

# PREDICTING RECIDIVISM FOR DEFENDANTS AND ITS CRITICISMS OF BIAS

Reed Ballesteros

Northwestern University, Introduction to Data Science

December 5, 2021

There has been a need for numerical models in the criminal justice system for almost one hundred years, with the first models to predict a successful parolee dating back to 1927 (Christian 2020, 51). Manuals from 1951 described punch-card-based systems that were used to determine parole at that time (Ohlin 1951, 63). Adoption of these systems was slow. Personal computing in the 1980s and 1990s brought about more sophisticated models which were used in various jurisdictions in the United States.

One such model was the development of the Correctional Offender Management Profiling for Alternate Sanctions, or COMPAS, in 1998 by Dave Wells and Tim Brennan through their company Northpointe. The software uses a linear model to determine recidivism, the likelihood of re-committing a crime, in a defendant. The system calculates three types of risk scales for an individual: pretrial release risk, general recidivism risk and violent recidivism risk. COMPAS was initially adopted as a pilot program by the state of New York in 2001 and was rolled out to the rest of the state in 2010, except for New York City (Angwin et al. 2016).

As COMPAS and other risk assessment tools are integrated into more criminal justice systems, the issue of potential biases in them has been a growing concern. Widespread awareness of the issue was led by a ProPublica article which describes how the COMPAS algorithm can be racially biased against African Americans (Angwin et al. 2016). Angwin and her team found a disparity which favored white defendants over African American defendants in both false-positive (labeled 'Higher Risk' for recidivism but did not re-offend) and false-negative (labeled 'Lower Risk' for recidivism but did re-offend) results. Northpointe has disputed the ProPublica article (Angwin et al. 2016).

Since the article was published, there have been several studies made to further investigate COMPAS, as well as provide commentary on the article. What follows are some methodologies used in these studies to reveal different insights. Such studies are appreciated as it is crucial for high-stakes systems like COMPAS to predict recidivism accurately, fairly and without bias (Angwin et al. 2016).

## 1. Methods – Analysis and Modeling

COMPAS risk predictions are based on multivariate logistic regression modeling where results are converted into a decile score (Brennan et al. 2009). The system contains recidivism risk scales which are linear functions using Lasso regression. Its “Internal Prison Classification” system used for managing offenders implements bootstrapped K-means methods. Random forest and support vector machine modelling are used to test system accuracy (Singh et al. 2018, 52).

A study conducted a test to observe how a crowdsourced evaluation of criminal records compares to COMPAS in determining recidivism (Dressel and Farid 2018). The crowdsourced results were used as training data to create a predictive linear regression model which was compared against COMPAS to find almost similar results.

The above study has been contested with another one also conducting a crowdsourced survey but did not give participants immediate feedback after each of their evaluation of criminal record vignettes (Lin et al. 2020). The study hypothesized by not offering feedback these recidivism algorithms can predict better than humans. Making such evaluations is more attune to the court environment where the outcomes based on using risk assessment tools by the judicial system might not be known for years. This survey data was used to create logistic regression models which was also compared against COMPAS and found that participants performed significantly worse without feedback.

Another study attempted to reconstruct the COMPAS model and concluded that the system factors more on age than race in determining recidivism risk for a defendant, thus disagreeing with the ProPublica article (Rudin et al. 2020). Rudin’s team ran logistic regression models like Angwin’s team but factored age into which led to different insights. Their main argument is since COMPAS is a proprietary black box system they cannot fully evaluate its complete functionalities for fairness, which they find is a major concern.

## **2. Data Review**

Several of the studies described above used the publicly open-source ProPublica dataset that was originally procured by Angwin's team for the "Machine Bias" article (Angwin et al. 2016; Larson et al. 2016). Two years' worth of COMPAS score data in 2013 and 2014 were obtained from the Broward County Sheriff's Office in Florida and used those 18,610 records to match criminal and incarceration records from Florida's available public resources online. The arduous effort to join and clean the data alone took almost a year (Christian 2020, 60).

Crowdsourced participants for the Dressal and Farid study as well as the Lin study came from Amazon Mechanical Turk. Both studies paid participants \$1.00 for completing their respective tasks but were also offered an incentive bonus as much as \$5.00 if they performed well (Dressel and Farid 2018; Lin et al 2020).

## **3. Applications to Management**

The concerns of racial biases in false-negative and false-positive decisions from risk assessment systems are an on-going discussion. These matters are important as these systems are used to make high-stakes decisions that can have a major effect on peoples' lives. Without full transparency on a proprietary black box system like COMPAS the public cannot be fully aware as to how fair it makes these decisions (Rudin et al. 2020). While there are benefits in efficiency with having these systems assist the criminal justice system, there needs to be open feedback for their transparent continuous improvement to reach the goal of fairness for all.

## References

- Angwin, Julia, Jeff Larson, Surya Mattu and Lauren Kirchner. 2016. "Machine Bias: There is software that is used across the country to predict future criminals. And it is biased against blacks." *ProPublica*, May 23, 2016. Retrieved October 7, 2021.  
<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
- Brennan, Tim, William Dieterich, and Beate Ehret. 2009. "Evaluating the Predictive Validity of the Compas Risk and Needs Assessment System." *Criminal Justice and Behavior* 36, no. 1 (January): 21–40. <https://doi.org/10.1177/0093854808326545>
- Christian, Brian. 2020. *The Alignment Problem: Machine Learning and Human Values*. New York: W.W. Norton & Company.
- Dressel, Julia, and Hany Farid. 2018. "The Accuracy, Fairness, and Limits of Predicting Recidivism." *Science Advances* 4, no. 1 (January): eaao5580.  
<https://doi.org/10.1126/sciadv.aao5580>
- Larson, Jeff, Surya Mattu, Lauren Kirchner, and Julia Anguin. 2016. "How We Analyzed the COMPAS Recidivism Algorithm." *ProPublica*, May 23, 2016. Retrieved October 7, 2021.  
<https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>
- Lin, Zhiyuan, Jongbin Jung, Sharad Goel, and Jennifer Skeem. 2020. "The Limits of Human Predictions of Recidivism." *Science Advances* 6, no. 7 (February): eaaz0652.  
<https://doi.org/10.1126/sciadv.aaz0652>
- Ohlin, Lloyd E. 1951. *Selection for Parole, A Manual of Parole Prediction*. New York: Russel Sage Foundation.
- Rudin, Cynthia, Caroline Wang, and Beau Coker. 2020. "The Age of Secrecy and Unfairness in Recidivism Prediction." *Harvard Data Science Review* 2, no. 1 (Winter).  
<https://doi.org/10.1162/99608f92.6ed64b30>
- Singh, Jay P., Daryl G. Kroner, J. Stephen Wormith, Sarah L. Desmarais, and Zachary Hamilton. 2018. *Handbook of Recidivism Risk/Needs Assessment Tools*. Hoboken, NJ: Wiley Blackwell.