Data, Algorithms, Fairness, and Justice

Cristopher Moore, Santa Fe Institute

*Complex Systems Summer School*
Le doute n'est pas un état bien agréable, mais l'assurance est un état ridicule.

Doubt is not a very pleasant state, but certainty is a ridiculous state.

Thank you for contributing
Your contribution benefits millions of Translate users

LEARN MORE
Facebook sends man animation featuring cartoon characters dancing on his mother's grave
Algorithms in the Justice System
On any given day, about 450,000 people in jail awaiting trial
More than 1/1,000 Americans, 70% of the jail population
Most on bail of $2,500 or less
Disproportionately Black and Hispanic
Costs $14 Billion per year
Disrupts families, marriages, jobs; can in fact increase crime
Hard to prepare defense: more guilty pleas, plea bargains
Wealthy can afford bail and be released
500 Women and Teenagers to Be Bailed Out From Rikers by Human Rights Group

“Allowing people who have neither been tried or convicted to languish in jail violates the spirit of the Constitution. Pretrial incarceration has become the penalty for poverty, for mental illness and for blackness.”
— Ritchie Torres, Bronx Councilman

In October, the Robert F. Kennedy Human Rights organization will begin spending up to $5 million to bail out more than 500 women and teenagers from Rikers Island.

Johnny Milano for The New York Times
Salerno v. United States (1987)

“In our society, liberty is the norm, and detention prior to trial or without trial is the carefully limited exception”
— Chief Justice Rehnquist

“This case brings before the Court for the first time a statute in which Congress declares that a person innocent of any crime may be jailed indefinitely... if the Government shows to the satisfaction of a judge that the accused is likely to commit crimes... at any time in the future”
— Justice Thurgood Marshall’s dissent
Big Data to the Rescue?

Why smart statistics are the key to fighting crime

Anne Milgram, former New Jersey Attorney General
What the algorithm thinks

<table>
<thead>
<tr>
<th>Ground truth</th>
<th>low risk</th>
<th>high risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>low risk</td>
<td>![false negatives](false negatives)</td>
<td>![false positives](false positives)</td>
</tr>
<tr>
<td>high risk</td>
<td>![false negatives](false negatives)</td>
<td>![false positives](false positives)</td>
</tr>
</tbody>
</table>
False Positives vs. False Negatives

“It is better that ten guilty persons escape than that one innocent suffer.”

—William Blackstone
False Positives vs. False Negatives

“I'm more concerned with bad guys who got out and released than I am with a few that in fact were innocent.”

— Dick Cheney
Two popular algorithms or “risk assessment tools”

COMPAS

Northpointe / equivant

137-item questionnaire and interview

Proprietary (secret) formula

Arnold Foundation Public Safety Assessment (PSA)

Rapidly growing, 38 jurisdictions so far

9 factors from criminal record

Simple, publicly known formula
What data goes into COMPAS?

Risk Assessment

<table>
<thead>
<tr>
<th>PERSON</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name:</td>
</tr>
<tr>
<td>Gender: Male</td>
</tr>
<tr>
<td>Marital Status: Single</td>
</tr>
<tr>
<td>Agency: DAI</td>
</tr>
<tr>
<td>Offender #:</td>
</tr>
<tr>
<td>DOR:</td>
</tr>
</tbody>
</table>

Current Charges

- [ ] Homicide
- [ ] Robbery
- [ ] Drug Trafficking/Sales
- [ ] Sex Offense with Force
- [X] Weapons
- [ ] Burglary
- [ ] Drug Possession/Use
- [ ] Sex Offense w/o Force
- [X] Assault
- [ ] Property/Larceny
- [ ] DUI/OUIL
- [ ] Arson
- [ ] Fraud
- [X] Other

Criminal History

Exclude the current case for these questions.

7. How many times has this person been arrested before as an adult or juvenile (criminal arrests only)?
   5

8. How many prior juvenile felony offense arrests?
   [ ] 0 [ ] 1 [ ] 2 [ ] 3 [X] 4 [ ] 5+

9. How many prior juvenile violent felony offense arrests?
   [ ] 0 [ ] 1 [X] 2+
What data goes into COMPAS?

Family Criminality

The next few questions are about the family or caretakers that mainly raised you when growing up.

31. Which of the following best describes who principally raised you?
   - ☐ Both Natural Parents
   - ☐ Natural Mother Only
   - ☐ Natural Father Only
   - ☐ Relative(s)
   - ☐ Adoptive Parent(s)
   - ☐ Foster Parent(s)
   - ☑ Other arrangement

32. If you lived with both parents and they later separated, how old were you at the time?
   - ☑ Less than 5 ☐ 5 to 10 ☐ 11 to 14 ☐ 15 or older ☐ Does Not Apply

33. Was your father (or father figure who principally raised you) ever arrested, that you know of?
   - ☑ No ☐ Yes

34. Was your mother (or mother figure who principally raised you) ever arrested, that you know of?
   - ☑ No ☐ Yes
What data goes into COMPAS?

Substance Abuse

What are your usual habits in using alcohol and drugs?

45. Do you think your current/past legal problems are partly because of alcohol or drugs?  
☐ No ☑ Yes

46. Were you using alcohol or under the influence when arrested for your current offense?  
☐ No ☑ Yes

47. Were you using drugs or under the influence when arrested for your current offense?  
☑ No ☐ Yes

48. Are you currently in formal treatment for alcohol or drugs such as counseling, outpatient, inpatient, residential?  
☑ No ☐ Yes

49. Have you ever been in formal treatment for alcohol such as counseling, outpatient, inpatient, residential?  
☑ No ☐ Yes

50. Have you ever been in formal treatment for drugs such as counseling, outpatient, inpatient, residential?  
☑ No ☐ Yes
What data goes into COMPAS?

Residence/Stability

54. How often do you have contact with your family (may be in person, phone, mail)?
   - No family
   - Never
   - Less than once/month
   - Once per week
   - Daily

55. How often have you moved in the last twelve months?
   - Never
   - 1
   - 2
   - 3
   - 4
   - 5+

56. Do you have a regular living situation (an address where you usually stay and can be reached)?
   - No
   - Yes

57. How long have you been living at your current address?
   - 0-5 mo.
   - 6-11 mo.
   - 1-3 yrs.
   - 4-5 yrs.
   - 6+ yrs.

58. Is there a telephone at this residence (a cell phone is an appropriate alternative)?
   - No
   - Yes
What data goes into COMPAS?

Social Environment

Think of the neighborhood where you lived during the past few (3-6) months.

65. Is there much crime in your neighborhood?
   ✔ No ☐ Yes

66. Do some of your friends or family feel they must carry a weapon to protect themselves in your neighborhood?
   ✔ No ☐ Yes

67. In your neighborhood, have some of your friends or family been crime victims?
   ☐ No ✔ Yes

68. Do some of the people in your neighborhood feel they need to carry a weapon for protection?
   ☐ No ✔ Yes

69. Is it easy to get drugs in your neighborhood?
   ✔ No ☐ Yes

70. Are there gangs in your neighborhood?
    ☐ No ✔ Yes
What data goes into COMPAS?

**Education**

Think of your school experiences when you were growing up.

71. Did you complete your high school diploma or GED?
   - [x] No  [ ] Yes

72. What was your final grade completed in school?
   - 9

73. What were your usual grades in high school?
   - [ ] A  [ ] B  [x] C  [ ] D  [ ] E/F  [ ] Did Not Attend

**Vocation (Work)**

Please think of your past work experiences, job experiences, and financial situation.

80. Do you have a job?
   - [x] No  [ ] Yes

81. Do you currently have a skill, trade or profession at which you usually find work?
   - [x] No  [ ] Yes

82. Can you verify your employer or school (if attending)?
   - [x] No  [ ] Yes

83. How much have you worked or been enrolled in school in the last 12 months?
   - [ ] 12 Months Full-time  [ ] 12 Months Part-time  [ ] 6+ Months Full-time  [x] 0 to 6 Months PT/FT

84. Have you ever been fired from a job?
   - [x] No  [ ] Yes
Objections to COMPAS

COMPAS doesn’t use race (it does use gender)
But it uses “proxies” that are correlated with race
Redlining
Objections to COMPAS

COMPAS doesn’t use race (it does use gender)

But it uses “proxies” that are correlated with race

Past arrests: people of color are more likely to get arrested

“Environment” questions, over which you have no control

Employment questions, correlated with income

Juvenile record?

Drug treatment?

Proprietary formula—simple (logistic regression),
but with unknown weights
What data goes into the Arnold PSA?

**RELATIONSHIP BETWEEN RISK FACTORS AND PRETRIAL OUTCOMES**

<table>
<thead>
<tr>
<th>Risk Factor</th>
<th>FTA</th>
<th>NCA</th>
<th>NVCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Age at current arrest</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Current violent offense</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Current violent offense &amp; 20 years old or younger</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Pending charge at the time of the offense</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>4. Prior misdemeanor conviction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Prior felony conviction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Prior conviction (misdemeanor or felony)</em></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Prior violent conviction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Prior failure to appear in the past two years</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Prior failure to appear older than two years</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>9. Prior sentence to incarceration</td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

FTA = Failure to Appear  
NCA = New Criminal Activity  
NVCA = New Violent Criminal Activity
The Arnold PSA

Publicly known
Simple linear point system
Past convictions, not arrests
Does not use juvenile record
Uses age, but not gender, employment, education, or environment

### Public Safety Assessment Risk Factors

<table>
<thead>
<tr>
<th>Risk Factor</th>
<th>Weights</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FAILURE TO APPEAR</strong></td>
<td></td>
</tr>
<tr>
<td>Maximum total weight = 7 points</td>
<td></td>
</tr>
<tr>
<td>Pending charge at the time of the offense</td>
<td>No = 0</td>
</tr>
<tr>
<td></td>
<td>Yes = 1</td>
</tr>
<tr>
<td>Prior conviction</td>
<td>No = 0</td>
</tr>
<tr>
<td></td>
<td>Yes = 1</td>
</tr>
<tr>
<td>Prior failure to appear pretrial in past 2 years</td>
<td>0 = 0</td>
</tr>
<tr>
<td></td>
<td>1 = 2</td>
</tr>
<tr>
<td></td>
<td>2 or more = 4</td>
</tr>
<tr>
<td>Prior failure to appear pretrial older than 2 years</td>
<td>No = 0</td>
</tr>
<tr>
<td></td>
<td>Yes = 1</td>
</tr>
<tr>
<td><strong>NEW CRIMINAL ACTIVITY</strong></td>
<td></td>
</tr>
<tr>
<td>Maximum total weight = 13 points</td>
<td></td>
</tr>
<tr>
<td>Age at current arrest</td>
<td>23 or older = 0</td>
</tr>
<tr>
<td></td>
<td>22 or younger = 2</td>
</tr>
<tr>
<td>Pending charge at the time of the offense</td>
<td>No = 0</td>
</tr>
<tr>
<td></td>
<td>Yes = 3</td>
</tr>
<tr>
<td>Prior misdemeanor conviction</td>
<td>No = 0</td>
</tr>
<tr>
<td></td>
<td>Yes = 1</td>
</tr>
<tr>
<td>Prior felony conviction</td>
<td>No = 0</td>
</tr>
<tr>
<td></td>
<td>Yes = 1</td>
</tr>
<tr>
<td>Prior violent conviction</td>
<td>0 = 0</td>
</tr>
<tr>
<td></td>
<td>1 or 2 = 1</td>
</tr>
<tr>
<td></td>
<td>3 or more = 2</td>
</tr>
<tr>
<td>Prior failure to appear pretrial in past 2 years</td>
<td>0 = 0</td>
</tr>
<tr>
<td></td>
<td>1 = 1</td>
</tr>
<tr>
<td></td>
<td>2 or more = 2</td>
</tr>
<tr>
<td>Prior sentence to incarceration</td>
<td>No = 0</td>
</tr>
<tr>
<td></td>
<td>Yes = 2</td>
</tr>
<tr>
<td><strong>NEW VIOLENT CRIMINAL ACTIVITY</strong></td>
<td></td>
</tr>
<tr>
<td>Maximum total weight = 7 points</td>
<td></td>
</tr>
<tr>
<td>Current violent offense</td>
<td>No = 0</td>
</tr>
<tr>
<td></td>
<td>Yes = 2</td>
</tr>
<tr>
<td>Current violent offense &amp; 20 years old or younger</td>
<td>No = 0</td>
</tr>
<tr>
<td></td>
<td>Yes = 1</td>
</tr>
<tr>
<td>Pending charge at the time of the offense</td>
<td>No = 0</td>
</tr>
<tr>
<td></td>
<td>Yes = 1</td>
</tr>
<tr>
<td>Prior conviction</td>
<td>No = 0</td>
</tr>
<tr>
<td></td>
<td>Yes = 1</td>
</tr>
<tr>
<td>Prior violent conviction</td>
<td>0 = 0</td>
</tr>
<tr>
<td></td>
<td>1 or 2 = 1</td>
</tr>
<tr>
<td></td>
<td>3 or more = 2</td>
</tr>
</tbody>
</table>

*Source: Laura and John Arnold Foundation*
What does “accuracy” mean anyway?
lower risk score

or

higher risk score
lower risk score or higher risk score?
lower risk score  or  higher risk score
AUC = 0.7
AUC = 0.7

lower risk score higher risk score
Receiver Operating Characteristic Curve

AUC = 0.7

% green accepted

% red accepted
AUCs in practice

COMPAS: 0.69–0.71
Arnold PSA: 0.64–0.66
tradeoff for transparency?
Random people: 0.71
Low compared to medicine, other AI tasks
Not too bad for social science... people are hard to predict
Lumping into categories

Bottom 10%: 2
Lower risk score: 3 4 5 6 7 8 9
Top 100%: 10
Higher risk score: 1 2 3 4 5 6 7 8 9
Finally...

<table>
<thead>
<tr>
<th>“low risk”</th>
<th>“medium”</th>
<th>“high risk”</th>
</tr>
</thead>
<tbody>
<tr>
<td>![Diagram](false negatives)</td>
<td>![Diagram](false positives)</td>
<td></td>
</tr>
</tbody>
</table>
Fairness and the ProPublica Debate
Machine Bias

There's software used across the country to predict future criminals. And it's biased against blacks.

by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica

May 23, 2016
False Positives, False Negatives, and False Analyses: A Rejoinder to “Machine Bias: There’s Software Used Across the Country to Predict Future Criminals. And it’s Biased Against Blacks.”

Anthony W. Flores, Ph.D.
California State University, Bakersfield

Christopher T. Lowenkamp, Ph.D.
Administrative Office of the United States Courts
Probation and Pretrial Services Office

Kristin Bechtel, M.S.
Crime and Justice Institute at CRJ
"low risk"  "high risk"
“low risk”

“high risk”

\[
\frac{21}{27} = 78\% \text{ correct}
\]
“low risk”

“high risk”

\[
\frac{21}{27} = 78\% \text{ correct}
\]

\[
\frac{6}{27} = 22\% \text{ incorrect}
\]
false discovery rate 3/12=25%
false positive rate 3/15 = 20%
false discovery rate 3/12 = 25%
false positive rate $2/6 = 33\%$
false positive rate 2/6 = 33%
false positive rate 1/9 = 11%
false discovery rate 2/8 = 25%
false discovery rate 1/4 = 25%
predictive parity = 25%
predictive parity

“low risk”

“high risk”

predictive parity
predictive parity
disparate impact

"low risk"

"high risk"

predictive parity
disparate impact
Which definition of fairness is correct? It depends on your point of view... Are you a judge or a defendant?
Judge: "low risk"

Defendant: "high risk"
We can’t have predictive parity and make the false positive rates the same for both groups!*

*unless the base rates are identical or the algorithm is perfect.
Where do we go from here?
Point #1: No Black Boxes

Should proprietary algorithms ever be used in the public sector?

Intellectual property is no excuse

Secrecy vs. security: voting machines

Need transparency throughout the pipeline:

How was the training data collected?

What kind of algorithm is it, and how was it trained?

Can anyone run it, reproduce it, explain it, or contest it?

Once it is deployed, how well does it work?
Point #2: Embrace Uncertainty

- Asset Intercepted: 86.09%
- Confrontation: Violent: 97.23%
- Intervention Necessary
Point #2: Embrace Uncertainty

should a judge see this:

“In a sample in Kentucky from 2014-2015, 26% of defendants in this category were re-arrested before their trial, and 3% were arrested for a violent offense. Studies of your local population are ongoing to see if similar statistics hold locally”
Point #2: Embrace Uncertainty

or this:
Point #3: Explaining and Contesting Decisions

General Data Protection Regulation (GDPR):

Article 22: The data subject shall have the right not to be subject to a decision based solely on automated processing.

The data controller shall safeguard... the right to obtain human intervention... and to contest the decision.

[Recital 71: to obtain an explanation of the decision... and to challenge the decision.]

Article 15: The data subject shall have the right to obtain from the controller... the existence of automated decision-making, including profiling... and meaningful information about the logic involved.
But what is an explanation anyway?

Not so clear for human decisions either...

[A]pplicants who have lived at their present address for less than six months are awarded 39 points, a level which they could not reach again until they had maintained the same residence for seven and one-half years. Furthermore, applicants who have been residents for between six months and 1 year 5 months (30 points) are considered more creditworthy than those who have been residents for between 1 and 1/2 years and 3 years 5 months (27 points).

Psychology: we make explanations up after the fact

Law: “preponderance of evidence”, “reasonable doubt”
Counterfactual explanations: What if the input were different?

“Why Should I Trust You?”
Explaining the Predictions of Any Classifier

Marco Tulio Ribeiro
University of Washington
Seattle, WA 98105, USA
marcotcr@cs.uw.edu

Sameer Singh
University of Washington
Seattle, WA 98105, USA
sameer@cs.uw.edu

Carlos Guestrin
University of Washington
Seattle, WA 98105, USA
guestrin@cs.uw.edu

(a) Husky classified as wolf  (b) Explanation

Requires black-box access
Point #4: Science and Politics

“Decision making framework” turns scores into actions; chosen by District Attorney, Police Chiefs, Public Defenders...

But politicians might tweak framework, algorithm, or definitions (Santa Cruz, New Jersey lawsuit)

In the 80s Congress overrode sentencing reform: crack cocaine

<table>
<thead>
<tr>
<th></th>
<th>NCA 1</th>
<th>NCA 2</th>
<th>NCA 3</th>
<th>NCA 4</th>
<th>NCA 5</th>
<th>NCA 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>FTA 1</td>
<td>ROR 11.8% of population</td>
<td>ROR 7.7% of population</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FTA 2</td>
<td>ROR 0.5% of population</td>
<td>ROR 6.9% of population</td>
<td>PML 1 11.7% of population</td>
<td>PML 2 6.4% of population</td>
<td>PML 3 0.1% of population</td>
<td></td>
</tr>
<tr>
<td>FTA 3</td>
<td></td>
<td></td>
<td>PML 1 2.0% of population</td>
<td>PML 1 8.8% of population</td>
<td>PML 2 6.0% of population</td>
<td>PML 3 2.9% of population</td>
</tr>
<tr>
<td>FTA 4</td>
<td></td>
<td></td>
<td>PML 1 0.6% of population</td>
<td>PML 1 1.5% of population</td>
<td>PML 2 3.5% of population</td>
<td>PML 3 4.8% of population</td>
</tr>
<tr>
<td>FTA 5</td>
<td></td>
<td></td>
<td>PML 2 0.0% of population</td>
<td>PML 2 0.4% of population</td>
<td>PML 3 2.0% of population</td>
<td>PML 3 + EM/HD 2.4% of population</td>
</tr>
<tr>
<td>FTA 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Release Not Recommended 0.1% of population</td>
<td>Release Not Recommended 0.3% of population</td>
</tr>
</tbody>
</table>
When Politics Pretends to be Science

ICE lowered threshold on their Classification Assessment” to zero, so it always recommends detention

Easy to say “the algorithm made us do it…”

THE IMMIGRATION DETENTION RISK ASSESSMENT

MARK NOFERI AND ROBERT KOU LIS H

current over-detention trends. The unique aspects of immigration enforcement, laws, and the impacted population will likely frustrate accurate calibration of the risk tool, and effective implementation of even a calibrated tool—in turn frustrating constructive impact of ICE’s risk assessment initiative on over-detention. Consequently, the immigration risk assessment may only add a scientific veneer to enforcement that remains institutionally predisposed towards detention and control.
Point #5: Don’t Predict the Future, Change It

If we train algorithms on biased decisions, they will learn these biases, but give them a sheen of objectivity “Tech-washing”

Algorithms should help us overcome our cognitive weaknesses, not encode them
She is a doctor

O bir doktor

She is a doctor (feminine)

He is a doctor (masculine)
Amazon Pulled the Plug on an AI Recruitment Tool That Was Biased Against Women
The current culture of machine learning
The current culture of machine learning
Don’t reproduce the statistics of the past:
Make a better future