

Machine learning Of prediction and policy

Governments have much to gain from applying algorithms to public policy, but controversies loom

Aug 20th 2016 | From the print edition

FOR frazzled teachers struggling to decide what to watch on an evening off, help is at hand. An online streaming service's software predicts what they might enjoy, based on the past choices of similar people. When those same teachers try to work out which children are most at risk of dropping out of school, they get no such aid. But, as Sendhil Mullainathan of Harvard University notes, these types of problem are alike. They require predictions based, implicitly or explicitly, on lots of data. Many areas of policy, he suggests, could do with a dose of machine learning.



Machine-learning systems excel at prediction. A common approach is to train a system by showing it a vast quantity of data on, say, students and their achievements. The software chews through the examples and learns which characteristics are most helpful in predicting whether a student will drop out. Once trained, it can study a different group and accurately pick those at risk. By helping to allocate scarce public funds more accurately, machine learning could save governments significant sums. According to Stephen Goldsmith, a professor at Harvard and a former mayor of Indianapolis, it could also transform almost every sector of public policy.

In hospitals, for instance, doctors try to predict heart attacks so they can act before it is too late. Manual systems correctly predict around 30%. A machine-learning algorithm created by Sriram Somanchi of Carnegie Mellon University and colleagues, and tested on historic data, predicted 80%—four hours in advance of the event, in theory giving time to intervene.

Policing may be helped, too. Last year a policeman in Texas, who had responded to two suicide calls that day, was dispatched to a children's pool party and ended up pulling out his gun. Ideally, the station

would have sent a less stressed officer. Many police chiefs already have a simple system to flag “at risk” officers. No one can be sure that machine learning would have prevented the Texas scare. But a system developed by Rayid Ghani at the University of Chicago and others increases the correctness of at-risk predictions by 12% and reduces the incorrect labelling of officers as being at risk by a third. It is now being used by the Charlotte-Mecklenburg police department in North Carolina.

Chicago’s Department of Public Health is another early adopter. It used to identify children with dangerous levels of lead in their bodies through blood tests and then cleanse their homes of lead paint. Now it tries to spot vulnerable youngsters before they are poisoned. And in India, Microsoft and the state government of Andhra Pradesh are helping farmers choose the best time to sow their seeds. This month, eyeing new government contracts, Microsoft held its first machine-learning and data-science conference in Bangalore.

But the case for code is not always clear-cut. Many American judges are given “risk assessments”, generated by software, which predict the likelihood of a person committing another crime. These are used in bail, parole and (most controversially) sentencing decisions. But this year *ProPublica*, an investigative-journalism group, concluded that in Broward County, Florida, an algorithm wrongly labelled black people as future criminals nearly twice as often as whites. (Northpointe, the algorithm provider, disputes the finding.)

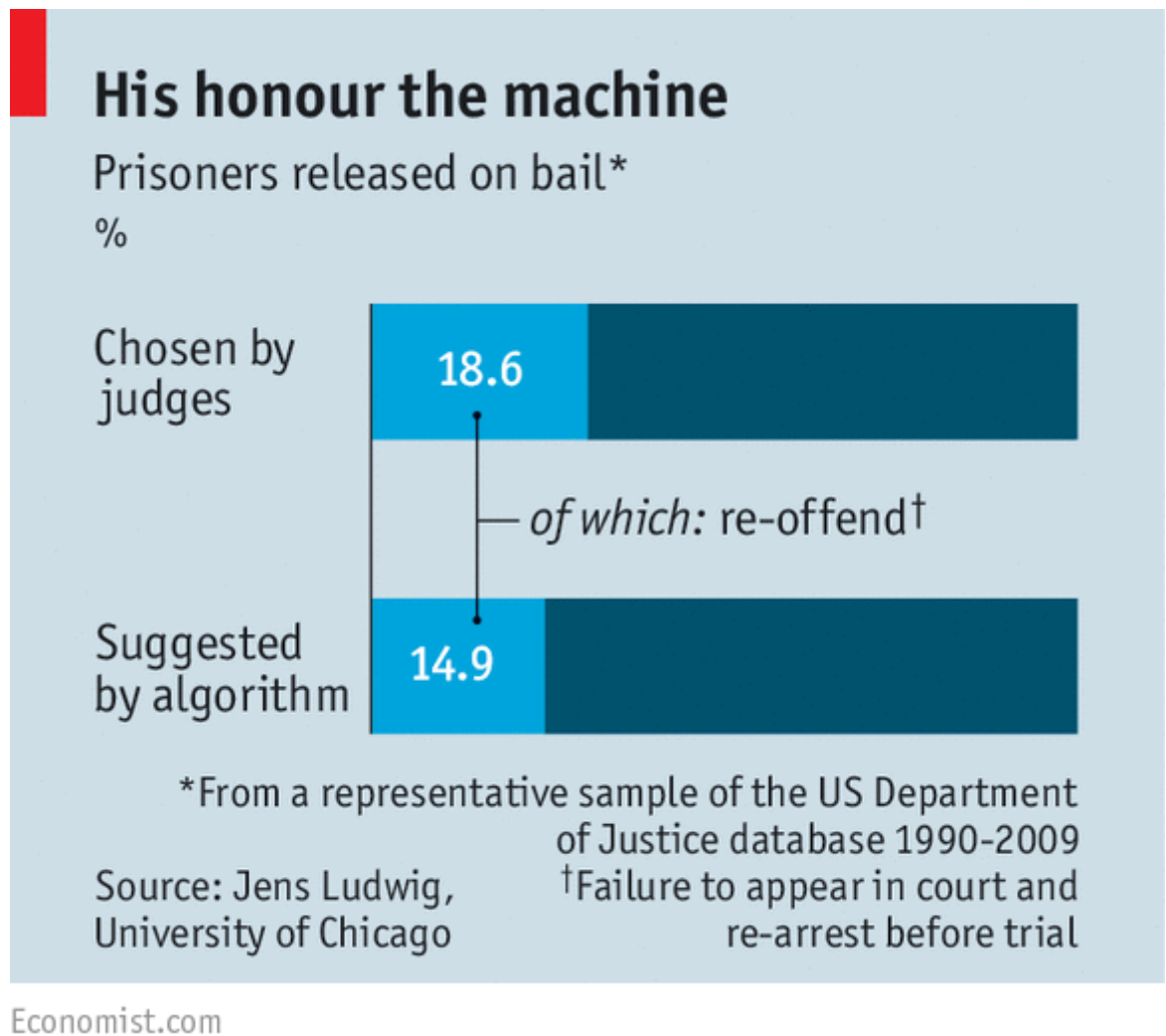
To limit potential bias, Mr Ghani says, avoid prejudice in the training data and set machines the right goals. Machines are trained to find patterns that predict future criminality from past data. They can therefore be told to find patterns that both predict criminality and avoid disproportionate false categorisation of blacks (and others) as future offenders. When a new defendant is tested against these patterns, the risk of racial skewing should be lower.

Bail decisions, in which judges estimate the risk of a prisoner fleeing or offending before trial, seem particularly ripe for help. Jens Ludwig of the University of Chicago and his colleagues claim that their algorithm, tested on a sample of past cases, would have yielded around 20% less crime (see chart), while leaving the number of releases unchanged. A similar reduction nationwide, they suggest, would require an extra 20,000 police officers at a cost of \$2.6 billion. The White House is taking notice. Better bail decisions are a big priority of its Data-Driven Justice Initiative, which 67 states, cities and counties signed in June.

Still, people want to know how decisions that affect them are made. The European Union is considering giving citizens affected by algorithmic decisions the right to an explanation. “Transparency, transparency, transparency” is needed, says Jay Stanley of the American Civil Liberties Union. But private companies may be loth to divulge their special sauce. For Boston’s chief information officer, Jascha Franklin-Hodge, that is a motivation to develop machine learning in-house. Analytical skills, however, are scarce.

Other obstacles may also slow adoption. Getting enough data for a project can be hard. Combining supposedly confidential data sets can heighten the risk of accidentally identifying individuals. Some

applications may be thought unethical. Mr Mullainathan and his colleagues show that machine learning can help predict the risk of death. That could, say, help focus hip replacements on those likely to live longest. Some may think that a step too far.



Prediction is anyway probabilistic, not perfect. Officials still have to act. Getting rid of lead paint may be easy; even with clever algorithms, stopping traumatised policemen from drawing their guns is not. For governments that embrace machine learning, the future will depend on how well they marry its predictive power with old-fashioned human wisdom.

From the print edition: Finance and economics