Machine Learning, Automated Algorithms, and Risk

Daniel B. Neill, Ph.D.

Associate Professor of Computer Science and Public Service
Associate Professor of Urban Analytics
Director, Machine Learning for Good (ML4G) Laboratory

New York University

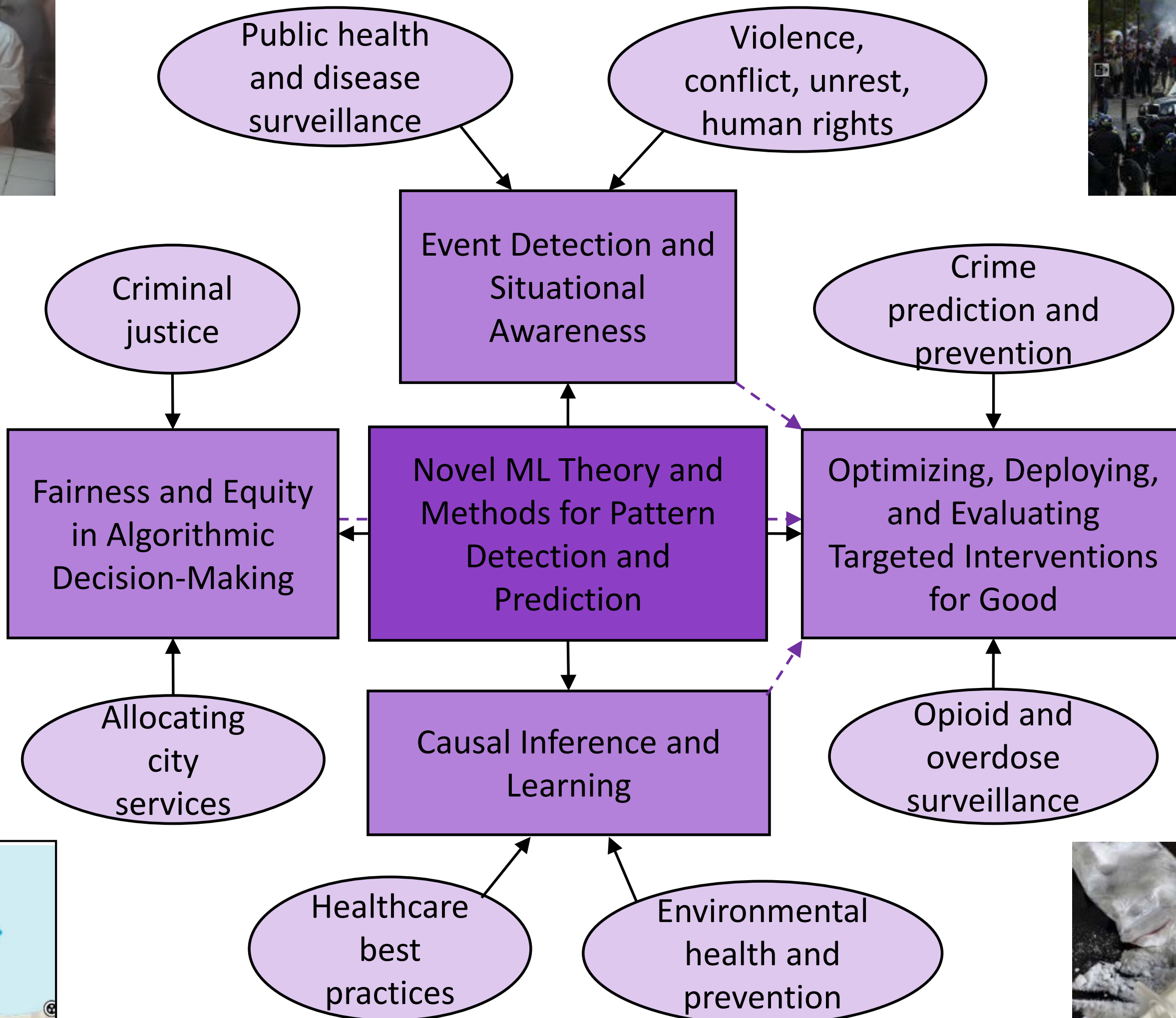
E-mail: daniel.neill@nyu.edu
Web: http://www.cs.nyu.edu/~neill

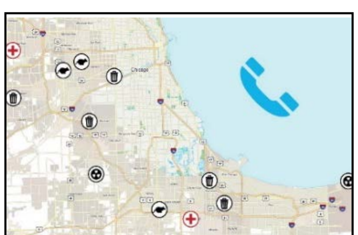
NYU Center for Urban Science + Progress

Disclaimer

The views, opinions, perspectives and outlooks contained herein are strictly for educational purposes, and represent those of the presenter only. They do not represent the views, opinions, perspectives and outlooks of Numerati Partners LLC, NYU, Guy Carpenter, the InsurTech Alliance ("ITA") Sandbox, Designated Affiliates, ITA Participants, ITA Cohort Firms, or any academic institution, governmental entity or other entity or organization with which the presenter may be associated.

The Machine Learning for Good Lab @ NYU





The use of **automated algorithms** for decision making has become increasingly ubiquitous across a wide variety of fields...

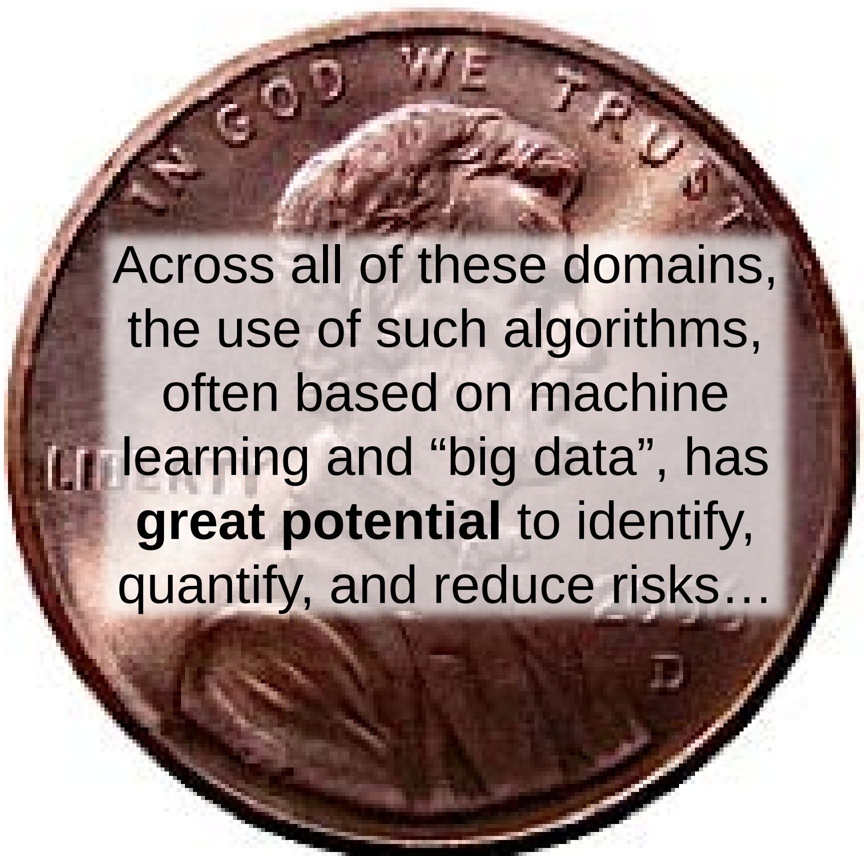
Online marketing

Lending Hiring

Health care diagnosis & treatment

Policing & criminal justice

Allocation of city services



Across all of these domains, the use of such algorithms, often based on machine learning and “big data”, has **great potential** to identify, quantify, and reduce risks…

But their use can also create **new types of risk** (e.g., algorithmic bias and discrimination)

Machine Learning in a nutshell

Machine Learning (ML) is the study of systems that improve their performance with experience (typically by **learning** from data).

ML algorithms have many advantages compared to classical methods: **flexible** and **interpretable** models, **robustness** to “messy” data, and the ability to **scale** to massive, complex datasets.

Supervised Learning

Data/input	Labels/output
x_1	y_1
x_2	y_2
...	...
x_N	y_N

Learn dependence:

$$y = f(x)$$

Discrete y = classification
Continuous y = regression

Example of supervised learning: Predicting risk of, and expected loss resulting from, a severe weather event.

This is most of what you learn in an intro ML class, but only a small fraction of today’s talk!

Machine Learning in a nutshell

Machine Learning (ML) is the study of systems that improve their performance with experience (typically by **learning** from data).

ML algorithms have many advantages compared to classical methods: **flexible** and **interpretable** models, **robustness** to "messy" data, and the ability to **scale** to massive, complex datasets.

Supervised Learning

Data/input	Labels/output
x_1	y_1
x_2	y_2
$\dots$	$\dots$
x_N	y_N

Learn dependence:

$$y = f(x)$$

Discrete y = classification
Continuous y = regression

In unsupervised learning problems, we have a set of data records with no labels.

Unsupervised ML tasks include **clustering** data into groups, and detecting **anomalies** or **outliers**.

What I do is mostly unsupervised learning, specifically, **detection of emerging events and patterns** in massive and complex data.

What are some potential benefits of ML to the (re-)insurance industry?

More accurate advance prediction of risks could improve **underwriting**: pricing, risk selection, new insurance products for emerging risks…

Early detection can provide detailed **situational awareness** prior to first notice of loss (reserving, claims management, fraud detection…)

Can target specific geographic or subpopulation-level **interventions** to reduce risks and mitigate losses.

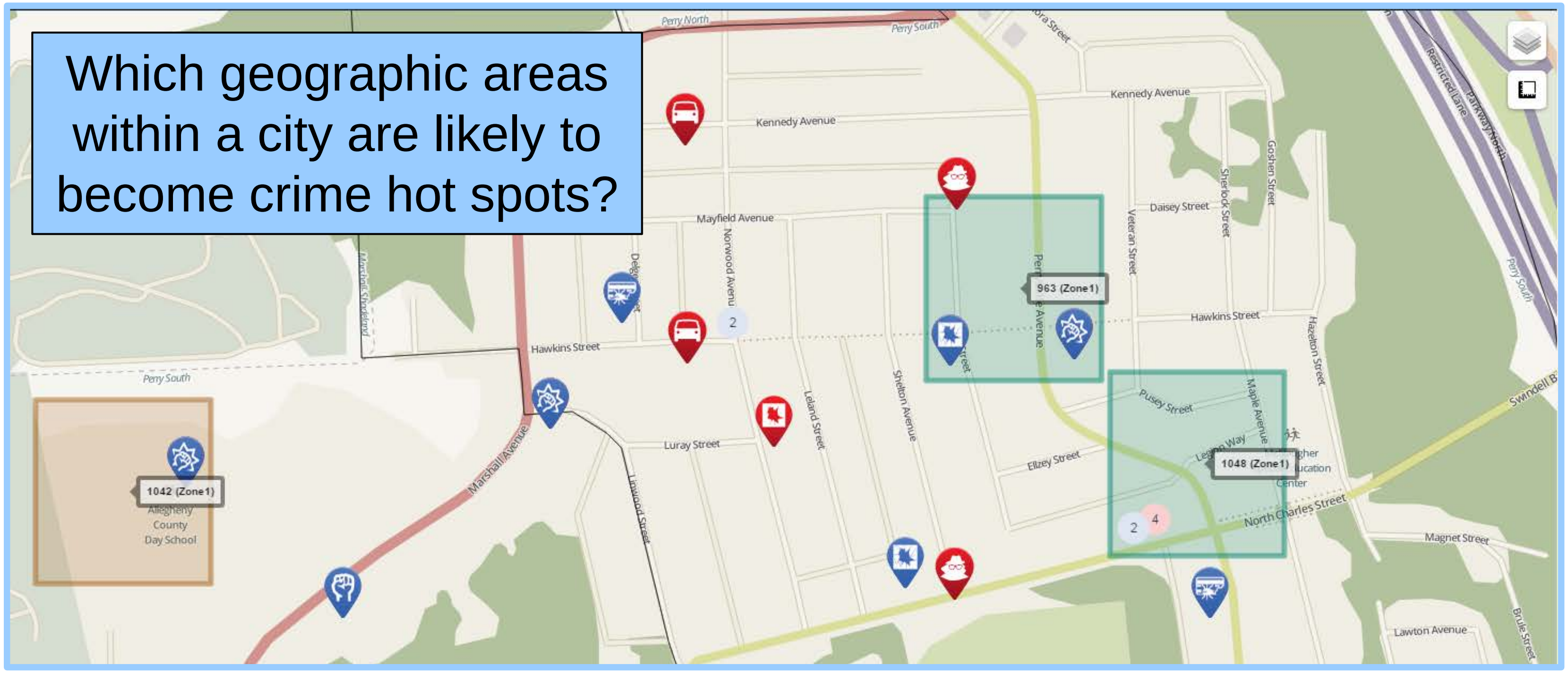
Can identify and fix underlying problems (e.g., hazardous conditions, or systematic bias and discrimination) which can result in legal **liability**.

Case studies: what types of risk does our work address?

Crime prediction and prevention (theft, assault, …) → property, health, life
Catastrophic **events** (natural disaster, disease pandemic, …) → property, health, life
Emerging patterns of **behavioral** risks (opioid overdoses, …) → health, workforce
Occupational risks (safety, training, external hazards, …) → workers comp, liability
Suboptimal **healthcare** practices (procedures, meds, …) → health, malpractice/liability
Environmental risks (poor housing, air/water quality, …) → health, liability
Systematic **bias** and discrimination (human and algorithmic) → liability

Better Models, Better Decisions

Better **predictive models** can lead to improved organizational decisions, allocation of public resources, and quantification of risks.



<u>Short time horizon:</u>
Targeted police patrols in these areas reduce crime.

<u>Longer time horizon:</u>
Quantify risk of loss from theft, assault. Also informs city planning and policy.

Early Warning for Critical Events

Disease pandemic

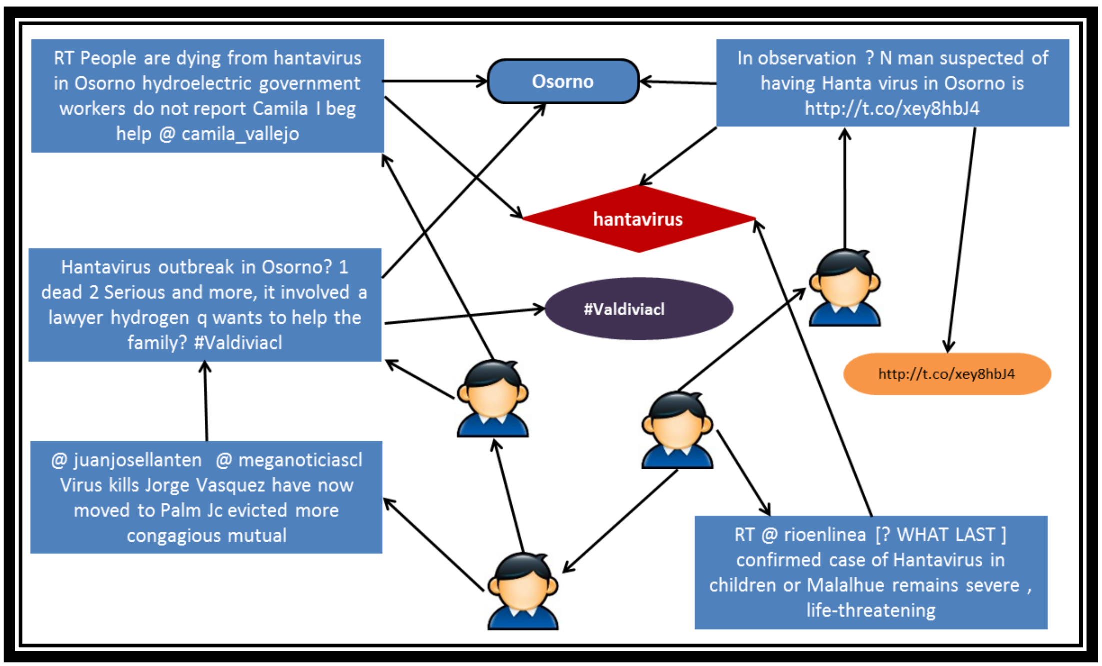
Natural disaster

Terrorist attack

Civil unrest

Early detection can reduce costs to society by enabling a targeted and effective response.

Advance prediction can both quantify and reduce risk.



(Chen and Neill, KDD 2014)

Twitter Event Surveillance

Accurately predicts **civil unrest, ~**1 week in advance

Enables earlier detection of emerging disease outbreaks

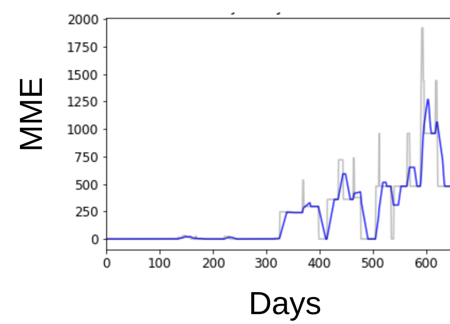
Can identify emerging human rights issues

Opioid and overdose surveillance

Identifying emerging patterns of **behavioral risk factors** and health outcomes can help to assess risks and target behavioral health interventions.

Example: opioid misuse $\nRightarrow$ fatal drug overdoses.

Loss mitigation by: prevention, treatment, rescue, recovery
(Where to target? Who to target? With which interventions?)



Individual level: using data from state prescription drug monitoring programs, we track individuals' total opioid prescription (milligrams morphine equivalent) over time, and correlate with overdoses and "red flag" risk behaviors.

Subpopulation level: we identify emerging clusters of overdoses among subpopulations defined by any combination of geography, demographics, and behaviors.

"elderly African-American males living downtown"

"individuals combining methadone
with the prescription drug Xanax"

Health surveillance for emerging events

Typical health surveillance systems group cases into known categories (e.g., flu-like illness) and identify clusters of disease in space and time.

But what about <u>newly emerging patterns</u> (e.g., a novel outbreak with an unusual combination of symptoms) that they were not already looking for?

Our **pre-syndromic surveillance** system monitors free-text chief complaints from clinics and hospital EDs, combining text analysis and cluster detection to find novel outbreaks and other emerging trends of interest to public health.

<u>Monitoring data from NYC Emergency Departments:</u>
Novel outbreaks (e.g., "nosebleed and stomach pain")
Disaster prep (acute injuries → mental health → chronic)
Unusual health events (e.g., "glass in coffee")

<u>Occupational health (NYC Dept. of Sanitation):</u>
Exposure to illness, external hazards (e.g., needle sticks)
Lapses in safety and training (e.g., lower back injuries)
→ Implications for worker's comp, sick leave mgmt., liability

Identifying causal effects of treatments and exposures

1. Healthcare treatments

We are using health insurance claims data from ~125K individuals to identify anomalous patterns of patient care that impact health outcomes.

* Correct suboptimal care *
* Identify new best practices *

Key idea: treatment effects may be **heterogeneous**; look for most positively and negatively affected subpopulations.

“**Glucocorticoids** significantly increase hospitalizations following treatment in the subpopulation of hypertensive, overweight males with endocrine disorders.”

2. Environmental health

We are using Medicaid data linked to detailed building characteristics in order to identify impacts of poor-quality housing on chronic health.

“Which housing conditions impact which health conditions, for which subpopulations, to what extent?”

Must adjust for known confounders, selection into treatment/exposure.

“**Crowded housing** is associated with increased respiratory conditions & injuries among Asians living in Manhattan.”

Implications for health, liability/malpractice, …

Subset scanning for pattern detection

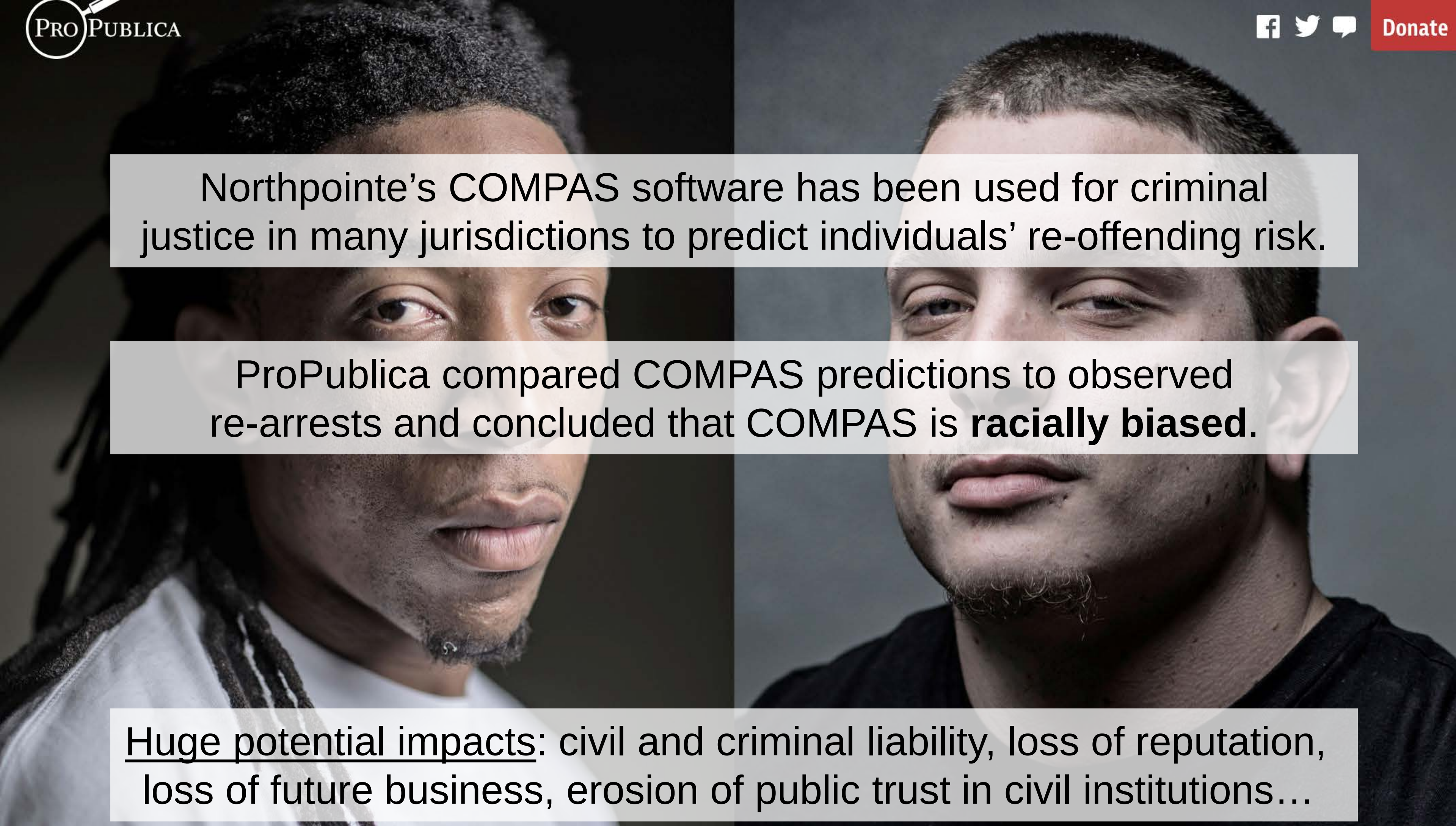
One key methodological idea of our work is **subset scanning**:

- We frame the pattern detection problem as a search over subsets of the data (e.g., spatial areas or subpopulations), maximizing some measure of the "interestingness" or "anomalousness" of a subset.
- This allows us to find **subtle patterns** which typical anomaly detection methods would miss ("needle in the haystack + connect the dots")
- Once we have found the highest scoring subset, we perform a statistical test to see whether it is unlikely to have occurred by chance.

Search over subsets would be computationally infeasible but…

- Our **fast subset scan** algorithm can efficiently identify the most interesting subsets of data records **without** exhaustive search.

This enables us to solve detection problems in milliseconds that would previously have required millions of years!



Northpointe’s COMPAS software has been used for criminal justice in many jurisdictions to predict individuals’ re-offending risk.

ProPublica compared COMPAS predictions to observed re-arrests and concluded that COMPAS is **racially biased**.

Huge potential impacts: civil and criminal liability, loss of reputation, loss of future business, erosion of public trust in civil institutions…

Bernard Parker, left, was rated high risk; Dylan Fugett was rated low risk. (Josh Ritchie for ProPublica)

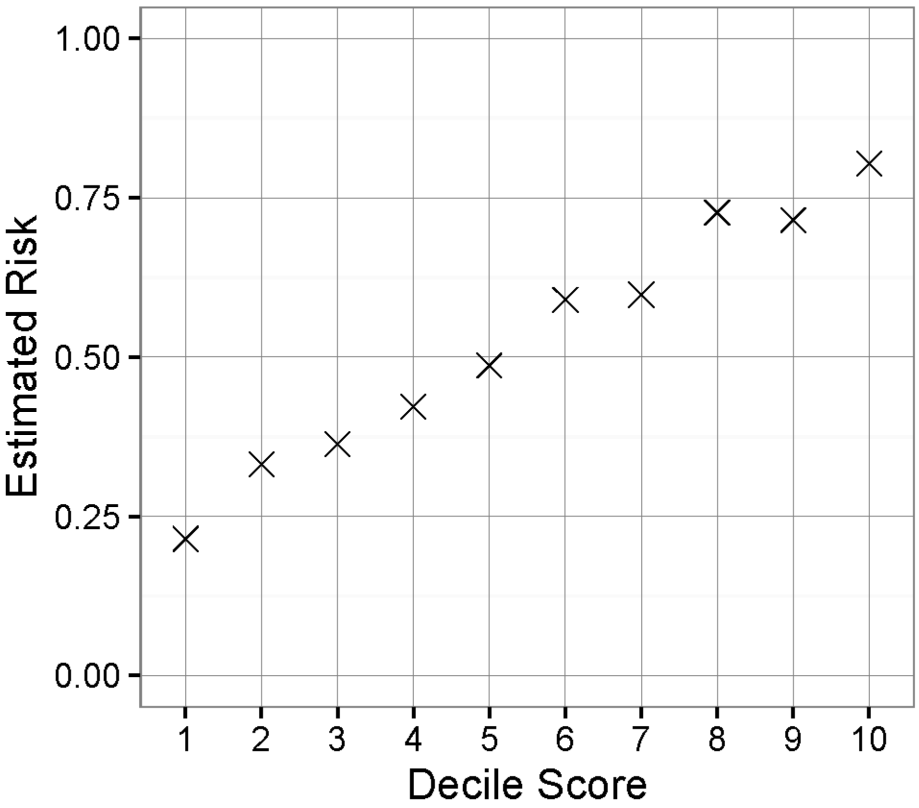
Machine Bias

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

Auditing algorithms for fairness

- Is the COMPAS algorithm for predicting re-offending risk **fair**, or is it **biased** against some subpopulation defined by observed characteristics?
  - **Black box** algorithm. All we observe is predictions vs. gold standard (re-offending) for a sample of individuals (ProPublica data from Broward County, FL).
  - Many possible biases: race, gender, age, past offenses…
  - Combinations of factors, e.g., "elderly white females"
- This led us to develop a new and general approach **("bias scan")** to audit black box algorithms for fairness and to correct systematic biases.
  - Uses fast subset scan to efficiently identify subpopulations for whom a classifier over- or under-estimates risk.

Results of bias scan on COMPAS

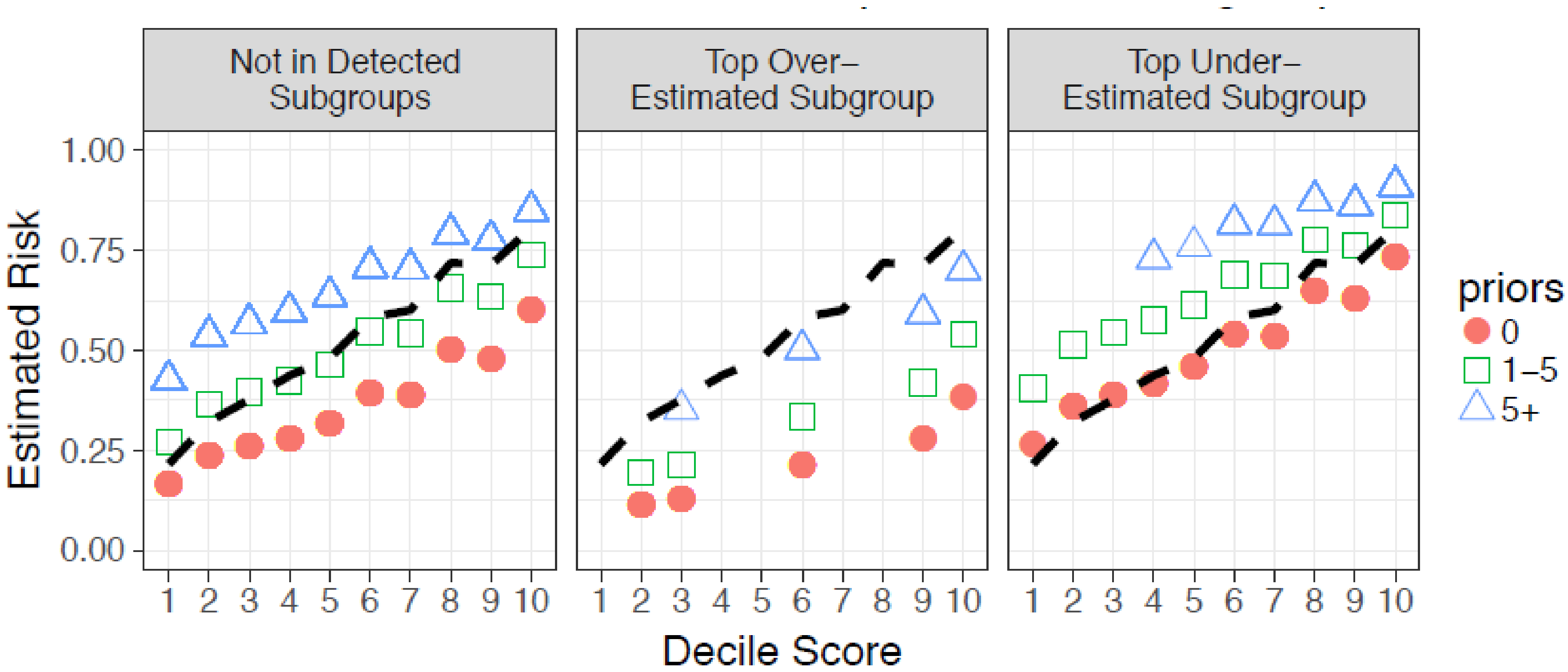


Start with maximum likelihood risk estimates for each COMPAS decile score.

<u>Detection result 1</u>: COMPAS underestimates the importance of prior offenses, overestimating risk for 0 priors, and underestimating risk for 5 or more priors.

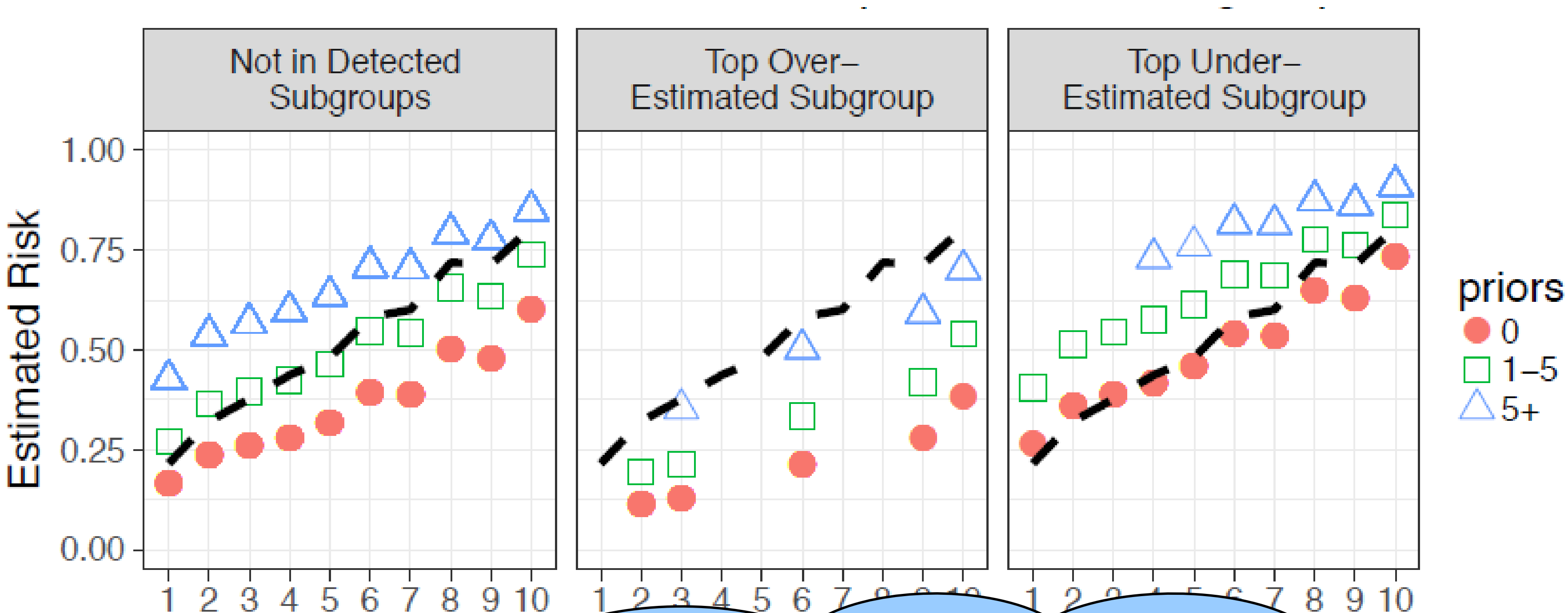
<u>Detection result 2</u>: Even controlling for prior offenses, COMPAS still underestimates risk for males under 25, and overestimates risk for females who committed misdemeanors.

Results of bias scan on COMPAS



After controlling for number of prior offenses and for membership in the two detected subgroups, there are no significant systematic biases in prediction.

Results of bias scan on COMPAS



The resulting probabilistic classifier has greater interpretability (though still based partially on a black box) and is less biased than the original COMPAS predictions… but does this mean it is "fair"?

Discussion: predictive fairness in context

- The method does not account for **target variable bias**: we predict re-offending risk but the gold standard is based on re-arrests not re-offenses.
  - Big problem with drug possession, weapon possession charges. Leads to feedback loops.
- How to avoid **disparate impacts** when making decisions based on even unbiased predictions?

The resulting probabilistic classifier has greater interpretability (though still based partially on a black box) and is less biased than the original COMPAS predictions… but does this mean it is “fair”?

Conclusions

Automated machine learning algorithms have great, mostly untapped potential to benefit the insurance and reinsurance fields:

Underwriting – advance prediction of crisis events, risk estimation

Risk mapping/surveillance – early detection of events, trends, patterns

Loss mitigation – reducing impact through early and targeted response

They inform many types of risk (e.g., events, behavioral, environmental, occupational), with relevance to property, health, liability, workers comp...

But they also create **new risks** of systematic failures and unforeseen consequences, including potential for algorithmic bias and discrimination.

These new risks should be carefully considered and mitigated, ideally through a combination of human oversight and algorithmic auditing.

Auditing approaches also have great potential to assess and mitigate biases in humans' risk-assessment decisions!

Thanks for listening!

More details on my web site:

http://www.cs.nyu.edu/~neill

Or e-mail me at:

daniel.neill@nyu.edu