Lecture 9: Data ethics continued

Criminology 250

Prof Maria Cuellar

University of Pennslyvania

To predict and serve

- In late 2013, Robert McDaniel – a 22-year-old Black man who lives on the South Side of Chicago – **received an unannounced visit** by a Chicago Police Department commander to warn him not to commit any further crimes.
- The visit **took McDaniel by surprise**. He had not committed a crime, did not have a violent criminal record, and had had no recent contact with law enforcement. So why did the police come knocking?
- It turns out that McDaniel was one of approximately 400 people to have been placed on Chicago Police Department's "heat list". These individuals had all been **forecast to be potentially involved in violent crime**, based on an analysis of geographic location and arrest data.
- The heat list is one of **a growing suite of predictive "Big Data" systems** used in police departments across the USA and in Europe to attempt what was previously thought impossible: to stop crime before it occurs.

https://www.youtube.com/watch?v=lG7DGMgfOb8&ab_channel=MovieclipsClassicTrailers

What is predictive policing?

- According to the RAND Corporation, predictive policing is defined as “the application of analytical techniques – particularly quantitative techniques – to identify likely targets for police intervention and prevent crime or solve past crimes by making statistical predictions”.
- Much like how Amazon and Facebook use consumer data to serve up relevant ads or products to consumers, police departments across the United States and Europe increasingly utilise software from technology companies, such as PredPol, Palantir, HunchLabs, and IBM to identify future offenders, highlight trends in criminal activity, and even forecast the locations of future crimes.

Bias in police-recorded data

- Decades of criminological research, dating to at least the nineteenth century, have shown that **police databases are not a complete census of all criminal offences**, nor do they constitute a representative random sample.
- Empirical evidence suggests that police officers – either implicitly or explicitly – **consider race and ethnicity in their determination** of which persons to detain and search and which neighbourhoods to patrol.
- Bias in police records can also be attributed to levels of **community trust in police**, and the desired amount of local policing – both of which can be expected to vary according to geographic location and the demographic make-up of communities.
- Nevertheless, it is clear that police records do not measure crime. They measure some **complex interaction between criminality, policing strategy, and community–police relations**.

ML algorithms reproduce data with which it's trained

- Machine learning algorithms of the kind predictive policing software relies upon are designed to learn and **reproduce patterns in the data they are given**, regardless of whether the data represents what the model's creators believe or intend.
- Even the best machine learning algorithms trained on police data **will reproduce the patterns and unknown biases in police data**.
- In this sense, predictive policing (see "What is predictive policing?") is aptly named: **it is predicting future policing, not future crime**.
- Because these predictions are likely to over-represent areas that were already known to police, **officers become increasingly likely to patrol these same areas** and observe new criminal acts that confirm their prior beliefs regarding the distributions of criminal activity.

Their approach to finding a "groud truth"

- They combine a demographically representative synthetic population of Oakland, California with survey data from the 2011 National Survey on Drug Use and Health (NSDUH).
- This approach allowed us to obtain high-resolution estimates of illicit drug use from a non-criminal justice, population-based data source which we could then compare with police records.
- In doing so, we find that drug crimes known to police are not a representative sample of all drug crimes.

Drug arrests

Number of drug arrests made by Oakland police department, 2010. (1) West Oakland, (2) International Boulevard.

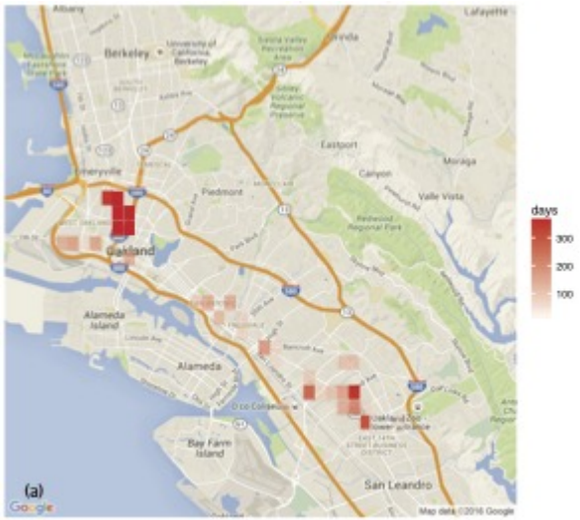
Estimated number of drug users, based on 2011 National Survey on Drug Use and Health.

Evaluate a predictive policing algorithm: PredPol

- Goal: To investigate the effect of police-recorded data on predictive policing models.
- Algorithm: This algorithm was developed by PredPol, one of the largest vendors of predictive policing systems in the USA and one of the few companies to publicly release its algorithm in a peer-reviewed journal.
- Methods: we apply a recently published predictive policing algorithm to the drug crime records in Oakland.

Results

Number of days with targeted policing for drug crimes in areas flagged by PredPol analysis of Oakland police data.



Results

- We find that rather than correcting for the apparent biases in the police data, the model reinforces these biases. The locations that are flagged for targeted policing are those that were, by our estimates, already over-represented in the historical police data.
- Using PredPol in Oakland, black people would be targeted by predictive policing at roughly twice the rate of whites. Individuals classified as a race other than white or black would receive targeted policing at a rate 1.5 times that of whites.

Discussion

- We have demonstrated that predictive policing of drug crimes results in **increasingly disproportionate policing of historically over-policed communities**.
- Although predictive policing is simply reproducing and magnifying the same biases the police have historically held, filtering this decision-making process through sophisticated software that few people understand lends unwarranted legitimacy to biased policing strategies.