Risk assessment and racial bias

Alexandra Chouldechova
Assistant Professor, Statistics and Public Policy
Heinz College, Carnegie Mellon University

Risk assessment
A brief history and taxonomy

• Dates back to the 1920s in the work of Ernest Burgess, who developed a tool to predict recidivism risk for offenders released in Illinois

• Early versions were based on clinical judgment
  • Take the factors that experts believe to be associated with reoffense risk
  • Develop a scoring system that tallies up risk factors

• Actuarial instruments (based on statistical models or “machine learning”) produce scores that have higher predictive accuracy
Actuarial instruments

- **First generation**
  - Unstructured
  - Based on “clinical” judgment

- **Second generation (STATIC-99)**
  - Actuarial tools (statistical models)
  - Based on static factors (e.g., age, criminal history)

- **Third generation (LSI-R)**
  - Mix of Actuarial and clinical judgment
  - Consider static and dynamic items (e.g., attitudes, behavioral health)
  - Risk-Needs assessment

- **Fourth generation (COMPAS, ORAS)**
  - Integrates case planning and risk management
  - Helps guide decisions about interventions and supervision
Predictive (racial) bias

• Experts can be biased in their assessments
  • E.g., overestimate risk for some groups relative to others
• Risk assessment models can have the same issue
• An unbiased tool would predict reoffense likelihood with equal accuracy across groups

• Immediate problems:
  • There are many different ways of measuring accuracy
  • We observe rearrest, not reoffense
    • Are we happy with an unbiased prediction of who gets caught?

• Let’s look at an example.

More individuals from the Purple group get flagged as high risk.

Does this mean that the risk assessment tool is biased?
Let’s look at outcomes.

Blue group has lower recidivism rate (4/10 vs. 5/10).

Reoffense rate among those flagged as high risk is the same across groups:

\[
\frac{2}{3} = 67\% \quad \text{vs.} \quad \frac{4}{6} = 67\%
\]

This is called predictive parity.

| PPV  | 67% | 67% |
Machine Bias

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

Source: Julia Angwin, Jeff Larson, SuryaMattu and Lauren Kirchner, ProPublica

Recidivism rate: 40%

Recidivism rate: 50%

<table>
<thead>
<tr>
<th></th>
<th>PPV</th>
<th>FPR</th>
<th>FNR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>67%</td>
<td>17%</td>
<td>50%</td>
</tr>
<tr>
<td></td>
<td>67%</td>
<td>40%</td>
<td>20%</td>
</tr>
</tbody>
</table>

FPR = Proportion of non-recidivists who are misclassified as High Risk

FNR = Proportion of recidivists who are misclassified as Low Risk
**Prediction Fails Differently for Black Defendants**

<table>
<thead>
<tr>
<th></th>
<th>WHITE</th>
<th>AFRICAN AMERICAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labeled Higher Risk, But Didn’t Re-Offend</td>
<td>23.5%</td>
<td>44.9%</td>
</tr>
<tr>
<td>Labeled Lower Risk, Yet Did Re-Offend</td>
<td>47.7%</td>
<td>28.0%</td>
</tr>
</tbody>
</table>

Overall, Northpointe’s assessment tool correctly predicts recidivism 61 percent of the time. But blacks are almost twice as likely as whites to be labeled a higher risk but not actually re-offend. It makes the opposite mistake among whites: They are much more likely than blacks to be labeled lower risk but go on to commit other crimes. (Source: ProPublica analysis of data from Broward County, Fla.)

---

**Non-recidivists come in two forms**

- Reoffends
- Doesn’t reoffend
- Doesn’t get caught
Reoffense rate: 50%

<table>
<thead>
<tr>
<th>PPV</th>
<th>100%</th>
<th>67%</th>
<th>67%</th>
</tr>
</thead>
<tbody>
<tr>
<td>FPR</td>
<td>0%</td>
<td>17%</td>
<td>40%</td>
</tr>
<tr>
<td>FNR</td>
<td>40%</td>
<td>50%</td>
<td>20%</td>
</tr>
</tbody>
</table>

Disentangling concerns

- Risk assessment tools make mistakes
  - So do human assessors

- When thinking about whether the tool could be useful, it can help to step away from issues of model inaccuracy

- Try the “Oracle test”
What questions remain?

Risk assessment tool

Oracle

Omitted objective bias

Any questions that remain may help clarify concerns about:

• Choice of target variable
  • Is rearrest the right target?

• Disconnect between prediction target and decision criteria
  • Should future dangerousness or failure to appear risk factor into bail decisions?

• Explainability
  • Is it enough to know that the individual is high risk, or do you also need to know why?

• Effects of interventions
  • What interventions are at your disposal to reduce risk, and are any of them likely to be effective for the given individual?
Fairness is a process property

An unbiased risk assessment tool may lead to biased outcomes and may have disparate impact depending on how it is used.

Thank you.

Carnegie Mellon University
Heinz College
Information Systems, Public Policy, Management
achould@cmu.edu
Some additional resources


