The Accuracy of Random Forest Analysis on Predicting Recidivism

Emma Gaydos, Nina Baker, Wagih Henawi

Abstract

Various sectors of the criminal justice system in the United States use algorithms to predict a defendant's chance of re-offending when determining the defendant's eligibility for parole, the length of their sentence, bail, and other important factors. However, a 2016 study by ProPublica demonstrated that one such algorithm, COMPAS, is often inaccurate in a way that causes racially disparate outcomes. To substantiate ProPublica's analysis, we ran a random forest analysis on recidivism data from the Broward County Clerk's office and created confusion matrices to examine the accuracy and false positive and negative rates for Black and white defendants. We found that, despite similar rates of overall accuracy, Black defendants had a 61.5% higher chance of being falsely predicted to re-offend. Additionally, a white defendant had a 52.7% higher chance of being incorrectly predicted to not reoffend compared to Black defendants. This is in line with ProPublica's research. We conclude with recommendations about the use of COMPAS in the criminal justice system.

Introduction

The United States criminal justice system has repeatedly come under fire for anti-Black prejudice within policing, sentencing, pre-trial practices, parole, and post-prison consequences (The Sentencing Project, 2018). Because of this broader context, any procedures used within the criminal justice framework must demonstrate consideration for racial inequity. There are concerns that one such procedure, called the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS), has not effectively done so (Angwin et al., 2016).

COMPAS is an artificial intelligence, or AI, intended to predict whether or not a given offender will recidivate based on a variety of categorical variables. These variables have not been made available by COMPAS's creator -- Northpointe -- but include factors such as criminal history and employment, among others. Race is not directly included in their analysis.

In a 2016 study, researchers at ProPublica concluded that COMPAS's algorithm was laden with racial bias (Angwin et al.). The false positive rate -- that is, the rate at which people were predicted as high risk for recidivism and did not recidivate -- was much higher for Black defendants (44.9%) than white defendants (23.5%) in the ProPublica study. In contrast, the false negative rate -- the rate at which people were predicted as low risk for recidivism and did recidivate -- was significantly higher for white defendants (47.7%) than Black defendants (28.0%). This is especially concerning given that COMPAS has influenced decisions regarding sentencing, pre-trial practices, parole, and bond in at least 10 different states (Angwin et al., 2016).

If racial bias in fact impacts COMPAS's risk assessment, then COMPAS should not be used. However, Northpointe has denied the claims of the ProPublica study, arguing that the algorithm predicts the recidivism of Black defendants and white defendants equally well (Dieterich et al., 2016, pp. 24-26). As evidence, they show two similar AUCs for white defendants (0.693) and for Black defendants (0.692). As a result, we decided to evaluate the accuracy of ProPublica's conclusions by conducting our own analysis on ProPublica's dataset.

The aim of this paper is to evaluate tools like COMPAS by creating our own recidivism prediction model. We will utilize a random forest to predict recidivism rates for all defendants before assessing the model for the accuracy, false positive rate, and false negative rate for Black and white defendants. In line with the arguments outlined above, we hypothesize that our model will show similar overall accuracy for Black and white defendants. However, we also expect, in line with ProPublica's work, that the model fails differently for different races. We predict that Black defendants will have a greater false positive rate than white defendants and white defendants will have a greater false negative rate than Black defendants.

Methodology

Our data originates from the Broward County Clerk's Office "COMPAS Recidivism Risk Score Data and Analysis" and is obtained from the ProPublica Datastore. We cleaned our data by removing variable columns which held information about the defendant which were duplicated in other columns and by removing columns which held no predictive information, such as the case number of the defendant. Our final dataset consisted of 27 explanatory variables. Descriptions and summary statistics of each variable can be found in Appendix A. After removing cases with missing variable entries, we were left with a total sample size of 7819 defendants.

To begin our analysis, we randomly partitioned the dataset in half to create a training dataset and a testing dataset where the sample size was 3910 and 3909, respectively. Using our

training dataset, we constructed a random forest consisting of 1000 classification trees. Each tree is constructed using a bootstrapped dataset where 5 variables are randomly selected at each level as candidates for that node, and the one with the highest predictive power towards recidivism takes that node position.

Utilizing our random forest model, we predict the probability of recidivation for each case in our training dataset and create a receiver operating characteristic (ROC) curve matching our results. This ROC curve determines the optimal threshold value that will maximize the accuracy of our model. At that threshold value, all probabilities higher than that value are determined to be true. We then predicted recidivism using the specified threshold on our testing dataset, and analyzed differences in predictive power based on a defendant's race.

Results

Our ROC curve (Figure 1) for our training dataset computed a recommended threshold of 0.281, meaning that every defendant with a predicted probability of recidivating at 28.1% or higher were determined to be at high risk for recidivating.

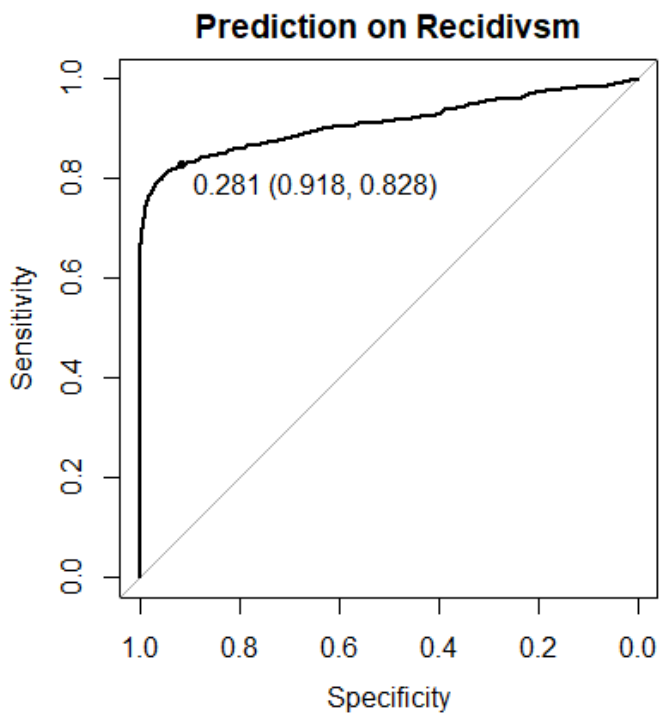

Figure 1: ROC on our training dataset

Using the recommended threshold of 0.281, we ran our testing dataset through the model to create a confusion matrix (Table 1) to display the number of defendants who did and did not re-offend, and whether the model predicted them as high-risk or low-risk re-offenders. The probability of a defendant who re-offended accurately being predicted to have a high risk for recidivism was 65.8%, and the probability of a defendant who did not re-offend accurately being predicted to have a low risk for recidivism was 64.3%. The overall accuracy of the model was 64.8%.

	Predicted low risk for recidivism	**Predicted high risk recidivism**
Did not re-offend	1648	915
Did re-offend	460	886

Table 1: Confusion matrix comparing the number of defendants in our testing data who did and did not re-offend and whether they were predicted as a high risk or low risk recidivator.

We bisected the testing data into Black defendants and white defendants to test the sensitivity and specificity of our model by race. Using the recommended threshold value of 0.281 for white defendants, the model created a 30.2% false positive rate for white defendants compared to a 49.1% false positive rate of Black defendants. To put this in context, a Black defendant had a 61.5% higher chance of being falsely classified as having a high risk of recidivism compared to a white defendant. On the other hand, white defendants had a 46.9% false negative rate compared to a 24.7% false negative rate for Black defendants. As a result, a white defendant had a 52.7% higher chance of being misclassified as having a low-risk for recidivism compared to Black defendants.

Discussion

Our results align with ProPublica's study. Despite the model's overall accuracy rate for Black defendants and white defendants being similar (62.7% and 64.9%, respectively), the types of inaccuracies presented in the model were different. False positives were much more common for Black defendants than white defendants while false negatives were much more common for white defendants than Black defendants. If we were to implement our model in a real-world setting, this might indicate fewer parole opportunities and longer sentences (among other consequences) for Black defendants than white defendants. This model would harm Black defendants and benefit white defendants, thereby introducing additional racial bias to the system.

Moreover, it is worth questioning whether artificial intelligence or statistical modeling should play any part in the criminal justice system. There is evidence that the criminal justice system is already biased against Black people (The Sentencing Project, 2018). That being the case, traits that are associated with Black people are likely associated with crime and recidivism. Any model that builds on such variables will inevitably double down on already-pervasive systemic racism. In conclusion, we agree with ProPublica's recommendation that the COMPAS system be either completely removed or temporarily halted. If halted, further research into recidivism must be done prior to reinstating the program. Any program taking the place of COMPAS should include no racial bias or, at the least, much less than a judge or jury.

Our study has some notable limitations. Our data comes entirely from one county, so there may be fundamental differences between our data and the US population at large. Also, due to Northpointe's lack of transparency, we did not have access to COMPAS to test it directly. Our model was instead constructed of variables that were likely to be components of COMPAS, so it is possible that there are discrepancies between the two programs. However, we believe that pairing our results with those of the ProPublica researchers strengthens the conclusions of each study.

References

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). *Machine bias*. ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 2). *Compas recidivism risk score data and analysis*. ProPublica Data Store. https://www.propublica.org/datastore/dataset/compas-recidivism-risk-score-data-and-analysis

Brennan, T., Dieterich, B., & Ehret, B. (2007). Research synthesis: Reliability and validity of COMPAS. Northpointe.

Dieterich, W., Mendoza, C., & Brennan, T. (2016). COMPAS risk scales: Demonstrating accuracy equity and predictive parity. Northpointe.

Larson, J., Mattu, S., & Kirchner, L., Angwin, J. (2016). *How we analyzed the COMPAS recidivism algorithm*. ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

The Sentencing Project (2018, April 19). *Report to the United Nations on racial disparities in the U.S. criminal justice system.* The Sentencing Project. https://www.sentencingproject.org/publications/un-report-on-racial-disparities/

Appendix A

Explanatory Variable:

Recidivism - Whether or not the case recidivated (True if they did)

True	False
2692	5127

Response Variables:

Race - The race of the case.

African-American	Caucasian	Hispanic	Other
3853	2771	677	518

Age Group - The age group the case belongs to

18-25	26-35	36-45	46-55	55+
2019	2772	1437	1030	561

Sex - The sex of the case

Male	Female
6225	1594

Marital Status - The marital status of the case

Single	Married	Significant Other	Divorced	Seperated	Unknown	Widowed
6020	935	260	337	179	50	38

Charges - The number of charges against the case.

0-5	6-9	10-19	20+
2877	1878	1659	1405

Escape - Whether or not the case attempted escape (True if they did)

True	False
322	7497

Physical - Whether or not the cases charges included physical assault (True if it did)

True	False
3772	4047

Parent - Whether or not the case had at least one child (True if they did)

True	False
220	7599

Drugs - Whether or not the cases charges included drugs (True if they did)

True	False
3289	4530

Marijuana - Whether or not the cases charges included marijuana (True if it did)

True	False
2158	5661

Alcohol - Whether or not the cases charges included Alcohol (True if it did)

True	False
1235	6584

Juvenile Felony Count - Number of felonies committed as a juvenile

Min	1st Quartile	Median	Mean	3rd Quartile	Max
0	0	0	0.067	0	20

Juvenile Misdemeanor Count - Number of misdemeanors committed as a juvenile

Min	1st Quartile	Median	Mean	3rd Quartile	Max
0	0	0	0.082	0	13

Priors - The number of prior offences

Min	1st Quartile	Median	Mean	3rd Quartile	Max
0	0	1	3.3	4	43

Weapon - The number of charges including a weapon

Min	1st Quartile	Median	Mean	3rd Quartile	Max
0	0	0	0.3	0	13

Burglary - The number of charges including burglary

Min	1st Quartile	Median	Mean	3rd Quartile	Max
0	0	0	0.4	0	28

Disrupt - The number of charges including disruption

Min	1st Quartile	Median	Mean	3rd Quartile	Max
0	0	0	0.8	0	17

Mischief - The number of charges including mischief

Min	1st Quartile	Median	Mean	3rd Quartile	Max
0	0	0	0.09	0	12

Obstruction - The number of charges including obstruction

Min	1st Quartile	Median	Mean	3rd Quartile	Max
0	0	0	0.36	0	14

Loitering - The number of charges including loitering

Min	1st Quartile	Median	Mean	3rd Quartile	Max
0	0	0	0.08	0	30

Accessory - The number of charges including accessory to a crime

Min	1st Quartile	Median	Mean	3rd Quartile	Max
0	0	0	0.05	0	6

Cocaine - The number of charges including cocaine

Min	1st Quartile	Median	Mean	3rd Quartile	Max
0	0	0	0.5	0	25

Heroin- The number of charges including heroin

Min	1st Quartile	Median	Mean	3rd Quartile	Max
0	0	0	0.013	0	5

Meth - The number of charges including meth

Min	1st Quartile	Median	Mean	3rd Quartile	Max
0	0	0	0.04	0	9

Pharmacy - The number of charges including pharmaceuticals

Min	1st Quartile	Median	Mean	3rd Quartile	Max
0	0	0	0.12	0	25

Stalking - The number of charges including Stalking

Min	1st Quartile	Median	Mean	3rd Quartile	Max
0	0	0	0.01	0	4

Tobacco - The number of charges including tobacco

Min	1st Quartile	Median	Mean	3rd Quartile	Max
0	0	0	0.016	0	7