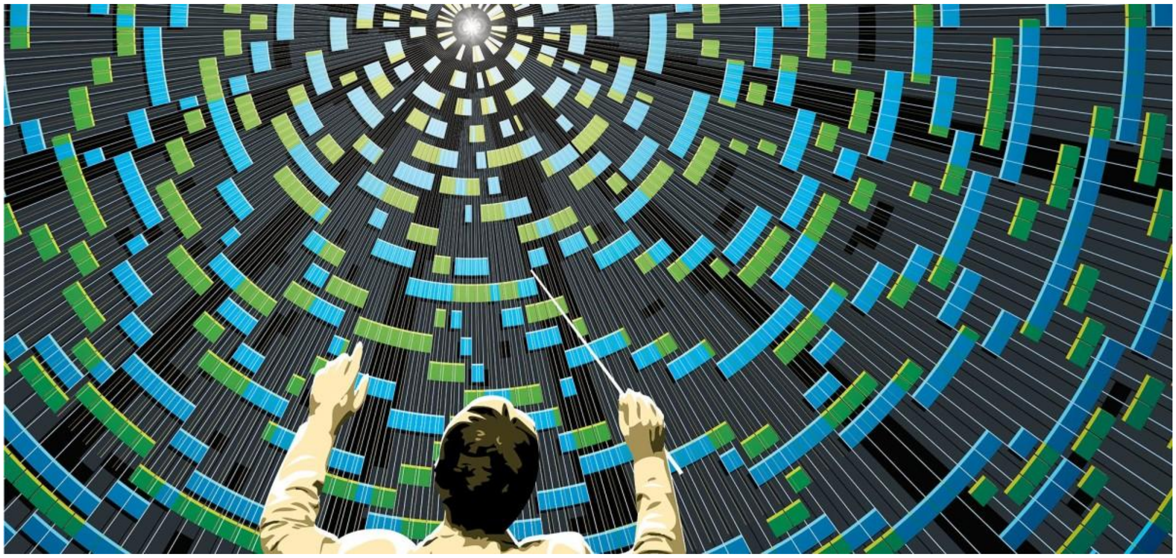
Harvard Business Review



ANALYTICS

Algorithms Need Managers, Too

by Michael Luca, Jon Kleinberg, and Sendhil Mullainathan

FROM THE JANUARY–FEBRUARY 2016 ISSUE

Most managers' jobs involve making predictions. When HR specialists decide whom to hire, they're predicting who will be most effective. When marketers choose which distribution channels to use, they're predicting where a product will sell best. When VCs determine whether to fund a start-up, they're predicting whether it will succeed. To make these and myriad other business predictions, companies today are turning more and more to computer algorithms, which perform step-by-step analytical operations at incredible speed and scale.

Algorithms make predictions more accurate—but they also create risks of their own, especially if we do not understand them. High-profile examples abound. When Netflix ran a million-dollar competition to develop an algorithm that could identify which movies a given user would like, teams of data scientists joined forces and produced a winner. But it was one that applied to DVDs—and as Netflix's viewers transitioned to streaming movies, their preferences shifted in ways that didn't match the algorithm's predictions.

Another example comes from social media. Today many sites deploy algorithms to decide which ads and links to show users. When these algorithms focus too narrowly on maximizing user click-throughs, sites become choked with low-quality "click-bait" articles. Click-through rates rise, but overall customer satisfaction may plummet.

Problems like these aren't inevitable. In our work designing and implementing algorithms and identifying new data sources with a range of organizations, we have seen that the source of difficulty often isn't bugs in the algorithms; it's bugs in the way we interact with them. To avoid missteps, managers need to understand what algorithms do well—what questions they answer and what questions they do not.

Why Do Smart Algorithms Lead Us Astray?

As a growing body of evidence shows, humanizing algorithms makes us more comfortable with them. This can be useful if, for example, you're designing an automated call function. A real person's voice is more likely than an electronic voice to get people to listen. The fundamental problem, however, is that people treat algorithms and the machines that run them the same way they'd treat an employee, supervisor, or colleague. But algorithms behave very differently from humans, in two important ways:

Algorithms are extremely literal.

In the latest Avengers movie, Tony Stark (also known as Iron Man) creates Ultron, an artificial-intelligence defense system tasked with protecting Earth. But Ultron interprets the task literally, concluding that the best way to save Earth is to destroy all humans. In many ways, Ultron behaves like a typical algorithm: It does exactly what it's told—and ignores every other consideration. We get into trouble when we don't manage algorithms carefully.

The social media sites that were suddenly swamped with click-bait fell into a similar trap. Their overall goal was clear: Provide content that would be most appealing and engaging to users. In communicating it to the algorithm, they came up with a set of instructions that seemed like a good proxy—find items that users will click on the most. And it's not a bad proxy: People typically click on content because it interests them. But making selections solely on the basis of clicks quickly filled sites with superficial and offensive material that hurt their reputation. A human would understand that the sites' designers meant "Maximize quality as measured by clicks," not "Maximize clicks even at the expense of quality." An algorithm, on the other hand, understands only what it is explicitly told.

Algorithms are black boxes.

In Shakespeare's *Julius Caesar,* a soothsayer warns Caesar to "beware the ides of March." The recommendation was perfectly clear: Caesar had better watch out. Yet at the same time it was completely incomprehensible. Watch out for what? Why? Caesar, frustrated with the mysterious message, dismissed the soothsayer, declaring, "He is a dreamer; let us leave him." Indeed, the ides of March turned out to be a bad day for the ruler. The problem was that the soothsayer provided *incomplete* information. And there was no clue to what was missing or how important that information was.

Like Shakespeare's soothsayer, algorithms often can predict the future with great accuracy but tell you neither what will cause an event nor why. An algorithm can read through every *New York Times* article and tell you which is most likely to be shared on Twitter without necessarily explaining why people will be moved to tweet about it. An algorithm can tell you which employees are most likely to succeed without identifying which attributes are most important for success.

Recognizing these two limitations of algorithms is the first step to managing them better. Now let's look at other steps you can take to leverage them more successfully.

Be Explicit About All Your Goals

Everyone has objectives and directives, but we also know that the end doesn't always justify the means. We understand that there are soft (often unspoken) goals and trade-offs. We may turn down a little profit today for a gain in reputation tomorrow. We may strive for equality—even if it causes organizational pain in the short term. Algorithms, on the other hand, will pursue a specified objective single-mindedly. The best way to mitigate this is to be crystal clear about everything you want to achieve.

If you care about a soft goal, you need to state it, define it, and quantify how much it matters. To the extent that soft goals are difficult to measure, keep them top of mind when acting on the results from an algorithm.

At Google (which has funded some of our research on other topics), a soft-goal problem emerged with an algorithm that determines which ads to display. Harvard professor Latanya Sweeney unearthed it in a study. She found that when you typed names that were typically African American, like “Latanya Farrell,” into Google, you were shown ads offering to investigate possible arrest records, but not when you searched on names like “Kristen Haring.” Google’s hard goal of maximizing clicks on ads had led to a situation in which its algorithms, refined through feedback over time, were in effect defaming people with certain kinds of names. It happened because people who searched for particular names were more likely to click on arrest records, which led these records to appear even more often, creating a self-reinforcing loop. This probably was not the intended outcome, but without a soft goal in place, there was no mechanism to steer the algorithm away from it.

> Algorithms don’t understand trade-offs; they pursue objectives single-mindedly.

We recently saw the importance of soft goals in action. One of us was working with a West Coast city to improve the efficiency of its restaurant inspections. For decades, the city had been doing them mostly at random but giving more-frequent scrutiny to places with prior violations. Choosing which establishments to inspect is an ideal job for an algorithm, however. Our algorithm found many more variables—not just past violations—to be predictive. The result was that the health department could identify probable offenders more easily and then find actual violations with far fewer inspections.

The officials loved the idea of making the process more efficient and wanted to move toward implementation. We asked if there were any questions or concerns. After an awkward silence, one person raised her hand. “I don’t know how to bring this up,” she said. “But there’s an issue we should discuss.” She explained that in some neighborhoods with tighter quarters, there tended to be more violations. These neighborhoods also happened to be home to higher percentages of minority residents with lower incomes. She did not want these neighborhoods to be excessively targeted by the algorithm. She was expressing a soft goal related to fairness. Our simple solution was to incorporate that objective into the algorithm by setting a ceiling on the number of inspections within each area. This would achieve the hard goal, identifying the restaurants most likely to have problems, while still respecting the soft one, ensuring that poor neighborhoods were not singled out.

Notice the extra step that allowed us to bake in soft goals: giving everyone an opportunity to articulate any concerns. We find that people often formulate soft goals as concerns, so asking for them explicitly facilitates more open and fruitful discussion. It's also critical to give people license to be candid and up-front—to say things that they wouldn't normally. This approach can surface a variety of issues, but the ones we see most commonly relate to fairness and to the handling of sensitive situations.

With a core objective and a list of concerns in hand, the designer of the algorithm can then build trade-offs into it. Often that may mean extending the objective to include multiple outcomes, weighted by importance.

Minimize Myopia

A popular consumer packaged goods company was purchasing products cheaply in China and selling them in the United States. It selected these products after running an algorithm that forecast which ones would sell the most. Sure enough, sales took off and cruised along nicely—until several months later, when customers started to return the items.

As it happens, the surprisingly high and steady return rate could have been predicted (even though the algorithm had failed to foresee it). The company obviously cared about quality, but it hadn't translated that interest into an algorithm that carefully projected consumer satisfaction; instead it had asked the algorithm to focus narrowly on sales. Ultimately, the company's new approach was to become great at forecasting not just how well products would sell but also how much people would enjoy and keep their products. The firm now looks for offerings that customers will rave about on Amazon and other platforms, and the product return rate has plummeted.

This company ran into a common pitfall of dealing with algorithms: Algorithms tend to be myopic. They focus on the data at hand—and that data often pertains to short-term outcomes. There can be a tension between short-term success and long-term profits and broader corporate goals. Humans implicitly understand this; algorithms don't unless you tell them to.

This problem can be solved at the objectivesetting phase by identifying and specifying long-term goals. But when acting on an algorithm's predictions, managers should also adjust for the extent to which the algorithm is consistent with longterm aims.

Myopia is also the underlying weakness of programs that produce low-quality content by seeking to maximize click-throughs. The algorithms are optimizing for a goal that can be measured in the moment—whether a user clicks on a link—without regard to the longer-range and more important goal of keeping users satisfied with their experience on the site.

Nearsightedness can similarly be an issue with marketing campaigns. Consider a run-of-the-mill Gap advertising campaign with Google. It would most likely lead to a spike in visits to Gap.com—because Google's algorithm is good at predicting who will click on an ad. The issue is, the real goal is increasing sales—not increasing website visits. To address this, advertising platforms can collect sales data through a variety of channels, such as partnerships with payment systems, and incorporate it into their algorithms.

> Algorithms focus on the data at hand—which often pertains to short-term outcomes.

What's more, website visits are a short-term behavior, whereas the long-term impact of advertisements includes the downstream effects on brand image and repeat business. While perfect data on such effects is hard to find, careful data audits can help a lot. Managers should systematically list all internal and external data that may be relevant to the project at hand. With a Google campaign, the Gap's marketers could begin by laying out all their objectives—high sales, low returns, good reputation, and so on—and then spell out ways to measure each. Product returns, online reviews, and searches for the term "Gap" would all be great metrics. The best algorithm could then build predictions from a combination of all these features, calibrating for their relative importance.

Choose the Right Data Inputs

Let's return to the example of health departments that are trying to identify restaurants at risk for causing foodborne illness. As mentioned earlier, cities historically have inspected either randomly or on the basis of prior inspection results. Working with Yelp, one of us helped the city of Boston use online reviews to determine which restaurants were most likely to violate local health codes, creating an algorithm that compared the text in reviews with historical inspection data. By applying it, the city identified the same number of violations as usual, but with 40% fewer inspectors—a dramatic increase in efficiency.

This approach worked well not just because we had a lot of restaurants to look at but because Yelp reviews provided a great set of data—something cities hadn't given much thought to. A Yelp review contains many words and a variety of information. The data is also diverse, because it's drawn from different sources. In short, it's quite unlike the inspector-created data cities were accustomed to working with.

When choosing the right data resources, keep in mind the following:

Wider is better.

One trap companies often fall into is thinking of big data as simply a lot of records—for example, looking at one million customers instead of 10,000. But this is only half the picture. Imagine your data organized into a table, with a row for each customer. The number of customers is the length of the table. The amount you know about each customer determines the width—how many features are recorded in each row. And while increasing the length of the data will improve your predictions, the full power of big data comes from gathering wide data. Leveraging comprehensive information is at the heart of prediction. Every additional detail you learn about an outcome is like one more clue, and it can be combined with clues you've already collected. Text documents are a great source of wide data, for instance; each word is a clue.

Diversity matters.

A corollary to this is that data should be diverse, in the sense that the different data sources should be relatively unrelated to one another. This is where extra predictive power comes from. Treat each data set like a recommendation from a friend. If the data sets are too similar, there won't be much marginal gain from each additional one. But if each data set has a unique perspective, a lot more value is created.

Understand the Limitations

Knowing what your algorithm can't tell you is just as important as knowing what it can. It's easy to succumb to the misguided belief that predictions made in one context will apply equally well in another. That's what prevented the 2009 Netflix competition from yielding more benefit to the company: The algorithm that accurately forecast which DVD a person would want to order in the mail wasn't nearly as good at pinpointing which movie a person would want to stream right now. Netflix got useful insights and good publicity from the contest, but the data it collected on DVDs did not apply to streaming.

Algorithms use existing data to make predictions about what might happen with a slightly different setting, population, time, or question. In essence, you're transferring an insight from one context to another. It's a wise practice, therefore, to list the reasons why the algorithm might not be transferable to a new problem and assess their significance. For instance, a health-code violation algorithm based on reviews and violations in Boston may be less effective in Orlando, which has hotter weather and therefore faces different food safety issues.

Also remember that correlation still doesn't mean causation. Suppose that an algorithm predicts that short tweets will get retweeted more often than longer ones. This does not in any way suggest that you should shorten your tweets. This is a prediction, not advice. It works as a prediction because there are many other factors that correlate with short tweets that make them effective. This is also why it fails as advice: Shortening your tweets will not necessarily change those other factors.

Consider the experiences of eBay, which had been advertising through Google for years. EBay saw that people who viewed those ads were more likely to shop at it than people who did not. What it didn't see was whether the advertisements (which were shown millions of times) were causing people to come to its site. After all, the ads were deliberately shown to likely eBay shoppers. To separate correlation from causation, eBay ran a large experiment in which it randomly advertised to some people and not others. The result? It turns out that the advertisements were for the most part useless, because the people who saw them already knew about eBay and would have shopped there anyway.

Algorithms capable of making predictions do not eliminate the need for care when drawing connections between cause and effect; they are not a replacement for controlled experiments. But what they can do is extremely powerful: identifying patterns too subtle to be detected by human observation, and using those patterns to generate accurate insights and inform better decision making. The challenge for us is to understand their risks and limitations and, through effective management, unlock their remarkable potential.

A version of this article appeared in the January–February 2016 issue (pp.96–101) of *Harvard Business Review*.

Michael Luca is the Lee J. Styslinger III Associate Professor of Business Administration at Harvard Business School.

Jon Kleinberg is a professor of computer science at Cornell University and the coauthor of the textbooks *Algorithm Design* (with Éva Tardos) and *Networks, Crowds, and Markets* (with David Easley).

Sendhil Mullainathan is a professor of economics at Harvard University and the coauthor (with Eldar Shafir) of *Scarcity: Why Having Too Little Means So Much*.

8 COMMENTS

Brandon Shelton 3 years ago

I appreciate the message of this article. With all of the seemingly-sustained hype around the power of big data, it's so important to call out straight-forward analytics' limitations. Good "data science" is equal parts science and art.

REPLY 0 0

JOIN THE CONVERSATION