

Objective unfairness: the impact of machine learning on mortgage lending decisions



Written by

[Tarun Ramadorai](#)

Published

30 October 2019

Category

[Finance](#)

Key topics

[Artificial Intelligence](#), [Big Data](#), [Finance](#), [Technology](#)

Technology has a tendency to amplify pre-existing inequities, but how far does machine learning extend these biases to mortgage approvals?

imperialcollege · Magnifying disparities with machine learning

In January 2019, US Congresswoman Alexandria Ocasio-Cortez made headlines when she claimed that algorithms [“have racial inequities that get translated”](#). In this case, she was referring primarily to facial recognition technology, but the point echoed existing concerns over the growing use of machine learning technologies.

By analysing huge datasets and identifying patterns and features within them, machine learning programs can accelerate and improve the accuracy of decisions in such areas as recruitment, marketing, healthcare and finance. The problem is the algorithms that underpin this technology are designed by humans and learn from historical human decision data, which means they risk taking on and, potentially, exacerbating human biases.

Learning from the past

What is undoubtedly true is that decisions made on the basis of algorithmic predictions have the ability to make finer judgements based on more subtle factors than conventional methods.

Take mortgage applications as an example: lenders primarily care about ensuring the loans they make are repaid, so they approve or deny applications (and set interest rates) based on the risk of default. Conventionally, this is done using a statistical model that benchmarks each applicant’s default risk based on income, job security, previous borrowing and a number of other factors.

Machine learning systems take this further, not just comparing applicants against statistical benchmarks but continuously combing vast amounts of historical data to find an applicant’s statistical neighbours (previous applicants with similar material factors).

Decisions can then be made on the basis of how reliable these neighbours were as borrowers. Over time, as the system learns more and receives more data, its decisions should become increasingly accurate in terms of avoiding likely defaulters and charging the ideal interest rate.

Hidden factors

In our research, we tested this by taking a dataset of nine million US mortgage approvals from 2009 to 2013 and tracking them over the next three years. Based on this data, we put together a conventional statistical model of default risk and a more sophisticated machine learning model.

The latter did indeed predict default with greater accuracy, identifying (among other things) marginal applicant “winners”, i.e. individuals who would have been deemed unacceptable credit risks or offered high interest rates by the conventional model, but are understood to be reliable borrowers at attractive rates by the algorithmic approach.

The proportion of winners was around 65 per cent for white and Asian borrowers, but only around 50 per cent for black and Hispanic borrowers

Proportionally, these winners increase across the board, but it’s here that racial inequality enters the discussion: in our model, the proportion of winners was around 65 per cent for white and Asian borrowers, but only around 50 per cent for black and Hispanic borrowers.

What’s driving this disparity? It goes without saying that race was not included as a variable in our research, but it is possible that the algorithm has effectively learned to triangulate race, i.e. to work out a borrower’s likely race based on the other factors included. Equally, it’s possible the well-documented social injustices that certain ethnic groups make them, through no fault of their own, more likely to pose an objective default risk, and the algorithm is picking up on this.

No simple solution

In order to work out which of these causes was at play, we tested the model again, this time including race as a factor. The results were only marginally different,

indicating that triangulation was most likely not the main factor at play – in other words, the algorithm is not inherently racist in the sense of calculating ethnic background and discriminating on the basis of it.

That's not to say we can rule out triangulation as a factor, as it does play a part, but rather it is primarily the flexibility of machine learning in predicting default based on permissible factors that leads to the racial disparity in the proportion of winners.

Can machine learning algorithms morally be relied on?

Of course, this knowledge does not solve the problem. At the end of the day, our results show Ocasio-Cortez's concern was valid: machine learning does reflect existing inequities, even if the cause is not inherent in the technology itself.

A lender genuinely concerned only with maximising financial gain and minimising loss will nonetheless see its approach have an inadvertent negative social effect. What's more, as this is the case, can machine learning algorithms morally be relied on? And what level of regulation is required to mitigate these effects?

These questions will rightly be at the vanguard of research as the technology develops, and we can only hope the knowledge of what needs to change will prove a valuable tool in making that change a reality, leading us into a future in which machine learning can be an equalising force.

This article draws on findings from the paper "[Predictably Unequal? The Effects of Machine Learning on Credit Markets](#)" by Andreas Fuster (Swiss National Bank), Paul Goldsmith-Pinkham (Federal Reserve Banks), Tarun Ramadorai and Ansgar Walther (Imperial College London).

This paper is winner of the [Wharton Research Data Services's Best Paper Award](#) and Best Empirical Finance Paper at the [Western Finance Association conference](#), and the American Finance Association's [Brattle Group Prize](#).

This article was updated on 9 January 2023 to include reference to the Brattle Group Prize.

Written by

[Tarun Ramadorai](#)

Published

30 October 2019

Category

[Finance](#)

Key topics

[Artificial Intelligence](#), [Big Data](#), [Finance](#), [Technology](#)

Main image: SiberianArt / iStock / Getty Images Plus via Getty Images.

[Download PDF](#)

Share



About Tarun Ramadorai

Professor of Financial Economics

Tarun Ramadorai is Professor of Financial Economics at Imperial College London. He has a broad range of research interests in the areas of household finance, asset pricing, and international finance. He is greatly interested in finance and economics issues in emerging markets, with a particular focus on India.

You can find the author's full profile, including publications, at their [Imperial Profile](#)

Monthly newsletter

Receive the latest insights from Imperial College Business School

[Sign up now](#)