

The promise and perils of predictive policing based on big data

November 16 2015, by H V Jagadish



Off to nab a would-be criminal? Credit: Steve Koukoulas, CC BY-NC-ND

Police departments, like everyone else, would like to be more effective while spending less. Given the tremendous attention to [big data](#) in recent years, and the [value it has provided](#) in fields ranging from astronomy to medicine, it should be no surprise that police departments are using data analysis to inform deployment of scarce resources. Enter the era of what is called "[predictive policing](#)."

Some form of predictive policing is likely now in force in a city near you. [Memphis](#) was an [early adopter](#). Cities from [Minneapolis](#) to [Miami](#) have embraced predictive policing. Time magazine named predictive policing (with particular reference to the city of Santa Cruz) one of [the 50 best inventions of 2011](#). New York City Police Commissioner William Bratton recently said that predictive policing is "[the wave of the future](#)."

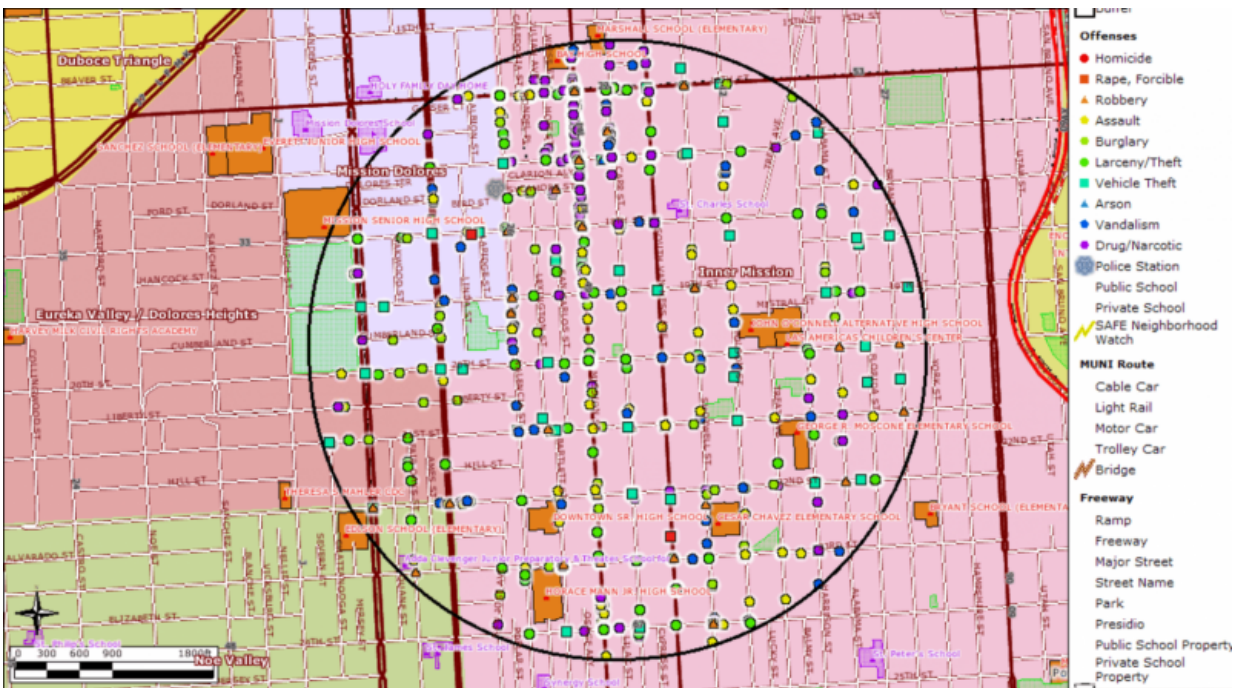
The term "predictive policing" suggests that the [police](#) can anticipate a crime and be there to stop it before it happens and/or apprehend the culprits right away. As the [Los Angeles Times points out](#), it depends on "sophisticated computer analysis of information about previous crimes, to predict where and when crimes will occur."

At a very basic level, it's easy for anyone to read a crime map and identify neighborhoods with higher crime rates. It's also easy to recognize that burglars tend to target businesses at night, when they are unoccupied, and to target homes during the day, when residents are away at work. The challenge is to take a combination of dozens of such factors to determine where crimes are more likely to happen and who is more likely to commit them. Predictive policing algorithms are getting increasingly good at such analysis. Indeed, such was the premise of the movie *Minority Report*, in which the police can arrest and convict murderers before they commit their crime.

Predicting a crime with certainty is something that science fiction can have a field day with. But as a data scientist, I can assure you that in reality we can come nowhere close to certainty, even with advanced technology. To begin with, predictions can be only as good as the input data, and quite often these input data have errors.

But even with perfect, error-free input data and unbiased processing, ultimately what the algorithms are determining are correlations. Even if

we have perfect knowledge of your troubled childhood, your socializing with gang members, your lack of steady employment, your wacko posts on social media and your recent gun purchases, all that the best algorithm can do is to say it is likely, but not certain, that you will commit a violent crime. After all, to believe such predictions as guaranteed is to deny free will.



Predictive policing goes beyond looking only at where crime has already occurred. Credit: Brett Lider, CC BY-SA

Feed in data, get out probabilities

What data can do is give us probabilities, rather than certainty. Good data coupled with good analysis can give us very good estimates of probability. If you sum probabilities over many instances, you can

usually get a robust estimate of the total.

For example, [data analysis](#) can provide a probability that a particular house will be broken into on a particular day based on historical records for similar houses in that neighborhood on similar days. An insurance company may add this up over all days in a year to decide how much to charge for insuring that house.

A police department may add up these probabilities across all houses in a neighborhood to estimate how likely it is that there will be a burglary in that neighborhood. They can then place more officers in neighborhoods with higher probabilities for crime with the idea that police presence may deter crime. This seems like a win all around: less crime and targeted use of police resources. Indeed the statistics, in terms of [reduced crime rates](#), support our intuitive expectations.

Likely doesn't mean definitely

Similar arguments can be used in multiple arenas where we're faced with limited resources. Realistically, customs agents cannot thoroughly search every passenger and every bag. Tax authorities cannot audit every tax return. So they target the "most likely" culprits. But likelihood is very far from certainty: all the authorities know is that the odds are higher. Undoubtedly many innocent individuals are labeled "likely." If you're innocent but get targeted, it can be a big hassle, or worse.

Incorrectly targeted individuals may be inconvenienced by a customs search, but predictive policing can do real harm. Consider the case of Tyrone Brown, recently [reported in The New York Times](#). He was specifically targeted for attention by the Kansas City police because he was friends with known gang members. In other words, the algorithm picked him out as having a higher likelihood of committing a crime based on the company he kept. They told him he was being watched and

would be dealt with severely if he slipped up.



Preventing crime is the goal, but are personal freedoms a casualty? Credit: ibmphoto24, CC BY-NC-ND

The algorithm didn't "make a mistake" in picking out someone like Tyrone Brown. It may have correctly determined that Tyrone was more likely to commit a murder than you or I. But that is very different from saying that he did (or will) kill someone.

Suppose there's a one-in-a-million chance that a typical citizen will commit a murder, but there is a one-in-a-thousand chance that Tyrone

will. That makes him a thousand times as likely to commit a murder as a typical citizen. So it makes sense statistically for the police to focus their attention on him. But don't forget that there is only a one-in-a-thousand chance that he commits a murder. For a thousand such "suspect" Tyrones, there is only one who is a murderer and 999 who are innocent. How much are we willing to inconvenience or harm the 999 to stop the one?

Kansas city is far from being alone in this sort of preemptive contact with citizens identified as "likely to commit crimes." Last year, there was [considerable controversy](#) over a similar program in Chicago.

Balancing crime reduction with civil rights

Such tactics, even if effective in reducing [crime](#), raise [civil liberty concerns](#). Suppose you fit the profile of a bad driver and have accumulated points on your driving record. Consider how you would feel if you had a patrol car follow you every time you got behind the wheel. Even worse, it's likely, even if you're doing your best, that you will make an occasional mistake. For most of us, rolling through a stop sign or driving five miles above the speed limit is usually of little consequence. But since you have a cop following you, you get a ticket for every small offense. In consequence, you end up with an even worse driving record.

Yes, data can help make predictions, and these predictions can help police expend their resources smarter. But we must remember that a probabilistic prediction is not certainty, and explicitly consider the harm to innocent people when we take actions based on probabilities. More broadly speaking, data science can bring us many benefits, but care is required to make sure that it does so [in a fair manner](#).

This story is published courtesy of [The Conversation](#) (under Creative Commons-Attribution/No derivatives).

Source: The Conversation

Citation: The promise and perils of predictive policing based on big data (2015, November 16)
retrieved 30 June 2024 from <https://phys.org/news/2015-11-perils-policing-based-big.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.