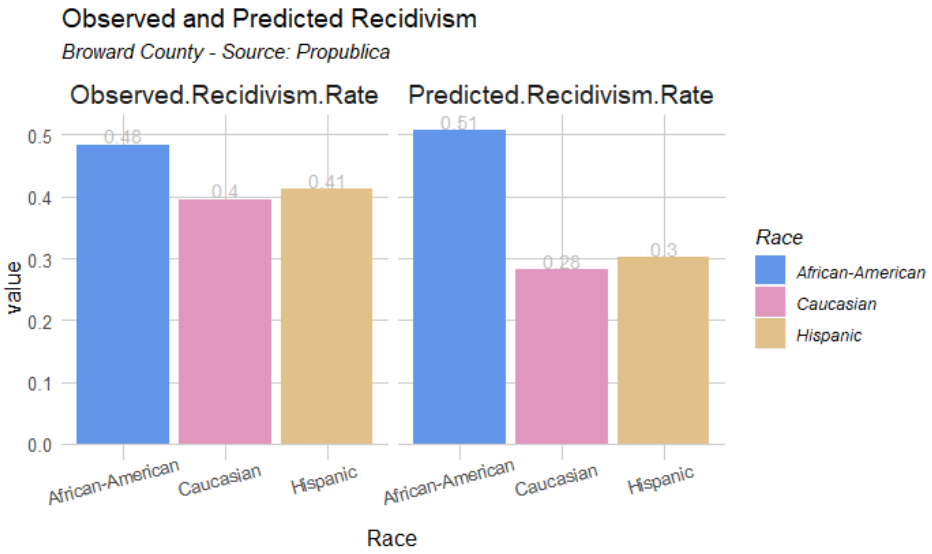
Table 1: AUC by Race and Model Specification

regression	Race	AUC
No Race	African-American	0.7286138
With Race	African-American	0.7285623
No Race	Caucasian	0.6928050
With Race	Caucasian	0.6925800
No Race	Hispanic	0.7549563
With Race	Hispanic	0.7552478

Accuracy is not the only measure to evaluate a model, the generalizability across different racial groups is also essential for social just, therefore, the observed and predicted recidivism rates by race are examined. The recidivism rate for Caucasians and Hispanics are underpredicted since their predicted rate is much lower than their observed rate.

Table 2: Observed vs Predicted Recidivism Rate by Race and Model Specification

Race	regression	Observed.Recidivism.Rate	Predicted.Recidivism.Rate
African-American	No Race	0.4841572	0.5082383
African-American	With Race	0.4841572	0.5259823
Caucasian	No Race	0.3958333	0.2821970
Caucasian	With Race	0.3958333	0.2784091
Hispanic	No Race	0.4117647	0.3025210
Hispanic	With Race	0.4117647	0.2773109



The confusion matrix displays the error rates for each group. Here, a true positive is an ex-offender predicted to recidivate actually recidivated; a true negative is an ex-offender predicted not to recidivate and actually did not; a false positive is an ex-offender predicted to recidivate and actually did not; a false negative is an ex-offender predicted to not recidivate and actually recidivated. The bar chart of this confusion matrix shows that though accuracy is similar for each group, the algorithm correctly predicts African-American recidivists much accurately than the other two groups, while it predicts African-Americans who do not recidivate with much lower rate. In other words, the algorithm tends to predict African-Americans to be more likely to recidivate than the other two groups. The tradeoff between accuracy and generalizability is that higher accuracy demands higher true negative and true positive rates, which are highly biased towards race. If one wants to ensure a very low recidivism rate for the

people having access to the job training program, the algorithm would give Caucasians and Hispanics more chance than African-Americans. The reason could be that African-Americans are targeted more by the police; therefore, they have higher chances of being rearrested and thus recidivate. Unless this underlying bias is eliminated from the cause, the model will always produce biased results. Thus, I do not recommend the adoption of the model.

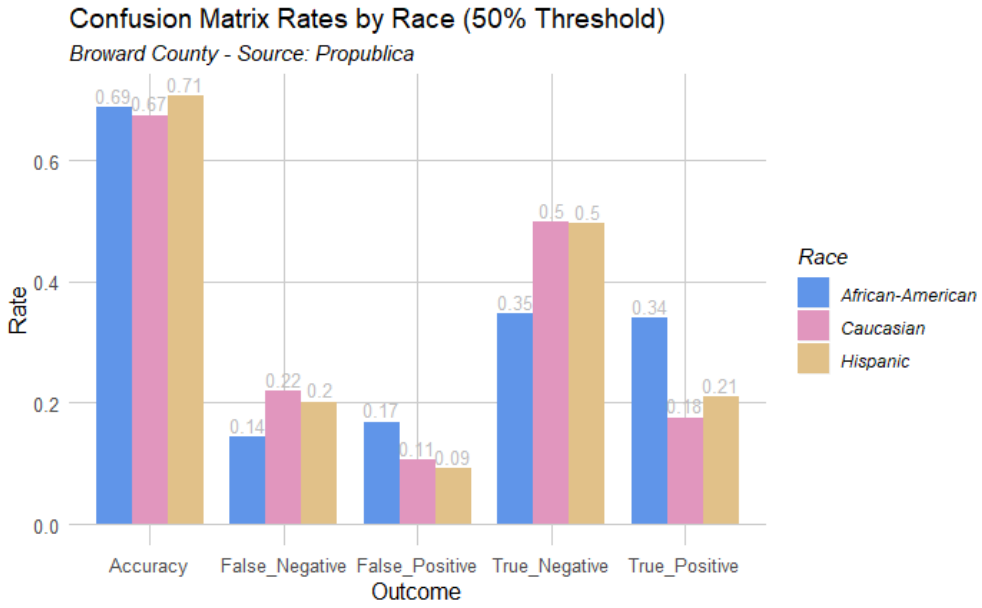


Table 3: Confusion Matrix

Race	True_Positive	True_Negative	False_Negative	False_Positive	Accuracy
African-American	0.3396705	0.3472750	0.1444867	0.1685678	0.6869455
Caucasian	0.1761364	0.4981061	0.2196970	0.1060606	0.6742424
Hispanic	0.2100840	0.4957983	0.2016807	0.0924370	0.7058824