Machine Learning in the Courtroom: Evaluating the Effectiveness of Algorithms to Predict the Likelihood of Recidivism

Bowen Mince and Wagih Henawi
Department of Mathematics and Statistics, Grinnell College

Abstract

The United States has both the largest prison population and the highest per-capita incarceration rate in the world. These metrics have increased in the past half-century, even outpacing population growth. To reduce the prison population in an unbiased manner, several algorithms have been created and implemented to predict the likelihood of recidivism of convicted people. In this work, we develop models using two types of algorithms (classification trees and random forests) and then evaluate their accuracy in predicting recidivism using a data set from Broward County, Florida that tracked 10,372 criminals convicted in 2013 and 2014 over a 5 year period. The accuracy for each model ranges from 60% - 70%. However, the false-positive rate and false-negative rate varied significantly in each model with regard to race. We then discuss the shortcomings of the current use of these algorithms in courtrooms throughout the United States and how they perpetuate systemic bias.

Background

In 2016, Pro-Publica published an article claiming the popular risk assessment tool, COMPAS, was limited in its accuracy to predict recidivism. Moreover, they found that the algorithm was racially biased, where African Americans were twice as likely to be predicted to recidivate compared to Whites. Similarly, Whites that did recidivate were twice as likely to be given a lower risk score compared to African Americans. In making its claims, the newsroom looked at COMPAS scores given to individuals in Broward County, Florida.

Methods

We developed our own models to see if we can predict recidivism and whether bias would still exist in the model. We developed two models using classification trees and random forests. The data we used comes from Broward County, Florida. It contains 10,428 individuals arrested during the years 2013-2014. The variable of interest, recidivism, tells us whether or not the individual recidivated within two years of receiving their COMPAS risk assessment.

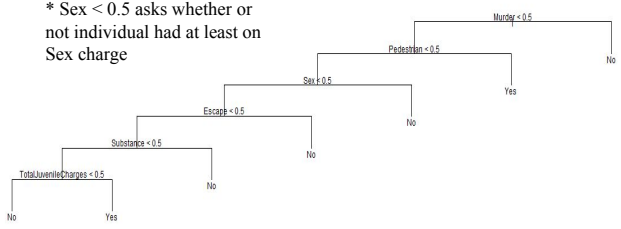
Steps Taken*

- Data cleaning.
- Separated the data into Training and Testing datasets.
- Developed the two models using the Training dataset.
- Evaluated the overall accuracy (and accuracy across racial groups) of each model.

*RMD file detailing these steps available upon request

Figure 1: Tree Diagram representing our Classification Tree Model

* Sex < 0.5 asks whether or not individual had at least on Sex charge



Acknowledgements

We would like to thank Dr. Shonda Kuiper for her guidance and mentorship. Please email mincebow@grinnell.edu for a comprehensive view of our results.

Results

Type	Observed	Predicted	Tree	Random Forest
True Negative	No	No	1885	1735
False Negative	Yes	No	879	656
False Positive	No	Yes	153	303
True Positive	Yes	Yes	204	427

Table 1: Confusion Matrix for Classification Tree and Random Forest on Testing Data. The accuracy for the Tree and Random Forest is (1885 +209)/3122 = 67% and (1735+427)/3122 = 69% respectively.

In terms of the training dataset, the random forest greatly outperformed the classification tree (97% accuracy to 67%). When considering the testing dataset, the random forest model slightly outperformed the classification tree (69% to 67%). Both models discriminate on the basis of race when evaluated on the testing dataset. Figure 2 shows that African Americans have a false positive rate of 20% compared to Whites who had a false positive rate of about 12%.

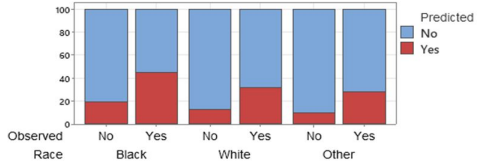


Figure 2: Evaluation of Random Forest on Testing Dataset Grouped by Race

Discussion & Future Work

Our algorithms show that it's very difficult to predict recidivism. Even when we think our model has high accuracy with the training dataset, it is imperative to test it with the testing dataset. Our findings confirm that bias with regard to race can inherently enter the model even when race is not in the model. Future work would include discerning which variables are correlated with race and lead to a racially biased model.

References

- Julia Angwin, Jeff Larson. "Machine Bias." *ProPublica*, ProPublica, 23 May 2016, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.
- Yong, Ed. "A Popular Algorithm Is No Better at Predicting Crimes than Random People." *The Atlantic*, Atlantic Media Company, 29 Jan. 2018, https://www.theatlantic.com/technology/archive/2018/01/equivant-compas-algorithm/550646/.