# Predictive Algorithms in The United States Justice System: What We See in the Dark

Avalyn Jasenn
COMP3190 - Intro to AI
University of Manitoba, 2018

**Predictive Algorithms in the United States Justice System: What We See in the Dark**

## Faith in the Age of Information

Predictive algorithms are rapidly becoming an accepted part of human life. These algorithms are designed to aid in human reasoning, at times replacing the role of human judgment completely. The adoption and full integration of these systems into our society - including our legal and financial systems - has raised ethical and legal concerns surrounding their effects, with evidence leading towards their potential to manifest mass generalizations and unfair treatment. Faith in the use of these predictive assessments to aid in judgment has been accepted widely throughout the United States Justice System, without the appropriate level of understanding obliged by their role in protecting the human rights and freedoms of the public.

## A Brief History of Justice

Recidivism is defined as the relapse into criminal behaviour by an individual who has been previously involved and charged with a criminal offense. It has been accepted by the majority of the public that predictive algorithms now control essential aspects of our lives such as our credit rankings, however controversy and debate have begun to flourish surrounding the fairness and accuracy of these practices, most notably in the practices of the United States Correctional System.

Scientifically calculated predictions are nothing new to the justice system, having been utilized by the U.S. Criminal Justice System for nearly a century. The system began to shift from capital and corporal punishment towards rehabilitation in the late 19th century and began applying the use of crime predictions in the early 1920s.

The extraordinary levels of freedom granted to judges to apply sentences and treatment with the ultimate goal of reentry into society quickly became discriminatory. This ‘freedom’ instead became an abuse of power with minorities sentenced with disproportionate severity in comparison to white offenders across the country(Kehl,

Guo, Kessler, 2017). Attempts at rehabilitation were reported (by methods now considered to be highly flawed) to have no effect, and widely accepted conclusions that "nothing works" became a powerful influence on correctional methods and criminological ideals(Mackenzie, 2017).

In the 1960s criminal prediction practices turned their focus to identifying the capacity of individuals to commit violent crimes, and in the 1970's sentencing methods began to reform back towards notions of retribution(Kehl, Guo, Kessler, 2017). "Tough on time" idealists began to attack the utilitarian goals of rehabilitation as "coddling criminals," pushing to instead control crime through incapacitation. The justification being that imprisoned criminals cannot cause crime, and the longer they are incarcerated the less threat they pose to the public(Mackenzie, 2017); and despite the predictions of violent criminal tendencies being fraught with overwhelming inaccuracies, this has led to the prolonged incarceration of those individuals deemed "high risk"(Kehl, Guo, Kessler, 2017).

Increases in the correctional populations in the United States have skyrocketed since 1975, ushering in the age of mass incarceration. The average rate of 106 inmates per 100,000 adults in the population during the age of individualized rehabilitation and sentencing(Mackenzie, 2017) has grown, as of July 2017, to more than 650 inmates per 100,000 adults(statista.com, 2017). A report by United States Sentencing Commission determined that within eight years nearly half(49.3%) of all offenders who were released on probation from a federal prison were rearrested, with nearly one-third(31.7%) also reconvicted, and one quarter(24.6%) reincarcerated(Dumville, Hunt, 2016).

In most recent years there has been an attempt to find a constructive middle ground between methods of criminal rehabilitation and retributivism using evidence-based practices.(Kehl, Guo, Kessler, 2017). The use of algorithms to assess individuals' risk of recidivism has become such common practice across the United States that some states now require their use by law. These predictive instruments vary greatly in their development methods(many privately owned, others generated by state-led sanctions) but they frequently focus on data collected from the offender's record, as well as their biological, psychological, and social characteristics. These

systems have been implemented in various forms at every stage of the justice system(from initial arrest, through pretrial, to sentencing), studies finding them to predict individuals with a higher risk to recidivate with “some degree of accuracy,” and crediting them “with greater predictive validity than that of professional clinical judgement”(Monahan, Skeet, 2014).

## A Compas to Lead the Way

The Correctional Offender Management Profiling for Alternative Sanctions or 'COMPAS', is a predictive black-box algorithm, first developed in 1998 it has become the most widely used risk assessment software in the United States Justice System today(Kehl, Guo, Kessler, 2017). Northpointe, the (for-profit)owner of this proprietary software, offers their core product as a set of scores assigning the individual a risk score of 1-10(10 being the highest risk individuals) on three potential risk scales(Pretrial Release Risk, General Recidivism, and Violent Recidivism), and an assignment to 1 of 8 criminal typology assignments. These are measured on a “criminogenic needs” scale, assessing 19 core attributes ranging from the subjects cognitive behaviour, history of violence and level of social isolation, to their family criminality, financial problems and level of education. The combination of risk scores and typology assignments provided with the intention to aid decisions on the individual's pretrial release, the opportunity for parole, and appropriate rehabilitation methods(Northpointe, Practitioners Guide to Compas, 2011).

Data is collected about each subject through a set of 137 questions, these to be answered by the subjects as well as pulling information from their criminal record. Some examples of these questions to be answered directly by the offenders are(Northpointe, Sample COMPAS Risk Assessment Score, 2016):

> 37. Did a parent or parent figure who raised you ever have a drug or alcohol problem?
>
> Yes/No

66. Do some of your friends or family feel they must carry a weapon to protect themselves in your neighbourhood?

Yes/No

(The following responses to be answered on a Likert Scale ranging from "Strongly Disagree" to "Strongly Agree")

104. "I feel lonely."

110. "I have a best friend I can talk with about everything."

127. "A hungry person has a right to steal."

In 2009 the creators of the COMPAS published a validation study of the algorithm's accuracy. Defining a 'correct' rating as the assignment of a high risk level to those who did re-offend, and a low risk level to those who did not reoffend, the system correctly rated an average of 68% of the 2,328 sample individuals. When comparing racial disparities the system was found to have comparable rates, 67% accuracy rate for black men and a 69% accuracy rate for white men. Overall their assessment concluded that COMPAS's predictive power was "comparable to, and in some cases higher than, similar risk predictive instruments in this field"(Brennan, Dieterich, Ehret, 2009).

"*To expose abuses of power and betrayals of the public trust by government, business, and other institutions, using the moral force of investigative journalism to spur reform through the sustained spotlighting of wrongdoing.*"

*-Mission state of the independently operated ProPublica newsroom*

Is the mission statement of ProPublica, a nonprofit, four-time Pulitzer Prize-winning, online publication. In their recent pursuit of an altruistic examination of the, as they phrase it "powerful, largely hidden effect of algorithms in American life,"(Angwin, Larson, Mattu, Kirchner, 2016) ProPublica set their sites on COMPAS.

ProPublica designed and conducted their own analysis of the of the algorithms risk level predictions and on May 23, 2016, published their findings alongside a

provocative article, aptly named “Machine Bias: There’s software used across the country to predict future criminals. And it’s biased against blacks.” Their cover(num subjects) in Broward County, Florida(COMPAS currently in use there) study concluded that, among other disparities ranking at least a further investigation, the treatment of black and white defendants varied dramatically. Their findings did not, however, contradict COMPAS’s assessment of the equal accuracy of correct predictions, but found consistent disparate treatment in the misclassifications. Black defendants were almost twice as likely to be considered high risk yet not actually recidivate(45% vs. 24%), while white defendants reoffended yet were considered low risk in nearly the same amounts(48% vs 28%). Due to this consistent misclassification “COMPAS under-classified white reoffenders as low risk 70.5% more often than black reoffenders(48% vs. 28%).” Beyond this, ProPublica considered the algorithm to be “somewhat more accurate than a coin flip”, with only 61% of this predicted to be of high risk of general recidivism reoffending within the next two years, and only 20% of those predicted to commit violent crimes(Larson, Mattu, Kirchner, Angwin, 2016).

This prompted Northpointe’s near-immediate formal response in the form of the technical report “COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity” published on July 8, 2016, rejecting ProPublica’s conclusion of racial bias in COMPAS risk scales. Northpointe’s research department discrediting ProPublica’s study to be conducted without a “solid understanding of the techniques and methodological nuances common to this type of work”(Equivant.com, 2016) and their false conclusions due to technical errors, mis-defined classifications, and incorrect interpretation and use of the models(Dieterich, Mendoza, Brennan, 2016).Behind all of Northpointe’s technical disproofs and discreditation of ProPublica’s methods stands their validation of COMPAS’s fairness through predictive parity. Northpointe exemplifies achieved predictive parity through COMPAS’s consistent accuracy(roughly 60%), insisting that a test that is correct in equal proportions cannot be biased(Dieterich, Mendoza, Brennan, 2016).

Regardless, ProPublica’s article had already claimed massive attention, spreading virally likely due to Journalists irresistible opportunity to publish stories with titles such as “State Governments Use Racially Biased Software to Predict Crimes”. A hailstorm of debates and scientific scrutiny of ProPublica’s findings has followed, and

ProPublica's study and the scholarly papers it has inspired now together represent "the most far-reaching critique to date of the fairness of [criminal risk assessment] algorithms"(Angwin, Larson, 2016).

## Law and Order

> *"...an explicit embrace of the otherwise-condemned discrimination, sanitized by scientific language."*
>
> - Dr.Starr on the use of risk assessment scores in criminal sentencing, Michigan Law Review(Starr, 2014)

State vs Loomis is one of the first major cases in the United States of America to address whether a judge's consideration of software-generated risk assessment scores during sentencing is a violation of a states requirement to respect all legal rights or 'due process'. February 2013 Eric Loomis was arrested for operating the vehicle used in a drive-by shooting. Ultimately Loomis pled guilty to the lesser charges of attempting to flee a traffic officer, and operation of a motor vehicle without the owner's consent(Lvovsky, 2017). Following the guilty plea, the court requested a pre-sentence investigation report, which included a risk score calculated using COMPAS. On all three scales, pretrial recidivism, general recidivism, and violent recidivism, Loomis(who's record also contained a previous conviction of third-degree sexual assault) was designated, as a high risk offender and in 2016 was sentenced to six years in prison(Kehl, Guo, Kessler, 2017). Loomis's lawyer challenged this sentence, arguing the violation of due process on account of the following three factors(Kehl, Guo, Kessler, 2017):

1. The proprietary nature of the COMPAS software prevented him from assessing the accuracy of the score, therefore violating his right to be sentenced on accurate information.
2. The risk assessment system's reliance on information inferred by the characteristics of a larger group violated his right to an individualized sentence.

3. The COMPAS software uses the discriminatory, and therefore, illegitimate variable of gender in determining its calculations.

Although all three judges on the panel appeared to have hesitations surrounding the potential for bias and other concerning implications, in conclusion, they strongly endorsed the use of risk assessment scores in criminal processing and denied Loomis' claims, responding to his claims with the following(Kehl, Guo, Kessler, 2017):

1. Loomis may not have been able to see 'how' his score was calculated due to COMPAS's proprietary nature, however, the risk scores were calculated from a combination of information available on public record and the questionnaire he himself completed. Concluding Loomis was able to insure their accuracy.

2. The COMPAS risk assessment score can assist in providing the courts with more complete information but it was neither the sole, nor, determinative factor in their sentencing decision.

3. The rehabilitation needs and rates of recidivism differ between men and women. They argued that in this case, the use of gender as a variable is solely for "statistical norming," and the promotion of accurate assessments is in the best interests of both institutions and defendants.

This case did, however, lead to the judges to make several recommendations to how COMPAS should be implemented in sentencing. They advised that COMPAS may be considered as a factor in issues of public safety as considering non-prison alternatives, probation and supervision terms. COMPAS should not, however, be counted as a mitigating or aggravating factor towards the sentence, it should not be used as a determinant of the length, nor severity of punishment. These seem to be fair advisement of ethical implementations of this software, however, they provide a judge

with no clear method as to how to appropriately account for a risk score as a factor in a defendant's case.(Kehl, Guo, Kessler, 2017).

There are currently no laws governing 'how' the United States legal system is justified in implementing predictive software, nor scientifically assessed regulations set on the producers of such systems. Although predictive software instruments designed to assess recidivism have been used in the United States Correctional System since 2002, the study "Risk Assessment Instruments Validated and Implemented in Correctional Settings in the United States" published in 2013 found there were very few US evaluations examining the predictive validity of assessments completed on instruments commonly used in US corrections. In many cases validity had only been examined in one or two studies and most often completed by the same group who developed the instrument; and although the risk assessment instruments varied widely in the information used to derive scores(from offender self-reported interviews, to use of official records) and the number of included instruments(from 4 -130) no one instrument emerged as systematically producing more accurate results than the others(Desmarias, Sigh, 2013). Despite the thoroughly underwhelming evidence towards the benefits of these complicated, expensive algorithmic systems, their use has been aggressively encouraged to judges in recent years by legal experts and legislatures. A current proposed revision of the Model Penal Code(S.1917- Sentencing Reform and Corrections act of 2017) even explicitly endorses their use, with only cautionary notes made on the requirements to assure their fair and accurate implementation through validated studies and research.

> *"One of the great accomplishments of the legal order was holding the sovereign accountable for decision making and giving subjects basic rights", "...if law and due process are absent from this field, we are essentially paving the way to a new feudal order of unaccountable reputational intermediaries."* (Citron, Pasquale III, 2014)

## Justified Concern

Over the past twenty years, a critical number of complaints have surfaced surrounding the use of credit scores, systems being found to engage in such controversial behaviour as penalizing "consumers for certain purchases, such as personal counseling." These credit scoring systems have proved to have a disparate impact on both women and minorities and insurers' have been challenged in court for their use, as in the 2007 case against Allstate Insurance. The deficiencies in these credit scoring procedures resulting in "discriminatory action against roughly 5 million African-American and Hispanic customers", for which Allstate ultimately agreed to pay a multi-million dollar settlement(Citron and Pasquale III, 2014).

Current laws do not require consumer reporting agencies to disclose the methods used to calculate an individual's credit score; just as there are no laws that require an explanation of reasoning by criminal risk assessment systems. The data on which these calculations are formed may be disputed but private companies are allowed to keep their methods proprietary.

> *"Our risk scales are able to identify groups of high risk offenders — not a particular high-risk individual"* - Northpointe, on their COMPAS risk assessment system

The concept of individualism, the idea that people have a right to treatment as individuals under the law, is at the heart of the Supreme Court's Equal Protection Jurisprudence. "The supreme court have consistently held that otherwise-impermissible discrimination cannot be justified by statistical generalizations about groups, such as race or gender — even if those generalizations are, on average, accurate"(Kehl, Guo, Kessler, 2017).

Published in the state of Wisconsin's "Department of Corrections Electronic Case Reference Manual" was an observation:

> *"An offender who is young, unemployed, has an early age-at-first-arrest and a history of supervision failure, will score a medium or high on the Violence Risk Scale even though the offender never had a violent offence."*

Now, consider the fact that Judge's in Wisconsin are elected, therefore they must campaign to win their position. Assigning a more lenient sentence to an algorithmically labeled "high risk" offender (who may or may not commit further crimes) may be seen as a risk to not just the public but to the Judge's own career(Kehl, Guo, Kessler, 2017). However, studies conducted on both high and low risk groups of prisoners found that all those who spent more time in prison had higher recidivism rates than those of their counterparts, the lower risk group fairing even worse than the high risk group, with a 4% higher recidivism rate(Gendreau, Goggin, Cullen, 1999).

This creates a situation where a judge may have to defend his or her own reasoning against the reasoning of a risk assessment system, whose creators have offered no explanation for and have no legal obligation to defend.

## Wisdom in the Information Age

> *"... seem to talk to you as though they were intelligent, but if you ask them anything about what they say, from a desire to be instructed, they go on telling you the same forever."*
>
> -Socrates on the radical new information technology, 'the written word'(documented in Phaedrus, Section 275d).

The torrential downpour of big data has fuelled us into the information age, but we have yet to find ways to infuse this data with the innate wisdom of human moral and ethical beliefs. I personally believe that there is no manner of justifying the mandatory

application of a system on a group of people without first conducting exhaustive research on; i) the needs of people ii) diligent, focused and specialized development of technological systems to assist these needs, and iii) thorough testing of the effects the suggested methods may have. Technological advances, however, seem to be first implemented and widely accepted by our governing systems prior to true understanding or concern for the potential effects.

Chairwoman Ramirez has stated the obligation to "evaluate a scoring system's negative, disparate impact on protected groups, arbitrary results, mischaracterizations, and privacy harms", insisting the methods of these predictive scoring systems need to be open to inspection(Kehl, Guo, Kessler).

Only recently has a study been initiated by the Justice Lab at Harvard Law School to evaluate the efficiency of risk assessment scores in pretrial assessments across the United States. Although the laws currently uphold a private company's informational rights to maintain these algorithms as proprietary, as a business concern, the conclusion many have come to- that the potential threat these systems pose to human dignity- raises significantly more concern.

An encouraging study recently published looking at the human decision-making processes in semiautomated decision-making as related to discrimination participants determining approval of loan applications, given fictitious data-mining engine results as reference. The results suggested that offering the explanations and justifications of the decision-making system supported more correct conclusions and reasoning than decisions made on data-mining results presented without justification(Berendt, Preibusch, 2017). As the ultimate goal of these risk assessment, or 'expert systems', is to aid the experts I believe studies of this manner should lead development of these systems, focused on the goal of aiding human judgment; yet, they currently seem to be developed with the aim to mimic, and perhaps not intentionally, but, ultimately overrule it. With the reforms proposed in the 2017 sentencing act, the use of these faceless, unaccountable systems of 'higher reasoning' in the United States Justice System will soon be required by law.

## Faith for the Faithless

There is however, a new radical concept gaining support in the technological community with the values of transparency, accuracy, accountability, participation, and fairness at its core; 'Technological Due Process' aims to ensure "there is ample opportunity to challenge the decisions made by algorithms." Some supporters of this concept argue that algorithms used must be public to ensure scores are fair, accurate and replicable, and that a government entity must be created to thoroughly evaluate all risk-assessment methods(Kehl, Guo, Kessler, 2017).

I personally believe the ignorance of The United States Government has led them to fall far behind in their understanding of the advances in technology and self-education surrounding the potential consequences. This has effectively resulted in the failure of due process, to ensure informed, responsible and justified treatment of their citizens.

## References


Angwin, J., Larson, J. and ProPublica (2016). *Bias in Criminal Risk Scores Is Mathematically Inevitable, Researchers Say — ProPublica*. [online] ProPublica. Available at: https://www.propublica.org/article/bias-in-criminal-risk-scores-is-mathematically-inevitable-researchers-say [Accessed 6 Nov. 2017].

Angwin, J., Larson, J., Mattu, S., Kirchner, L. and ProPublica (2016). *Machine Bias - ProPublica*. [online] ProPublica. Available at: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [Accessed 25 Oct. 2017].

Berendt, B. and Preibusch, S. (2017). *Toward Accountable Discrimination-Aware Data Mining: The Importance of Keeping the Human in the Loop-and Under the Looking Glass*. [ebook] Big Data. Available at: https://www.ncbi.nlm.nih.gov/pubmed/28586238 [Accessed 27 Nov. 2017].

Brennan, T., Dieterich, W., and Ehret, B. (2009). *Evaluating the Predictive Validity of the COMPAS Risk and Needs Assessment System*. [ebook] Criminal Justice and Behavior, Vol. 36 No. 1, January 2009 21-40. International Association for Correctional and Forensic Psychology. Available at: http://www.northpointeinc.com/files/publications/Criminal-Justice-Behavior-COMPAS.pdf [Accessed 29 Nov. 2017].

Citron, D. and Pasquale III, F. (2014). *The Scored Society: Due Process for Automated Predictions. Washington Law Review*, [online] 89. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2376209 [Accessed 27 Nov. 2017].

Desmarias, S. and Singh, J. (2013). *Risk Assessment Instruments Validated and Implemented in Correctional Settings in the United States*. [ebook] Available at: https://jpo.wrlc.org/bitstream/handle/11204/4006/Risk%20Assessment%20Instruments%20Validated%20and%20Implemented%20in%20Correctional%20Settings%20in%20the%20United%20States.pdf?sequence=3&isAllowed=y [Accessed 1 Dec. 2017].

Dieterich, W., Mendoza, C. and Brennan, T. (2016). *COMPAS Risk Scales: Demonstrating Accuracy Equity and Predictive Parity*. [ebook] Northpointe Inc. Research Department. Available at: http://go.volarisgroup.com/rs/430-MBX-989/images/ProPublica_Commentary_Final_070616.pdf [Accessed 29 Nov. 2017].

Dumville, R., and Hunt, K. (2016). *Recidivism Among Federal Offenders: A Comprehensive Overview*. [ebook] United States Sentencing Commission. Available at: https://www.ussc.gov/sites/default/files/pdf/research-and-publications/research-publications/2016/recidivism_overview.pdf [Accessed 1 Dec. 2017].

Equivant.com. (2016). *Demonstrating Accuracy Equity & Predictive Parity*. [online] Available at: http://www.equivant.com/blog/response-to-propublica-demonstrating-accuracy-equity-and-predictive-parity [Accessed 9 Dec. 2017].

Gendreau P., Goggin, C., Cullen, F. (1999). *The Effects of Prison Sentences on Recidivism*. [online] Department of the Solicitor General of Canada. Available at: https://www.prisonpolicy.org/scans/gendreau.pdf [Accessed 1 Dec. 2017]

Kehl, D., Guo, P. and Kessler, S. (2017). *Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing*. [ebook] Responsive Communities Initiative, Berkman Klein Center for Internet & Society, Harvard Law School. Available at: http://nrs.harvard.edu/urn-3:HUL.InstRepos:33746041 [Accessed 7 Dec. 2017].

Larson, J., Mattu, S., Kirchner, L. and Angwin, J. (2016). *How We Analyzed the COMPAS Recidivism Algorithm — ProPublica*. [online] ProPublica. Available at: https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm [Accessed 26 Oct. 2017].

Lvovsky, A. (2017). *State v. Loomis*. [online] Harvardlawreview.org. Available at: https://harvardlawreview.org/2017/03/state-v-loomis/ [Accessed 6 Dec. 2017].

Mackenzie, D.(2001). *Sentencing and Corrections in the 21st Century: Setting the Stage for the Future*.[ebook] Available at: https://www.ncjrs.gov/pdffiles1/nij/189106-2.pdf [Accessed 1 Dec.2017]

Monahan, J. and Skeet, J.(2014). *Risk Redux: The Resurgence of Risk Assessment in Criminal Sanctioning*. [ebook] Federal Sentencing Reporter, Vol. 26, No. 3, Critical Issues in the Use of

Risk Assessments, Prior Record Enhancements, and Probation/Parole Revocation (February 2014), pp. 158-166 Published by: University of California Press on behalf of the Vera Institute of Justice Stable URL: http://www.jstor.org/stable/10.1525/fsr.2014.26.3.158 . [Accessed 1 Nov. 2017]

Northpointe (2011). *Practitioners Guide to COMPAS*. [ebook] Contributed to public document cloud by Julia Angwin, ProPublica (2016). Available at https://www.documentcloud.org/documents/2840784-Practitioner-s-Guide-to-COMPAS-Core.html#document/p30/a334028 [Accessed 20 Nov. 2017].

Northpointe (2016). *Sample COMPAS Risk Assessment Score*. [image] Contributed to public document cloud by Julia Angwin, ProPublica. Available at: https://www.documentcloud.org/documents/2702103-Sample-Risk-Assessment-COMPAS-CORE.html#document/p3/a296596 [Accessed 20 Nov. 2017].

Congress.gov (2017). *S.1917 - 115th Congress (2017-2018): Sentencing Reform and Corrections Act of 2017*. [online] Available at: https://www.congress.gov/bill/115th-congress/senate-bill/1917/text [Accessed 6 Dec. 2017].

Starr, S. (2013). *Evidence-Based Sentencing and the Scientific Rationalization of Discrimination*. Law and Economics Research Paper Series, Paper No. 13-014 Published By: University of Michigan Law School Scholarship Repository.[ebook] Available at: https://repository.law.umich.edu/cgi/viewcontent.cgi?referer=https://www.google.ca/&httpsredir=1&article=1200&context=law_econ_current [Accessed 25 Nov. 2017]

Statista(2017). *Ranking: most prisoners per capita by country 2017 | Statistic*. [online] Available at: https://www.statista.com/statistics/262962/countries-with-the-most-prisoners-per-100-000-inhabitants/ [Accessed 1 Dec. 2017].