Penn
University of Pennsylvania

Discussion of *Lessons and Limits of Predicting Shooting Victimization*

Greg Ridgeway

Department of Criminology

Department of Statistics

The Washington Post

Democracy Dies in Darkness

As driverless-car crashes mount, fear of riding in them rises, too

By Ashley Ha

July 26, 2018

Americans
surveys sho
the vehicles

11 traffic fatalities took place over the July 4th holiday weekend in Michigan

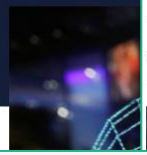
by Mid-Michigan Now Newsroo



SECURITY

Facial recognition leads to first wrongful U.S. arrest, activists say

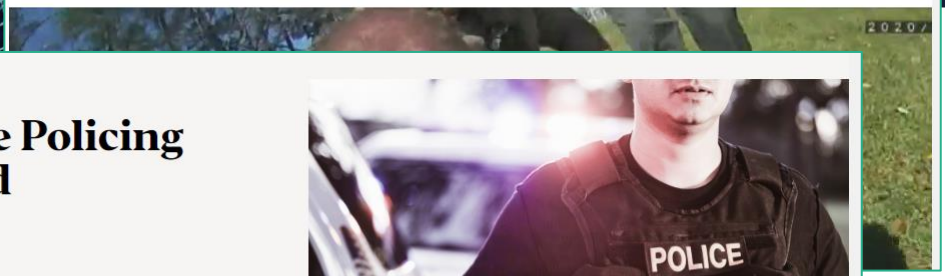
Robert Williams spent over a day in custody in January after face recognition software matched his driver's lic



Police body slam Black man in case of mistaken identity

Brittany Miller

Posted Jun 26, 2020 | 0



EXPLAINER

Predictive Policing Explained

Getty

otect Liberty &
ecurity
Privacy & Free
Expression

Table 1. Successfully predicted crimes under nondeployed conditions.

	ETAS				Analyst					
	Success	Total	Rate	PAI	Success	Total	Rate	PAI	Boost	*P*-value
Maidstone	60	615	9.8%	85.2	42	615	6.8%	59.6	1.4	0.0314
Sevenoaks	39	576	6.8%	55.7	23	576	4.0%	32.8	1.7	0.0409
Southwest	46	986	4.7%	3.5	20	933	2.1%	1.6	2.2	0.0184

Biased Algorithms Are Easier to Fix Than Biased People

Sendhil Mullainathan
New York Times, December 6, 2019

"Changing algorithms is easier than changing people: software on computers can be updated; the 'wetware' in our brains has so far proven much less pliable.

None of this is meant to diminish the pitfalls and care needed in fixing algorithmic bias. But compared with the intransigence of human bias, it does look a great deal simpler."

Collect data, predict, evaluate

- Gather extensive data on 12 million CPD homicide, shooting, victimization, and arrest event records
- 1.2M training, 320K individuals in test set
- 1,300 features containing demographics, arrest and victimization histories, and network features
- Gradient boosting to predict shooting victimization
- Evaluated on calibration, precision (positive predictive value), recall (true positive rate)
- Assess measures of fairness

Proposed prediction model appears to be of high quality

- Model is calibrated
  - Of those predicted to have 5% chance of being a shooting victim, 5% were shot
  - Model is also calibrated within race groups
- Higher predicted risk for black victims among shooting and non-shooting
  - but mathematically impossible to be calibrated and equal predicted risk with imbalance in base rates

Paper undersells the value of the results

- Among top 100 people, 20 shot in 18 months
- Top 1% included ~~"only"~~ 9% of eventual victims
  - Nearly 300,000 cases have low rates
  - But with 300,000 cases with probability 0.005, 1500 will be shot
  - Not cost-effective to target resources to 300,000 people each with low risk

Variable Importance
What is it Doing or What Matters?

- Feature importance describes what this specific model is doing
  - Why is my credit score lower now?
  - Why did I get predicted to be a high recidivism risk?
  - Is a prediction worse if I feature X is randomly replaced?
- Performance reduction after dropping features from model fitting gets at the value of information
  - Does knowing the ZIP code provide better precision?
  - Allows for correlated features to replace lost information

Network data has little predictive value

- Social network research is concluding that friends' victimization is infectious

"every 1 percent increase in exposure to gunshot victims in one's immediate network increases the odds of victimization by roughly 1.1 percent, holding all else constant"

Papachristos, Wildeman, and Roberto (2015). "Tragic, but not random: The social contagion of nonfatal gunshot injuries," *Social Science & Medicine* 125:139-150.

Arrest data capture something important about the risk of serious violent victimization

- Heller et al find that excluding arrest data reduces precision for the top 100 from 20% to 5%
- Papers and popular press argue that arrests are police self-perpetuating their own biases in a feedback loop
- Healthy reminder that arrests have information

Linking Prediction and Prevention

- In a *Criminology & Public Policy* article I wrote:

The best prediction models will be useless unless we have effective prevention strategies to implement in response to predictions

- Chicago PD's Strategic Subject List predicted shooting victimization, but prevention step was limited
- Numerous examples of prediction models outperforming human judgement, but never integrated into resource assignment

Penn
University of Pennsylvania

Discussion of *Lessons and Limits of Predicting Shooting Victimization*

Greg Ridgeway

Department of Criminology

Department of Statistics