Invisible algorithms in criminal justice

Sunny Ughareja <sunnyu@cmu.edu>,
Omar Nasr <onasr@andrew.cmu.edu>

Carnegie Mellon University

Introduction

- Risk assessment scores are used at every step of the criminal justice system from when a police officer stops someone, to assigning bail, and even during sentencing.
- These scores are increasingly common and influence fundamental decisions about a defendant's freedom.

Objective

- To determine to what extent do preconceived biases infiltrate these invisible algorithms.
- To reach a conclusion regarding the viability of these algorithms based on statistical studies.
- To propose potential solutions for existent issues.

Findings

- The algorithms to predict risk and recidivism are almost always kept secret and not open to public scrutiny.
- Northpointe's COMPAS, the most widely used algorithm, was found to be 61% accurate in predicting recidivism (Broward County, FL), only slightly better than a coin flip.
- Removing variables that can be correlated with race reduce the accuracy of the assessment, defeating the purpose of the risk score.

Compelling arguments in favor

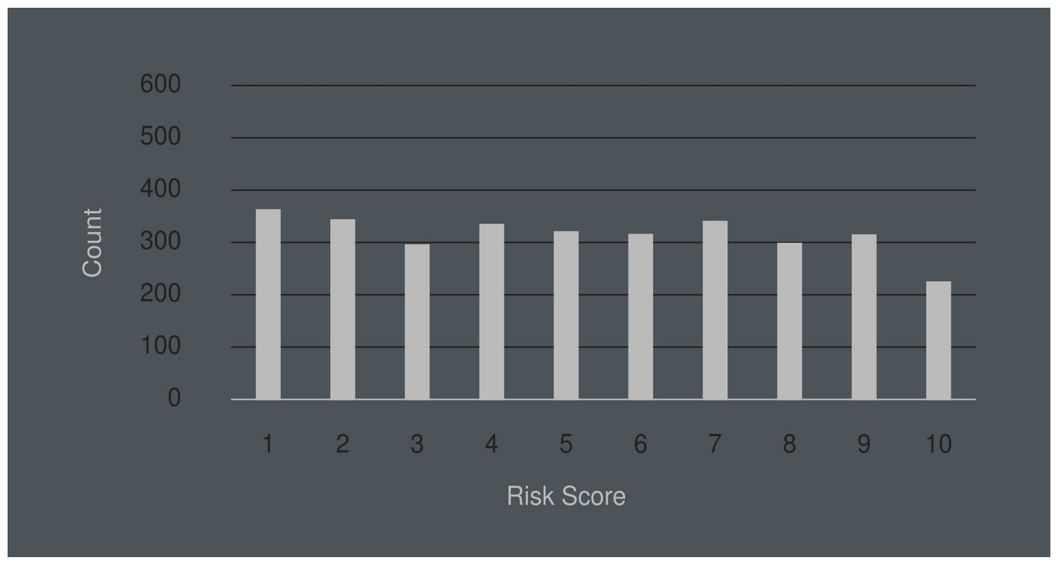
Used to effectively reduce incarceration rates, fill limited seats in rehabilitation programs and reduce human bias.

Possibility of incorrect assessments and bias

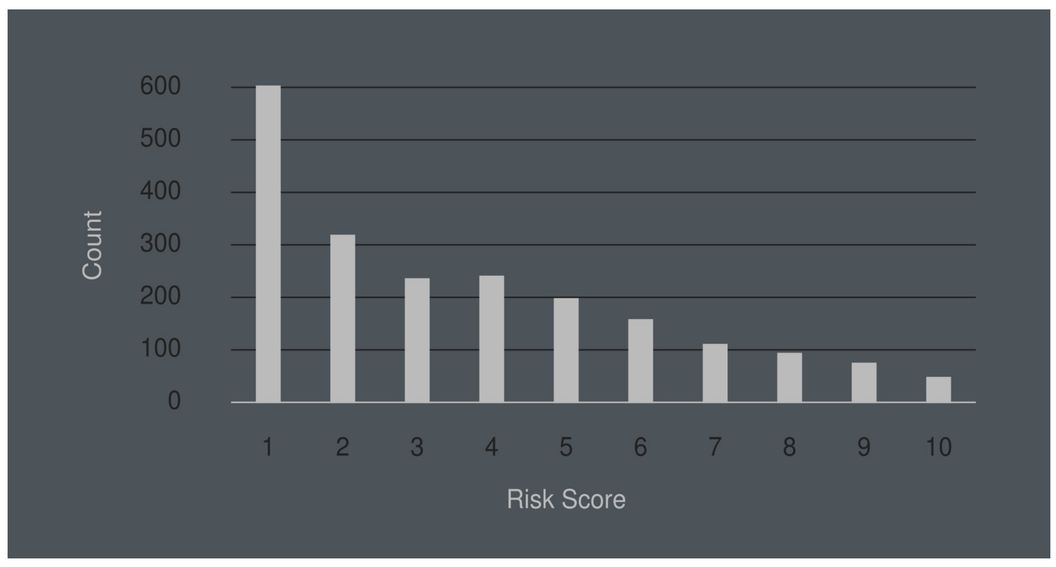
Could negatively impact fairness of decisions (by police and judges) and there is evidence suggesting algorithms are biased against black defendants.

	White	African American
Labeled Higher Risk, But Didn't Re-Offend	23.50%	*44.90%*
Labeled Lower Risk, Yet Did Re-Offend	47.70%	28.00%

Black Defendants' Risk Scores



White Defendants' Risk Scores



Recommendations

- Companies must be more transparent regarding weighted variables.
- Risk factors must only play a small role in overall sentencing.
- More statistical studies on other similar algorithms must be continuously conducted.

Carnegie Mellon University