DECISION MAKING

Using Algorithms to Understand the Biases in Your Organization

by Jennifer M. Logg

AUGUST 09, 2019

BY WUNDERLAND/GETTY IMAGES

Algorithms have taken a lot of heat recently for producing biased decisions. People are outraged over a recruiting algorithm Amazon was developing that overlooked female job applicants. Likewise, they are outraged over predictive policing and predictive sentencing that disproportionately penalize people of color. Importantly, race and gender were not included as inputs into any of these algorithms.

Should we be outraged by bias reflected in algorithmic output? Yes. But the way organizations respond to their algorithms determines whether they make strides in debiasing their decisions or further perpetuate their biased decision making.

So far, the typical response is for the media to scapegoat the algorithm while the company reverts to human decision making. But this is the wrong approach for identifying and addressing bias. Rather, organizations should use statistical algorithms for the magnifying glasses they are: Algorithms can aggregate individual data points with the purpose of unearthing patterns that people have difficult detecting. When algorithms surface biases, companies should seize on this "failure" as an opportunity to learn when and how bias occurs. This way, they're better equipped to debias their current practices and improve their overall decision making.

The Problem with Blaming Algorithms

Calling algorithms biased anthropomorphizes them. Consider, for example, headlines such as, "Why it's totally unsurprising that Amazon's recruitment AI was biased against women", "Amazon scraps secret AI recruiting tool that showed bias against women," and "Amazon's sexist hiring algorithm could still be better than a human." Even researchers impart agency to algorithms by questioning, "Why Does Artificial Intelligence Discriminate?" and labeling output as "Machine Bias" and "Algorithmic Bias." This simple choice of wording may seem unremarkable, but anthropomorphizing algorithms shifts blame to the tool, ultimately relieving the actual decision makers of their accountability. (In machine learning, "bias" has a different meaning; the problem is with the more colloquial application of the term.)

The actual decision makers are the people who make the hiring decisions. Misdirecting our outrage means that these decision makers, and their managers, are not held accountable for solving the issues unearthed by algorithmic processing. The input data to Amazon's algorithms consisted of historical data: previous hiring decisions made by people; this is where the bias originated and where organizations should focus efforts to debias. Blaming an algorithm for producing biased output is as counterproductive as blaming a mirror for reflecting a bruise on your forehead. Trashing the mirror does not heal the bruise, but it could prolong the time it takes to fix the problem and detect future ones.

Reverting to Human Judgment Doesn't Solve the Problem

When companies scrap these algorithms in response to backlash, they revert to their original, faulty decision-making processes. For most decisions, organizations have historically relied on human judgment. Years of research shows that human judgment is often predictably biased.

Not only are people inconsistent (what researchers consider "low reliability"), but we also get distracted by irrelevant information (that with "low predictive validity"). Take hiring and promotions: even after controlling for gender and age, researchers found that taller people make more money. An inch of height is worth an additional $789 per year of salary. It's unlikely that managers intend to hire or promote based on height, but this information seems to influence their judgments. Additionally, we grow tired as we process more information, increasing the probability that we make these mistakes.

If that weren't enough, the human thought process is also frustratingly opaque. Ask a manager to describe how they recruit high-performing employees and they may explain that they look for "team players." But what exactly does that mean? They may say they look for someone who works well with others. But what information do they look for in a resume or during an interview to signal that? People may rely on subjective criteria to make decisions and not even realize it until they try to explain their thought process. This makes it difficult to create a transparent decision process, making consistency near impossible. That's why it's dangerous to walk away from algorithms in favor of human judgment. It ultimately buries our biases deeper, making them more difficult to detect.

The Case for Algorithms

People get tired and distracted. Algorithms do not. By definition, mathematical equations carry out rules created for them. They remain consistent. This is why even the simplest algorithm, the regression, is often more accurate than experts.

Whereas people often find it difficult to explain their thought processes, most algorithms are transparent — at least to their creator. For simple linear regressions, a person needs to specify how much weight, or importance, each input variable receives in the equation. The equation requires input and output variables which are objective enough to quantify. Thus, numerics introduce transparency to a decision process. (Certain forms of machine learning are exceptions. Though a person decides which dataset is used, the decision-rules used by the trained algorithm aren't easily explainable.)

Of course, there's a legitimate concern about *blindly* following all algorithmic output, regardless of specific circumstances, because algorithms can efficiently compound bias that is present in the input data. An algorithm will magnify any patterns in the input data, so if bias is present, the algorithm will also magnify that bias.

Not surprisingly, this concern is particularly relevant when organizations give little consideration to the data variables used as input. And even more concerning is when organizations fail to put the algorithm through iterations of testing. Human judgment is necessary to assess the accuracy of algorithmic output, and algorithms need feedback to improve. The willingness of an organization to invest in algorithms without including feedback as part of the process has spurred a call for algorithmic auditing.

In fact, Amazon did check the output of its algorithms. And, luckily, they shared their "failure." That output told us something surprisingly specific about how bias infiltrated the company's hiring processes. Amazon utilized 500 models to identify which cues predicted success, defined as whether someone was hired at the company. In discovering the existence of bias, the company also uncovered clues as to where it originated. Certain words in people's resumes were associated with getting hired - verbs expressing confidence and describing how tasks were carried out, including "executed" and "captured." Most of the applicants who used those words happened to be male; statistically, those cues were correlated with gender.

This takeaway allowed Amazon to pinpoint bias in their past hiring decisions. Hiring managers were likely unaware that this particular language influenced them. Or perhaps they did perceive such language as a signal of someone's confidence. They may have relied more on that than other information in the resume, thinking that confidence is a more useful indicator of competence than it actually is.

Discovering this kind of association allows a company to improve its current hiring practices. For instance, Amazon can redact these irrelevant words on resumes before they are reviewed if they know that they are associated with gender and are otherwise not informative. Additionally, programmers can statistically account for this wording so that an algorithm does not use it as a predictive cue.

Using Algorithms as Magnifying Glasses

Organizations can use algorithms to purposely magnify potential biases in order to identify and address them. Detection is the first step in fixing the issue. When algorithms surface biases, companies learn about their past decision processes, what drives biases, and which irrelevant information distracts us from useful information. Companies can apply this magnifying glass strategy to any important decision process that involves predictions, from hiring to promotions.

Leveraging algorithms as magnifying glasses can save organizations time. For instance, if a department hires two people each year, it may take a while to realize that the department of ten consistently only includes one woman. But when an algorithm aggregates infrequent decisions, it finds patterns we wouldn't have seen for years. Making bias glaringly obvious gives organizations the opportunity to address the problem. The alternative is that organizations continue business as usual, letting bias seep into virtually every hire or promotion.

A Checklist for Leveraging Algorithms as Magnifying Glasses

Preparing to Build the Algorithm

Team: How diverse is the team building the algorithm? Diversity in thinking is key to unearthing our own blind spots.

Output: Have we clarified and quantified our goals for making our prediction? Have we specified what we want to predict? Statistics focuses on one dependent variable at the expense of others, and algorithms don't understand trade-offs.

Once biases are detected, organizations can correct biased decisions in three main ways. The first may be the most difficult. It involves creating better input data for the algorithm, which starts with changing current hiring practices. Second, we can continue to use the same historical data but create new rules for the algorithm, such as including a variable that specifies diversity. Third, we can examine how existing input variables may introduce bias or consider new, more appropriate input variables.

Ask "What's the Alternative?"

Input: How subjective is the information we will use to make our prediction? If your input variables are difficult to quantify, chances are that your goals are not clear enough. Attempt to better specify your output variable (e.g., break down what it means to be a high performing employee).

Building the Algorithm

Input goals: Did we choose the right data? Rich data should include multiple factors which describe each instance of observation (wide data) that you have good reason to believe are relevant to the prediction. For example, if there possible, ask potential employees the same 10 questions instead of just 4.

Input pitfalls to avoid: Consider the source and demographic makeup of the input data. Is it representative of the population we would like to produce for future decisions? Is it diverse? If not, can we weed out variables that are proxies for demographics? (Remember that geography may correlate with race and socio-economic status.)

Testing: Did we run multiple iterations of testing phases? Feedback is necessary to developing an algorithm, so examine the output produced by the feedback loops. If the output is biased, examine the relationships aggregated by the algorithm to understand why the bias occurs (ie, which specific input produces the biased output). Take time to consider if changes should be made to the algorithm.

Interpreting Output from the Algorithm

No algorithm is perfect. But neither are humans. If we were, we'd know the future. When faced with less than perfect algorithmic output, people may reflexively want to trash it.

During discussions in my class, "The Psychology of Big Data," students read about an algorithm built to predict which students are most likely to drop out of college. The algorithm was accurate about 85% of the time. The discussion centered around whether to trust less-than-perfect results. I encouraged them to consider the alternative when thinking about how much they should trust the algorithm. How well would a person predict the same outcome? Would they even be able to reach 60% accuracy? Compared to a benchmark of 60% accuracy, 85% starts to look much better.

When algorithmic and human accuracy are directly compared, the predictive accuracy of algorithms consistently blows even expert judgment out of the water. That's why we need to consider the alternative to algorithmic judgments. In fact, in my research with colleagues Julia Minson of Harvard University and Don Moore of University of California, Berkeley, we found that experts who ignore the algorithm's advice make less accurate predictions relative to lay people who are willing to follow the advice.

In the end, algorithms are tools. People build them, determine if their output is accurate, and decide when and how to act on that output. Data can provide insights, but people are responsible for the decisions made based on them.

Sanity check: Did we check that the algorithm predicts what we expect it to with a new sample (called “out-of-sample” predictions)?

Audit: Did we check that the output looks unbiased? A separate team should audit the process to question the appropriateness of the output and whether common sense backs up the relationship between the input and output variable. Importantly, the team should consider if there are unaccounted for variables that could explain the output. Finally, have we considered whether potential proxy variables are at play?

Data-Driven Decision Making: What actions do the results suggest we should take? Perhaps the output shed light on assumptions made when the algorithm was built or suggest that changes should be made to the team’s decision-making process. For instance, the output may lead the team to consider excluding or including specific input variables.

Jennifer M. Logg is an Assistant Professor of Management at Georgetown University’s McDonough School of Business. She was previously a Post-Doctoral Fellow at Harvard University and received her Ph.D. from the Haas School of Business at the University of California at Berkeley. Her primary research uses experiments to test how people respond to the increasing prevalence of information produced by algorithms. Her program of research examines how people expect algorithmic and human judgment to differ (research she calls, Theory of Machine, a twist on the classic “theory of mind”).

This article is about DECISION MAKING

⊕ Follow This Topic

Related Topics: Technology | Personnel Policies | Organizational Culture

Loading...

Loading...

Comments

Leave a Comment

Post Comment

0 COMMENTS

Join The Conversation

POSTING GUIDELINES