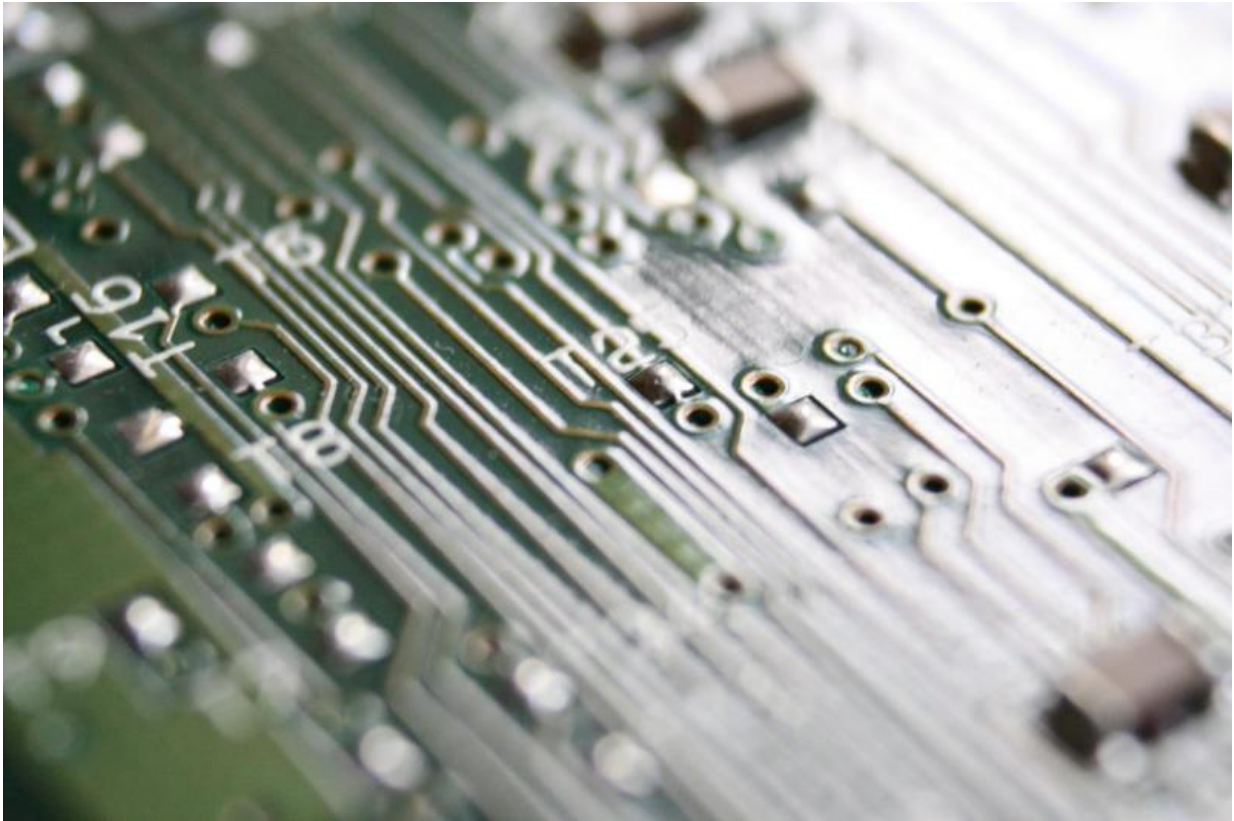


# Team finds new method to improve predictions

November 30 2016

---



Credit: Public Domain

Researchers at Princeton, Columbia and Harvard have created a new method to analyze big data that better predicts outcomes in health care, politics and other fields.

The study appears this week in the journal *Proceedings of the National Academy of Sciences*.

In previous studies, the researchers showed that significant variables might not be predictive and that good predictors might not appear statistically significant. This posed an important question: how can we find highly predictive variables if not through a guideline of statistical significance? Common approaches to prediction include using a significance-based criterion for evaluating variables to use in models and evaluating variables and models simultaneously for prediction using cross-validation or independent test data.

In an effort to reduce the error rate with those methods, the researchers proposed a new measure called the influence score, or I-score, to better measure a variable's ability to predict. They found that the I-score is effective in differentiating between noisy and predictive variables in [big data](#) and can significantly improve the [prediction](#) rate. For example, the I-score improved the [prediction](#) rate in breast cancer data from 70 percent to 92 percent. The I-score can be applied in a variety of fields, including terrorism, civil war, elections and financial markets.

"The practical implications are what drove the project, so they're quite broad," says lead author Adeline Lo, a postdoctoral researcher in Princeton's Department of Politics. "Essentially anytime you might be interested in predicting and identifying highly predictive variables, you might have something to gain by conducting variable selection through a statistic like the I-score, which is related to variable predictivity. That the I-score fares especially well in high dimensional data and with many complex interactions between variables is an extra boon for the researcher or policy expert interested in predicting something with large dimensional data."

**More information:** Adeline Lo et al, Framework for making better

predictions by directly estimating variables' predictivity, *Proceedings of the National Academy of Sciences* (2016). [DOI: 10.1073/pnas.1616647113](https://doi.org/10.1073/pnas.1616647113)

Provided by Princeton University

Citation: Team finds new method to improve predictions (2016, November 30) retrieved 9 April 2024 from <https://phys.org/news/2016-11-team-method.html>

<p>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.</p>
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------