PREDICTIVE ANALYTICS IN PRACTICE

A HARVARD BUSINESS REVIEW INSIGHT CENTER REPORT

Harvard Business Review

Sponsored by

SAS

PREDICTIVE ANALYTICS IN PRACTICE

The HBR Insight Center highlights emerging thinking around today's most important ideas. We'll investigate how to use predictive analytics in decision-making and planning. We'll look at emerging best practices in the field as well as examples of companies that are using analytics in particularly innovative ways.

A PREDICTIVE ANALYTICS PRIMER

BY TOM DAVENPORT

No one has the ability to capture and analyze data from the future. However, there is a way to predict the future using data from the past. It's called predictive analytics, and organizations do it every day.

Has your company, for example, developed a customer lifetime value (CLTV) measure? That's using predictive analytics to determine how much a customer will buy from the company over time. Do you have a "next best offer" or product recommendation capability? That's an analytical prediction of the product or service that your customer is most likely to buy next. Have you made a forecast of next quarter's sales? Used digital marketing models to determine what ad to place on what publisher's site? All of these are forms of predictive analytics.

Predictive analytics are gaining in popularity, but what do you—a manager, not an analyst—really need to know in order to interpret results and make better decisions? How do your data scientists do what they do? By understanding a few basics, you will feel more comfortable working with and communicating with others in your organization about the results and recommendations from predictive analytics. The quantitative analysis isn't magic—but it is normally done with a lot of past data, a little statistical wizardry, and some important assumptions. Let's talk about each of these.

The Data: Lack of good data is the most common barrier to organizations seeking to employ predictive analytics. To make predictions about what customers will buy in the future, for example, you need to have good data on who they are buying (which may require a loyalty program, or at least a lot of analysis of their credit cards), what they have bought in the past, the attributes of those products (attribute-based predictions are often more accurate than the "people who buy this also buy this" type of model), and perhaps some demographic attributes of the customer (age, gender, residential location, socioeconomic status, etc.). If you have multiple channels or customer touchpoints, you need to make sure that they capture data on customer purchases in the same way your previous channels did.

All in all, it's a fairly tough job to create a single customer data warehouse with unique customer IDs on everyone, and all past purchases customers have made through all channels. If you've already done that, you've got an incredible asset for predictive customer analytics.

The Statistics: Regression analysis in its various forms is the primary tool that organizations use for predictive analytics. It works like this in general: An analyst hypothesizes that a set of independent variables (say, gender, income, visits to a website) are statistically correlated with the purchase of a product for a sample of customers. The analyst performs a regression analysis to see just how correlated each variable is; this usually requires some iteration to find the right combination of variables and the best model. Let's say that the analyst succeeds and finds that each variable in the model is important in explaining the product purchase, and together the variables explain a lot of variation in the product's sales. Using that regression equation, the analyst can then use the regression coefficients—the degree to which each variable affects the purchase behavior—to create a score predicting the likelihood of the purchase.

Voila! You have created a predictive model for other customers who weren't in the sample. All you have to do is compute their score, and offer the product to them if their score exceeds a certain level. It's quite likely that the high scoring customers will want to buy the product—assuming the analyst did the statistical work well and that the data were of good quality.

The Assumptions: That brings us to the other key factor in any predictive model—the assumptions that underlie it. Every model has them, and it's important to know what they are and monitor whether they are still true. The big assumption in predictive analytics is that the future will continue to be like the past. As Charles Duhigg describes in his book The Power of Habit, people establish strong patterns of behavior that they usually keep up over time. Sometimes, however, they change those behaviors, and the models that were used to predict them may no longer be valid.

What makes assumptions invalid? The most common reason is time. If your model was created several years ago, it may no longer accurately predict current behavior. The greater the elapsed time, the more likely customer behavior has changed. Some Netflix predictive models, for example, that were created on early Internet users had to be retired because later Internet users were substantially different. The pioneers were more technically-focused and relatively young; later users were essentially everyone.

Another reason a predictive model's assumptions may no longer be valid is if the analyst didn't include a key variable in the model, and that variable has changed substantially over time. The great—and scary—example here is the financial crisis of 2008-9, caused largely by invalid models predicting how likely mortgage customers were to repay their loans. The models didn't include the possibility that housing prices might stop rising, and even that they might fall. When they did start falling, it turned out that the models became poor predictors of mortgage repayment. In essence, the fact that housing prices would always rise was a hidden assumption in the models.

Since faulty or obsolete assumptions can clearly bring down whole banks and even (nearly!) whole economies, it's pretty important that they be carefully examined. Managers should always ask analysts what the key assumptions are, and what would have to happen for them to no longer be valid. And both managers and analysts should continually monitor the world to see if key factors involved in assumptions might have changed over time.

With these fundamentals in mind, here are a few good questions to ask your analysts:

- **Can you tell me something about the source of data you used in your analysis?**
- **Are you sure the sample data are representative of the population?**
- **Are there any outliers in your data distribution? How did they affect the results?**
- **What assumptions are behind your analysis?**
- **Are there any conditions that would make your assumptions invalid?**

Even with those cautions, it's still pretty amazing that we can use analytics to predict the future. All we have to do is gather the right data, do the right type of statistical model, and be careful of our assumptions. Analytical predictions may be harder to generate than those by the late-night television soothsayer Carnac the Magnificent, but they are usually considerably more accurate.

LEARN FROM YOUR ANALYTICS FAILURES

BY MICHAEL SCHRAGE

By far, the safest prediction about the business future of predictive analytics is that more thought and effort will go into prediction than analytics. That's bad news and worse management. Grasping the analytic "hows" and "whys" matters more than the promise of prediction.

In the good old days, of course, predictions were called forecasts and stodgy statisticians would torture their time series and/or molest multivariate analyses to get them. Today, brave new data scientists discipline k-means clusters and random graphs to proffer their predictions. Did I mention they have petabytes more data to play with and process?

While the computational resources and techniques for prediction may be novel and astonishingly powerful, many of the human problems and organizational pathologies appear depressingly familiar. The prediction imperative frequently narrows focus rather than broadens perception. "Predicting the future" can—in the spirit of Dan Ariely's *Predictably Irrational*—unfortunately bring out the worst cognitive impulses in otherwise smart people. The most enduring impact of predictive analytics, I've observed, comes less from quantitatively improving the quality of prediction than from dramatically changing how organizations think about problems and opportunities.

Ironically, the greatest value from predictive analytics typically comes more from their unexpected failures than their anticipated success. In other words, the real influence and insight come from learning exactly how and why your predictions failed. Why? Because it means the assumptions, the data, the model and/or the analyses were wrong in some meaningfully measurable way. The problem—and pathology—is that too many organizations don't know how to learn from analytic failure. They desperately want to make the prediction better instead of better understanding the real business challenges their predictive analytics address. Prediction foolishly becomes the desired destination instead of the introspective journey.

In pre-Big Data days, for example, a hotel chain used some pretty sophisticated mathematics, data mining, and time series analysis to coordinate its yield management pricing and promotion efforts. This ultimately required greater centralization and limiting local operator flexibility and discretion. The forecasting models—which were marvels—mapped out revenues and margins by property and room type. The projections worked fine for about a third of the hotels but were wildly, destructively off for another third. The forensics took weeks; the data were fine. Were competing hotels running unusual promotions that screwed up the model? Nope. For the most part, local managers followed the yield management rules.

Almost five months later, after the year's financials were totally blown and HQ's credibility shot, the most likely explanation materialized: The modeling group—the data scientists of the day—had priced against the hotel group's peer competitors. They hadn't weighted discount hotels into either pricing or room availability. For roughly a quarter of the properties, the result was both lower average occupancy and lower prices per room.

The modeling group had done everything correctly. Top management's belief in its brand value and positioning excluded discounters from their competitive landscape. Think this example atypical or anachronistic? I had a meeting last year with another hotel chain that's now furiously debating whether Airbnb's impact should be incorporated into their yield management equations.

More recently, a major industrial products company made a huge predictive analytics commitment to preventive maintenance to identify and fix key components before they failed and more effectively allocate the firm's limited technical services talent. Halfway through the extensive—and expensive—data collection and analytics review, a couple of the repair people observed that, increasingly, many of the subsystems could be instrumented and remotely monitored in real time. In other words, preventive maintenance could be analyzed and managed as part of a networked system. This completely changed the design direction and the business value potential of the initiative. The value emphasis shifted from preventive maintenance to efficiency management with key customers. Again, the predictive focus initially blurred the larger vision of where the real value could be.

When predictive analytics are done right, the analyses aren't a means to a predictive end; rather, the desired predictions become a means to analytical insight and discovery. We do a better job of analyzing what we really need to analyze and predicting what we really want to predict. Smart organizations want predictive analytic cultures where the analyzed predictions create smarter questions as well as offer statistically meaningful answers. Those cultures quickly and cost-effectively turn predictive failures into analytic successes.

To paraphrase a famous saying in a data science context, the best way to predict the future is to learn from failed predictive analytics.

TO MAKE BETTER DECISIONS, COMBINE DATASETS

BY RITA MCGRATH

A complicated system is somewhat like a complicated recipe.

You know what the outcome will be because you understand what will cause what—combine a given number of ingredients together in a certain way, put them in the oven, and the results will be consistent as long as you repeat the same procedure each time.

In a *complex* system, however, elements can potentially interact in different ways each time because they are interdependent. Take the airline control system—the outcomes it delivers vary tremendously by weather, equipment availability, time of day, and so on.

So being able to predict how increasingly complex systems (as opposed to merely complicated systems) interact with each other is an alluring premise. Predictive analytics increasingly allow us to expand the range of interrelationships we can understand. This in turn gives us a better vantage point into the behavior of the whole system, in turn enabling better strategic decision-making.

This idea is not new, of course. Firms have been developing models that predict how their customers will behave for years. Companies have developed models that indicate which customers are likely to defect, what advertising pitches they will respond to, how likely a debtor is to default (and what can be done to avoid making loans to that person), what will prompt donors to up the ante on their giving, and even who is likely to pay more for services like car insurance. Organizations such as Blue Cross Blue Shield have used their considerable databases about chronically ill people to target and influence their care, reducing to some extent the total cost of care, much of which is concentrated in helping a small portion of their total consumer base.

What is new is that the advent of predictive analytics, in which disparate information that was never before considered as or looked at as related parts of a system, is giving us new ways to see interrelationships across, and think comprehensively about, entire systems. Rather than arguing about what various kinds of activities will drive which outcomes, the questions can now be answered quantitatively. Indeed, as I argued in *Harvard Business Review*, complex systems with their continually changing interrelationships often defy understanding by using conventional means. This in turn creates the opportunity for strategic action.

An example of exactly this kind of action caught my eye in an unlikely setting—city government. New York City Comptroller Scott Stringer, in an effort to help the city reduce its considerable cost of defending against and paying out legal claims made against the city, has turned to predictive analytics to help. The program is called ClaimStat and is modelled after Richard Bratton's famous CompStat program of collecting crime data. The system tracks the incidences that led to the city's paying out $674 million in payments for claims. Stringer's website observes that "These costs are projected to rise over the next four years to $782 million by FY 2018, a figure that is greater than the FY 2015 budget for the Parks Department, Department of Aging, and New York Public Library combined."

Using analytics, the city found a non-obvious systemic relationship—one where the dots may never have been connected otherwise—with costly unintended consequences: In fiscal year 2010, the budget allocated to the Parks and Recreation department for tree pruning was sharply reduced. Following the budget reductions, tree-related injury claims soared, as the Comptroller reports, leading to several multi-million dollar settlements with the public. One settlement actually cost more than the Department's entire budget for tree pruning contracts over a three-year period! Once funding was restored for tree-pruning, claims dropped significantly. Such a relationship might never have been spotted absent the connected database, as the budget for Parks and the budget for lawsuits are managed as separate and unrelated resources. By bringing them together, the system-wide consequences of individual decisions becomes obvious and something that can be tackled in a strategic way.

In the coming years, we can expect to see smart organizations increasingly leveraging the power of multiple databases to get a real vantage point on their strategic challenges.

PREDICT WHAT EMPLOYEES WILL DO WITHOUT FREAKING THEM OUT

BY JOHN BOUDREAU

Imagine one of your managers walks into their subordinate's office and says, "Our data analysis predicts that you will soon get restless and think of leaving us, so we want to make you an offer that our data shows has retained others like you." Would your employees welcome the offer, marveling at the value of your HR analytics? Or, might they see images of Big Brother, and be repelled by a company snooping on the data they generate as they work? Predictive analytics can enable a customized employment value proposition that maximizes mutual benefit for organizations and their talent; but at what point do predictive analytics become too creepy?

For example, predictive analytics can reduce employee turnover costs. In 2009, *The Wall Street Journal* reported on Google's algorithm that crunched data from employee reviews and promotion and pay histories to determine which employees are most likely to quit, and more recently Google was lauded for pioneering the use of big data to predict employee turnover. Laszlo Bock said this helped Google "get inside people's heads even before they know they might leave." This month, Credit Suisse said it calculates who is likely to quit, and proactively offers them new career roles. Will Wolf, the Global Head of Talent Acquisition & Development said that even if employees are not interested in the offered roles, "they are blown away that we're going out of our way to try to find them something interesting and new."

Creepy? Or, perhaps not so much. Yet.

But companies are looking beyond cost savings—to driving outcomes. HR predictive analytics is touted as transforming HR from retrospective and reactive administrative reporting to strategically integrated modeling to predict behaviors, attitudes and capabilities that drive tangible organizational outcomes. Some evidence shows a correlation between HR predictive analytics and organizational performance. Companies like Google are taking this even further. Google is launching a new firm called "Calico"designed to use search tools to improve life expectancy, and it was previously reported that a question considered by the Google People Analytics group was "what if working at Google increased your life span by a year?" In the quest to improve productivity and work life, the information that companies can analyze about you at work is limited only by software.

This insight has produced a common mantra for HR analytics: "to know our employees as well as we know our customers." It's no coincidence that this sounds like consumer marketing. Marketing concepts like brands, segments, value propositions and engagement are fertile metaphors for retooling HR, but there is also a more subtle lesson here.

Marketing often influences consumers through unconscious habits, as described in Charles Duhigg's book, "The Power of Habit." Duhigg describes his own habit of buying a cookie in the company cafeteria at 3:30 p.m. each day. He realized this was a combination of mid-afternoon boredom, and a desire to get away from his desk and to gossip. The cookie was incidental to the actual reward, but that made it no less a culprit in weight gain. Once he realized that, he could break the cookie habit. Suppose predictive analytics found such cookie-eating employees using your data on work schedules and cafeteria purchases, and you shared it with them, to help them be healthier? Would they be delighted or disturbed?

Consider this object lesson from marketing. Pregnancy is an event that changes otherwise stubborn purchasing habits, so retailers want to know about a pregnancy as early as possible. Duhigg's *New York Times* story reports that Target marketing analysts built a predictive algorithm to identify pregnant customers based on their purchasing habits and other demographic information. They sent those customers ads for pregnancy related products. What could be wrong with helping pregnant women be aware of products or services they need, as early as possible?

Apparently, women responded negatively if it was obvious that they received pregnancy ads before they revealed their pregnancy. They responded more positively if they received "an ad for a lawn mower next to diapers." Duhigg reports one executive saying, "as long as a pregnant woman thinks she hasn't been spied on, she'll use the coupons...As long as we don't spook her, it works." Duhigg also reports that Target company executives said the article contained "inaccurate information," so the story may exaggerate, but the lesson remains: Effective predictive analytics depends on how real people react, not just on the elegance of the analytics.

Organization leaders will increasingly confront such situations with their employees, not only their customers. Consider the potential to influence employee behaviors in arenas such as employee benefits, health care and wellness.

In the rush to ask "What *can* HR analytics predict?" perhaps the more vital question is "What *should* HR analytics predict?"

Legal compliance may not be a sufficient answer. A business law journal article, "The Eavesdropping Employer" concludes that "The American legal system's effort to protect employee privacy is... not properly equipped to defend against excessive invasions of privacy that come from increasingly-sophisticated monitoring practices." Appropriate standards may vary across companies and demographic groups. Google employees have said to me, "as long as our data is held and analyzed by our own HR Department, we trust them." Google's employees may be unique because they work for an organization dedicated to changing the world through personal data and analytics. Yet, one study reports that one-third of employees are comfortable sharing personal data with their employer, particularly millennials who will become a larger share of the future workforce. Mark Berry, the Vice President of Human Capital Analytics and Reporting at ConAgra Foods has said, "we want to know our employees as well as our customers," but added that the company has safeguards for types of data that can and cannot be collected.

How should those safeguards be constructed? What is the balance between predictive feasibility and predictive acceptability? These questions require artfully combining analytical rigor with sensitivity and insight into the humanity and ethics of work.

HR is a discipline well-suited to answering these questions, but are HR leaders prepared? Encouraged by constituents, product vendors and compelling stories, HR leaders understandably rush to increase analytic and data skills. Yet, an even more vital and unique role for HR is to help leaders balance what *can* be predicted against what *should* be predicted.

WE CAN'T ALWAYS CONTROL WHAT MAKES US SUCCESSFUL

BY PETER CAPPELLI

The 2002 movie *Minority Report* told the story of a future in which law enforcement could tell who would commit crimes in the future. The police then arrested those people before they could commit the crimes.

A good deal of work in the social sciences tries to do the same thing, albeit without clairvoyance or Tom Cruise. The idea is to identify the attributes of individuals that cause them to act in certain ways in the future: What causes some students to do well in school, why are some patients bad at taking their medicine, and, for our purposes, what causes some candidates to perform well in jobs?

Most of the studies in the workplace have been done by psychologists. Despite the new hype about big data as a means to build such models, psychologists have been studying questions like what predicts who will be a good performer since WWI. Over the generations, we've gotten used to the fact that tests most applicants don't understand, examining attributes such as personality, determine who gets hired.

Of course, there are lots of other tests that are not as well known but sometimes used by HR in the workplace, such as "integrity" tests that try to determine who will steal at work, something very much like the *Minority Report* movie.

But a few things have changed with the rise of big data. The psychologists have lost control of the effort. It's now done by economists, data engineers, IT operatives, and anyone who has access to the data. It also migrated outside the firm to an ever-growing crowd of vendors who offer enticing claims about the benefits of their prediction software. Rather than giving tests, the new studies look for associations with background data. The better ones worry about actual causation.

As data has gotten easier to access and software has made analyzing it simpler, we can examine every aspect of employee behavior. My colleague Monika Hamori and I did a study of what determines whether executives say "yes" when headhunters call to recruit them; in work underway, a colleague recently identified the attributes of individuals who get laid off in a consulting firm based on email traffic; another is looking at the attributes of supervisors that predict which ones do the best job of training. You name it, it's being studied now.

The promise of big data means that we are likely to get better at prediction in the future. Even a small improvement in predictive accuracy can be worth millions to companies that hire tens of thousands of people per year. These tools are especially attractive to retail and service companies because they have so many employees and such high turnover, which means they are hiring all the time.

Here's the issue, which is not new but it has grown more important with the developments above: Many of the attributes that predict good outcomes are not within our control. Some are things we were born with, at least in part, like IQ and personality or where and how we were raised. It is possible that those attributes prevent you from getting a job, of course, but may also prevent you from advancing in a company, put you in the front of the queue for layoffs, and shape a host of other outcomes.

So what, if those predictions are right?

First is the question of fairness. There is an interesting parallel with the court system where predictions of a defendant's risk of committing a crime in the future are in many states used to shape the sentence they will be given. Many of the factors that determine that risk assessment, some of which include things like family background that are beyond the ability of the defendant to control. And there has been pushback: is it fair to use factors that individuals could not control in determining their punishment?

Some of that fairness issue applies to the workplace as well. Even if it does predict who will steal, what if, for example, being raised by a single parent meant that you did not get a job, all other things being equal?

Second is the effect on motivation. If I believe that decisions about my employment such as promotions, layoff decisions, and other outcomes are heavily influenced by factors that I cannot control, such as personality or IQ, how does this affect my willingness to work hard?

And finally, unlike the *Minority Report* movie, our predictions in the workplace are nowhere close to perfect. Many times they are only a bit more accurate than chance, and they explain only a fraction of the differences in behavior of people.

The field of psychology has long thought about the ethical issues and moral consequences of their tests. At least as of yet, the new big data studies and the vendors selling them have not. How we balance the employer's interest in getting better employees and making more effective workplace decisions with broader concerns about fairness and unintended consequences is a pretty hard question.

HOW TO CLONE YOUR BEST DECISION-MAKERS

BY MICHAEL C. MANKINS AND LORI SHERER

Any company's decisions lie on a spectrum. On one end are the small, everyday decisions that add up to a lot of value over time. Amazon, Capital One, and others have already figured out how to automate many of these, like whether to recommend product B to a customer who buys product A or what spending limit is appropriate for customers with certain characteristics.

On the other end of the spectrum are big, infrequent strategic decisions, such as where to locate the next $20 billion manufacturing facility. Companies assemble all the data and technology they can find to help with such decisions, including analytic tools such as Monte Carlo simulations. But the choice ultimately depends on senior executives' judgment.

In the middle of the spectrum, however, lies a vast and largely unexplored territory. These decisions—both relatively frequent and individually important, requiring the exercise of judgment and the application of experience—represent a potential gold mine for the companies that get there first with advanced analytics.

Imagine, for example, a property-and-casualty company that specializes in insuring multinational corporations. For every customer, it might have to make risk-assessment decisions about hundreds of facilities around the world. Armies of underwriters make these decisions, each underwriter more or less experienced and each one weighing and sequencing the dozens of variables differently.

Now imagine that you employ advanced analytics to codify the approach of the best, most experienced underwriters. You build an analytic model that captures their decision logic. The armies of underwriters then use that model in making their decisions. This is not so much crunching data as simulating a human process.

What happens? The need for human knowledge and judgment hasn't disappeared—you still require skilled, experienced employees. But you have changed the game, using machines to replicate best human practice. The decision process now leads to results that are:

- **Generally better.** The incorporation of expert knowledge makes for more accurate, higher-quality decisions.
- **More consistent.** You have reduced the variability of decision outcomes.
- **More scalable.** You can add underwriters as your business grows and bring them up to speed more quickly.

In addition, you have suddenly increased your organization's test-and-learn capability. Every outcome for every insured facility feeds back into the modeling process, so the model gets better and better. So do the decisions that rely on it.

Using analytics in this way is no small matter. You'll find that decision processes are affected. And not only do you need to build the technological capabilities, you'll also need to ensure that your people adopt and use the new tools. The human element can sidetrack otherwise promising experiments.

We know from extensive research that decisions matter. Companies that make better decisions, make them faster, and implement them effectively turn in better financial performance than rivals and peers. Focused application of analytic tools can help companies make better, quicker decisions—particularly in that broad middle range—and improve their performance accordingly.

LET ALGORITHMS DECIDE–AND ACT–FOR YOUR COMPANY

BY BILL FRANKS

In the near future, simply having predictive models that suggest what might be done won't be enough to stay ahead of the competition. Instead, smart organizations are driving analytics to an even deeper level within business processes—to make real-time operational decisions, on a daily basis. These operational analytics are embedded, prescriptive, automated, and run at scale to directly drive business decisions. They not only predict what the next best action is, but also cause the action to happen without human intervention. That may sound radical at first, but it really isn't. In fact, it is simply allowing analytics to follow the same evolution that manufacturing went through during the industrial revolution.

Centuries ago everything was manufactured by hand. If you needed a hammer, for example, someone would manually produce one for you. While manually manufacturing every item on demand allows for precise customization, it doesn't allow for scale or consistency. The industrial revolution enabled the mass production of hammers with consistent quality and lower cost. Certainly, some customization and personal touches were lost. But the advantages of mass production outweigh those losses in most cases. It remains possible to purchase custom made items when the expense is deemed appropriate, but this usually only makes sense in special situations such as when the purchaser desires a one-of-a-kind piece.

The same revolution is happening in analytics. Historically, predictive analytics have been very much an artisanal, customized endeavor. Every model was painstakingly built by an analytics professional like me who put extreme care, precision, and customization into the creation of the model. This led to very powerful, highly-optimized models that were used to predict all sorts of things. However, the cost of such efforts only makes sense for high-value business problems and decisions. What about the myriad lower value decisions that businesses face each day? Is there no way to apply predictive analytics more broadly?

There is.

Operational analytics recognize the need to deploy predictive analytics more broadly, but at a different price point. An assembly line requires giving up customization and beauty in order to achieve an inexpensive, consistent product. So, too, operational analytics require forgoing some analytical power and customization in order to create analytics processes that can increase results in situations where a fully custom predictive model just doesn't make sense. In these cases, it is better to have a very good model that can actually be deployed to drive value than it is to have no model at all because only an optimal model will be accepted.

Let me illustrate the difference with a common example. One popular use of predictive models is to identify the likelihood that a given customer will buy a specific product or respond to a given offer. An organization might have highly robust, customized models in place for its top 10-20 products or offers. However, it isn't cost effective to build models in the traditional way for products or offers that are far down the popularity list. By leveraging the learnings from those 10-20 custom models, it is possible to create an automated process that builds a reasonable model for hundreds or thousands of products or offers rather than just the most common ones. This enables predictive analytics to impact the business more deeply.

Operational analytics are already part of our lives today, whether we realize it or not. Banks run automated algorithms to identify potential fraud, websites customize content in real time, and airlines automatically determine how to re-route passengers when weather delays strike while taking into account myriad factors and constraints. All of these analytics happen rapidly and without human intervention. Of course, the analytics processes had to be designed, developed, tested, and deployed by people. But, once they are turned on, the algorithms take control and drive actions. In addition to simply predicting the best move to make or product to suggest, operational analytics processes take it to the next level by actually prescribing what should be done and then causing that action to occur automatically.

The power and impact of embedded, automated, operational analytics is only starting to be realized, as are the challenges that organizations will face as they evolve and implement such processes. For example, operational analytics don't replace traditional analytics, but rather build upon them. Just as it is still necessary to design, prototype, and test a new product before an assembly line can produce the item at scale, so it is still necessary to design, prototype, and test an analytics process before it can be made operational. Organizations must be proficient with traditional analytics methods before they can evolve to operational analytics. There are no shortcuts.

There are certainly cultural issues to navigate as well. Executives may not be comfortable at first with the prospect of turning over daily decisions to a bunch of algorithms. It will also be necessary to get used to monitoring how an operational analytics process is working by looking at the history of decisions it has made as opposed to approving up front a series of decisions the process is recommending. Pushing through such issues will be a necessary step on the path to success.

The tools, technologies, and methodologies required to build an operational analytics process will also vary somewhat from those used to create traditional batch processes. One driver of these differences is the fact that instead of targeting relatively few (and often strategic) decisions, operational analytics usually target a massive scale of daily, tactical decisions. This makes it necessary to streamline a process so that it can be executed on demand and then take action in the blink of an eye.

Perhaps the hardest part of operational analytics to accept, especially for analytics professionals, is the fact that the goal isn't to find the best or most powerful predictive model like we're used to. When it is affordable and the decisions being made are important enough to warrant it, we'll still put in the effort to find the best model. However, there will be many other cases where using a decent predictive model to improve decision quality is good enough. If an automated process can improve results, then it can be used with confidence. Losing sleep over what additional power could be attained in the process with a lot of customization won't do any good in situations where it just isn't possible due to costs and scale to actually pursue that customization.

If your organization hasn't yet joined the analytics revolution, it is time that it did. Predictive analytics applied in batch to only high value problems will no longer suffice to stay ahead of the competition. It is necessary to evolve to operational analytics processes that are embedded, automated, and prescriptive. Making analytics operational is not optional!

TRACK CUSTOMER ATTITUDES TO PREDICT THEIR BEHAVIORS

BY WERNER REINARTZ AND RAJKUMAR VENKATESAN

CRM is typically all about customer behavior: you track customers' behavior in terms of where, when and in what context they interacted with your company. But the increasing ease with which you can track behavior and the ability to build and maintain extensive behavioral databases has encouraged many marketers to de-emphasize the collection and interpretation of "soft" attitudinal information: that is, data around customer satisfaction, attitudes towards brands, products, and sales persons, and purchase intentions.

The argument is that in-depth behavioral data already encapsulates underlying attitudes, and because decision makers are mainly concerned with customer behavior, there is not much need (any more) to worry about underlying attitudes. There's a similar assumption underlying much of the discussion around how to measure the return on marketing investment, where it seems to be tacitly accepted that attitudinal insights are insufficient at senior decision-making levels, and behavioral insights represent today's benchmarks.

But downplaying attitudinal data seems rather too convenient. After all, it's hard work to capture attitudes. Purchases, customer inquiries, or mailing contacts are collected by firms continuously for all customers through CRM software systems, but attitudinal information rests in the hearts and minds of customers, who have to be explicitly prompted and polled for that information through customer surveys and textual analysis of customer reviews and online chatter. What's more, some customers might not want to give that information, even if firms wanted to collect it.

Bottom line, you can maybe hope to get strong attitudinal information about a few customers, but it is unrealistic that you can get it about a lot of them—and you certainly can't be polling everyone all that often just to get information from a possibly unrepresentative subset. Much easier, therefore, to pretend that attitudes are just not that important.

This is actually a cop-out. In fact, respectable analytic techniques exist that allow to you *impute* attitudes from a small group about which you have complete information (attitudes, behavior, demographics) to a larger group where the attitudinal information is missing, and then test whether the imputation of those attitudes produces better predictions of the larger group's subsequent behavior (which you are tracking all the time).

Basically, what you do is analyze the relationships between attitudes, behavior, and demographics for customers in the small group so that you can express attitude as a derivative of the other observable factors: a male customer who is X years old and does Y will have Z attitude. You assign customers in the larger group with the attitudes that their behavior and demographics imply, using the relationships derived from the small group analysis. You then make predictions about their future behavior, which you can compare to the predictions you make on the basis of demographics and past behavior only.

We tested the approach with a company in the pharma industry. Our large dataset included the prescription history of more than six thousand physicians for a leading cardiovascular drug over 45 contiguous months. Physicians were surveyed on their attitudes toward the main drugs in the relevant therapeutic category, as well as their attitudes towards the firm's salespeople. The survey asked the doctors, for example, to rate the product's performance and to assess to what level they agreed or disagreed with statements made by the firm's salespeople during sales calls in light of their experience with the drug. Our goal was to explore how the pharmaceutical firm's customer lifetime value (CLV), customer retention, and sales were affected by the physicians' experience of the drug coupled with their attitude regarding the salespeoples' credibility and knowledge.

The results were startling. We found that for this company, a $1 million investment in collecting customer attitudes would deliver projected annual returns of more than 300% from providing more accurate behavioral predictions. It also revealed that attitude information for mid-tier customers (in terms of future profit potential) would produce the highest relative benefit. In other words, incorporating attitudes provides a forward-looking measure that helps to discriminate between the customers that will likely contribute to increasing profitability and those whose profitability will likely decline. In this case, it appeared that the firm was overspending on top-tier customers with regard to their CRM campaigns and that it could improve the ROI from CRM by rebalancing resources across top-tier and mid-tier customers.

Of course, there is no guarantee that the inclusion of customer attitude information in predictive CRM modeling will always yield returns. But our findings do make a very strong case that firms should explore avenues for tracking customer attitudes and to assess their predictive potential in order to adjust CRM strategies accordingly.

YOUR COMPANY'S ENERGY DATA IS AN UNTAPPED RESOURCE

BY ROB DAY

Most companies are unprepared for the emerging revolution in predictive energy analytics. In fact, many readers' eyes will have already glazed over at the preceding sentence, with the natural initial reaction that energy-related data isn't relevant to their jobs.

But what happens when every single light fixture in all of your company's facilities becomes a networked mini-computer with an array of sensors? Who at your company will be put in charge of turning buildings operations from a cost center to a revenue center? These examples are not hypothetical capabilities; these are now real options for companies. And yet few corporate managers are asking such questions, much less taking advantage.

Cost Savings

Chances are, energy-related spending has a significant impact on your company's profitability. There are over five million commercial and industrial facilities in the U.S. alone, according to the US. Energy Information Administration, with a combined annual energy cost of over $200 billion. The U.S. EPA estimates that around 30% of that energy is used inefficiently or unnecessarily. And many companies also face additional energy-related costs from their commercial vehicles, of which there are over 12 million in operation in the U.S. according to IHS, incurring fuel costs in the billions annually.

So there are some big potential savings out there to be gained, but for most companies the responsibility for capturing them is relegated to facilities and fleet managers. Furthermore, many of these managers are focused more on productivity and safety goals than energy savings, nor are they allocated budgets to acquire new energy-saving systems even when paybacks would be compelling. And of course, few such managers have a background in information technology.

But as computing and networking costs have fallen over the past few decades, it has opened up a host of new ways that data and IT could be applied to drive significant cost savings in company-owned buildings and vehicle fleets. Startups like First Fuel and Retroficiency are able to perform "virtual energy audits" by combining energy meter data with other basic data about a building (age, location, etc.) to analyze and identify potential energy savings opportunities. Many Fortune 500 companies have also invested in "energy dashboards" such as those offered by Gridium and EnerNOC, among numerous others, which give them an ongoing look at where energy is being consumed in their buildings, and thus predict ways to reduce usage.

Many companies use telematics (IT for vehicles) to track their fleets for safety and operational purposes, and some startups are now using these capabilities to also help drive fuel savings. XLHybrids, for instance, not only retrofits delivery vehicles with hybrid drivetrains for direct fuel savings, they also provide remote analysis to help predict better driving patterns to further reduce fuel consumption. Transportation giants like FedEx and UPS already use software-based optimization of fleet routes with cost savings in mind.

Operational Improvements

The benefits of tracking energy usage aren't limited just to energy savings. Because energy usage is an integral part of all corporate facilities and operations, the data can be repurposed for other operational improvements.

Take lighting, for example. Boston-based Digital Lumens offers fixtures for commercial and industrial buildings that take advantage of the inherent controllability of solid-state lighting, by embedding intelligence and sensors and adjusting consumption based upon daylight levels, occupancy, and other inputs to drive energy savings of 90% or more. But along the way to achieving these direct energy cost reductions, many of their customers find additional benefits from having a network of data-gathering mini-computers all over their facilities. For example, manufacturers and warehouse operators who've installed Digital Lumens systems have the ability to generate "heat maps" showing which locations in their facilities get the most traffic, which allows the facilities managers to reposition equipment or goods so that less time is wasted by workers moving around unnecessarily. And now retailers are starting to leverage the same information to better position higher-margin product where traffic is highest within their stores.

Another use of energy data is in predictive maintenance. When a critical piece of equipment breaks in a commercial setting, it can have a significant financial impact. If the refrigerator compressor breaks in a restaurant, for instance, it can force a halt to operations of the entire facility. But often, long before such equipment fully stops working, the early signs of a problem can be discerned in its energy usage signal. Startups like Powerhouse Dynamics and Panoramic Power are finding that their small-commercial customers get as much value out of such fault-detection and predictive maintenance as the get out of the overall energy monitoring services their systems are designed to provide.

Don't have a capital budget for energy savings projects? Well, other companies like SCIenergy and Noesis are now using predictive analytics to help underwrite energy-efficiency loans and even more creative financing which helps companies capture savings from day one, in some cases even guaranteeing system performance.

New Sources of Revenue

What really has the potential to radically change how corporate managers view predictive energy analytics, however, is how it can be used to turn existing "cost centers" into sources of new, high-margin revenue.

Electric utilities must keep the grid balanced at all times, and this challenge is only growing more acute. They can expensively purchase power from other sources at times of high demand, but it's often better for them to avoid such peaks by reducing consumption when needed. Thus, many such utilities are willing to pay commercial customers to participate in so-called "demand response" or "frequency regulation" programs in which customers periodically reduce their electricity usage so the utility doesn't have to bring another power plant online.

Imagine a big box retail store in the future: It has solar panels on the roof. A large-scale battery in the basement. Plus an intelligent load-control software system that deploys the battery's power as needed, and also adjusts the air conditioning, lighting, and other energy-consuming devices in the building in incremental ways so that when such loads are shifted around minute to minute, no one in the building feels any impact on comfort or operations. The combination of these systems would not only reduce the facility's bill from the local electric utility, it would also enable the building to automatically participate in that utility's demand response program and generate revenue.

Does this sound like a pipe dream? Seattle-based Powerit Solutions offers such intelligent automation today, and they already control 800 megawatts of load in the marketplace.

Unfortunately, most corporations aren't making the necessary investments in energy data analytics—they're not providing budgets or the cross-functional teams to identify the available cost savings, much less the new revenue opportunities. To be done right, integrating such solutions into the enterprise requires not just knowledge about buildings, but also IT and financial leadership. The effective "facilities management" team of the future will have all of these capabilities. Leading companies across all industries will have to start viewing energy data analytics as a core shareholder value activity, prioritizing it accordingly.

(Disclosure: Black Coral Capital, where I am a partner, is an investor in Digital Lumens, Noesis, and Powerit.)

HOW CITIES ARE USING ANALYTICS TO IMPROVE PUBLIC HEALTH

BY BECHARA CHOUCAIR, JAY BHATT AND RAED MANSOUR

From clean water supplies to the polio vaccine, the most effective public health interventions are typically preventative policies that help stop a crisis before it starts. But predicting the next public health crisis has historically been a challenge, and even interventions like chlorinating water or distributing a vaccine are in many ways reactive. Thanks to predictive analytics, we are piloting new ways to predict public health challenges, so we can intervene and stop them before they ever begin.

We can use predictive analytics to leverage seemingly unrelated data to predict who is most susceptible to birth complications or chronic diseases or where and when a virulent outbreak is most likely to occur. With this information, public health officials should be able to respond before the issue manifests itself – providing the right prenatal treatments to mitigate birth complications, identifying those most likely to be exposed to lead or finding food establishments most at risk for violations. With this information, data becomes actionable. Predictive analytics has the potential to transform both how government operates and how resources are allocated, thereby improving the public's health.

While the greatest benefits have yet to be realized, at the Chicago Department of Public Health (CDPH), we are already leveraging data and history to make smarter, more targeted decisions. Today, we are piloting predictive analytic models within our food protection, tobacco control policy, and lead inspection programs.

Recently, CDPH and the Department of Innovation and Technology engaged with local partners to identify various data related to food establishments and their locations – building code violations, sourcing of food, registered complaints, lighting in the alley behind the food establishment, near-by construction, social media reports, sanitation code violations, neighborhood population density, complaint histories of other establishments with the same owner and more.

The model produced a risk score for every food establishment, with higher risk scores associated with a greater likelihood of identifying critical violations. Based on the results of our pilot and additional stakeholder input, we are evaluating the model and continue to make adjustments as needed. Once it is proven successful, we plan to utilize the model to help prioritize our inspections, and by doing so, help improve food safety.

To be clear, this new system is not replacing our current program. We continue to inspect every food establishment following our current schedule, ensuring the entire food supply remains safe and healthy for our residents and tourists. But predictive analytics is allowing us to better concentrate our efforts on those establishments more likely to have challenges. In time, this system will help us work more closely with restaurateurs so they can improve their business and decrease complaints. In short, businesses and their customers will both be happier and healthier.

Building on the work of the food protection predictive model, we developed another key partnership with the Eric & Wendy Schmidt Data Science for Social Good Fellowship at University of Chicago (DSSG) to develop a model to improve our lead inspection program.

Exposure to lead can seriously affect a child's health, causing brain and neurological injury, slowed growth and development, and hearing and speech difficulties. The consequence of these health effects can be seen in educational attainment where learning and behavior problems are often the cause of lower IQ, attention deficit and school underperformance. Furthermore, we've seen a decrease in federal funding over the past several years for our inspectors to go out and identify homes with lead based paint and clearing them. But thanks to data science, we are now engaging on a project where we can apply predictive analytics to identify which homes are most likely to have the greatest risk of causing lead poisoning in children – based on home inspection records, assessor value, past history of blood lead level testing, census data and more.

Predictive models may help determine the allocation of resources and prioritize home inspections in high lead poisoning risk areas (an active approach), instead of waiting for reports of children's elevated blood lead levels to trigger an inspection (the current passive approach). An active predictive approach shortens the amount of time and money spent in mitigation by concentrating efforts on those homes that have the greatest risk of causing lead poisoning in children.

Incorporating predictive models into the electronic medical record interface will serve to alert health care providers of lead poisoning risk levels to their pediatric and pregnant patient populations so that preventive approaches and reminders for ordering blood lead level lab tests or contacting patients lost to follow-up visits can be done.

There is a great opportunity in public health to use analytics to promote data-driven policies. We need to use our data better, share it with the public and our partners, and then leverage that data to create better policies, systems and environmental changes.

Public institutions should increasingly employ predictive analytics to help advance their efforts to protect the health of their residents. Furthermore, large, complex data sets should be analyzed using predictive analysis for improved pattern recognition, especially from diverse data sources and types, ultimately leading to significant public health action. For the Chicago Department of Public Health, predictive analytics is not the future, it is already here.

USE DATA TO FIX THE SMALL BUSINESS LENDING GAP

BY KAREN MILLS

Access to credit is a key constraint for entrepreneurs. And limited credit is in part caused by the difficulty of predicting which small businesses will and won't succeed. In the past, a community bank would have a relationship with the businesses on Main Street, and when it came time for a loan, there would be a wealth of informal information to augment the loan application. Today, community banks are being consolidated and larger banks are relying more and more on data-driven credit scoring to make small business loans—if they are making them at all.

With larger volumes of data being used to analyze everything from the genome to traffic patterns and lunch choices, it is natural to ask whether big data can crack the code on small business credit risk. There is reason for optimism.

My recent Harvard Business School Working Paper on small business credit explores new technology-driven entrants in the world of small business lending. These innovative players, such as OnDeck, Funding Circle, and Fundera are disrupting the market by using technology to solve problems that have made small business lending costly for traditional banks. For example, they use online marketplaces to reduce the search costs for willing lenders to find creditworthy borrowers. And they are allowing new sources of capital such as peer-to-peer lending to replace traditional bank capital. However, all these online models depend on developing accurate new predictive models of credit assessment, often using new sources of data.

At first blush, it seems relatively easy to build an algorithm that has greater predictive power than the personal credit scores that some lenders continue to use as their primary small business credit indicator. Personal credit scores like FICO consider a combination of metrics such as payment history, current level of indebtedness, and types of credit used by potential small business borrowers.

In the high flying days of 2005-2007, banks around the country relied heavily on these scores to make quick decisions on millions of uncollateralized small business loans, with disastrous results. Since the crisis, banks have reconsidered their overreliance on personal credit scores in small business lending. Many lenders have built their own predictive models that incorporate key metrics about the borrower's business – such as industry trends and number of employees – in addition to personal scores. Some lenders – as well as the Small Business Administration, which provides a partial guarantee on some loans made by lenders – have also incorporated third-party credit scores like those produced by Dun & Bradstreet, which use propriety predictive models that contain a blend of personal and business data to better assess borrower risk.

New entrants to small business lending have been taking this blended model even one step further. Online lending platforms like OnDeck have been using information on cash flows and direct deposits from small businesses' bank accounts as a key indicator of credit health since 2006. Intuit has been experimenting with using companies' QuickBooks data (with their permission) to create a credit score that the business can then show to lenders via a QuickBooks platform that includes several of the large banks and online lenders. Others have even gone as far as to use data from social media sites like Yelp in their predictive formulas. After all, isn't the customer's voice relevant if you are going to finance a plumber or restaurant?

NEW DATA SOURCES FOR ASSESSING SMALL BUSINESS CREDIT

		RATIONALE	SOURCES	USERS *Alternatives*	USERS *Banks*
More predictive	CASH FLOW	Best measure of firm fundamentals, gauging ability to repay and managing risk repayment cycle during life of loan	intuit, Bill.com	×	
	CREDIT BUREAU	Useful for consumer risk profile, but limited business risk profile information and weak predictor of credit-worthiness	FICO, TransUnion, EQUIFAX, YODLEE	×	×
	FIRMOGRAPHIC	Risk correlates to revenue and profit level, firm size, age, industry, geography, customer size, owner education	Census, salesforce, BEA	×	×
Less predictive	SOCIAL	Highly volatile day to day, but viewed holistically can be a predictor of riskiness, esp. in retail/restaurant	yelp, twitter, foursquare, Angie's list	×	

SOURCE THE STATE OF SMALL BUSINESS LENDING: CREDIT ACCESS DURING THE RECOVERY AND HOW TECHNOLOGY MAY CHANGE THE GAME

HBR.ORG

Some worry that social media is unreliable and can often be manipulated by an aggressive competitor or by the small business itself. And early reports from the architects of these newer algorithms caution how long it takes to thoughtfully incorporate new metrics into the models. For now, the blended models based on personal scores and business-specific data continue to be the industry standard.

However, as new entrants increasingly experiment with cash-flow and direct-deposit data as a means of better predicting the ability of a small business to repay its loans, those with easy access to that data could have a real advantage.

Currently, large banks such as Wells Fargo and JP Morgan Chase, as well as credit card companies such as American Express and Capital One, have access to vast quantities of this type of data, and are beginning to incorporate it into their predictive models more often.

It is early days in the use of predictive modeling to reduce risk and create new markets for small business loans. But the likelihood for some success seems good. As new players enter the small business lending market and unveil new opportunities, large banks with both troves of data and teams experienced in this type of modeling are beginning to take note. What seems novel and niche in small business credit scoring today has the potential to be ubiquitous tomorrow.

In August, OnDeck announced an IPO valued at $1.5 billion. Some, at least, believe that new entrants and their innovative predictive approaches can change the game in small business lending. And if that's the case, the ultimate winners will be America's small businesses and entrepreneurs.

ALGORITHMS MAKE BETTER PREDICTIONS —EXCEPT WHEN THEY DON'T

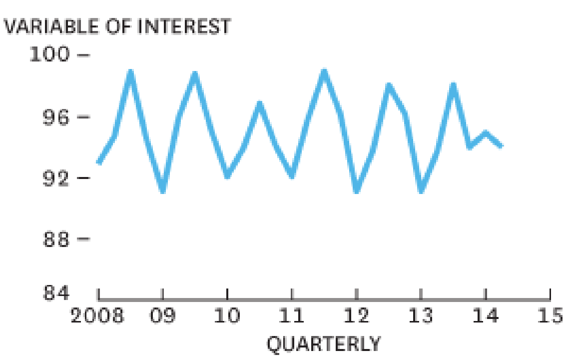
BY THOMAS C. REDMAN

Predictive analytics is proving itself both powerful and perilous. Powerful, because advanced algorithms can take a near-unlimited number of factors into account, provide deep insights into variation, and scale to meet the needs of even the largest company. Perilous, because bad data and hidden false assumptions can seriously mislead. Further, algorithms cannot (yet, anyway) tap intuition—the soft factors that are not data inputs, the tacit knowledge that experienced managers deploy every day, nor the creative genius of innovators.

So what should managers, especially leaders, do? The obvious answer is employ both computer-based programs and your own intuition. In this post, I'll use a series of simple plots to explain how to tap the potential of predictive analytics, sidestep the perils, and bring both the data and your good judgment to bear.

To start, consider the figure below, "Performance since 2008," a quarter-by-quarter time-series plot of results on a variable of interest (for example, it could be sales of a particular item, estimated resources to complete a certain project, etc). We need to predict performance for the first quarter of 2015 (1Q15).

PERFORMANCE SINCE 2008

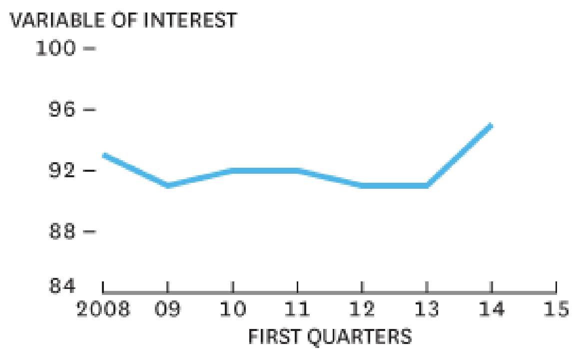


SOURCE THOMAS C. REDMAN HBR.ORG

A quick glance only might yield, "Wow, I don't know. Performance is bouncing up and down. How would I even guess?"

After staring at the plot a bit longer, most individuals (and all good analytics programs) will spot seasonality: down in first quarters, up in thirds. The next figure is a simpler plot, featuring first quarters only.

PERFORMANCE SINCE 2008, FIRST QUARTERS ONLY

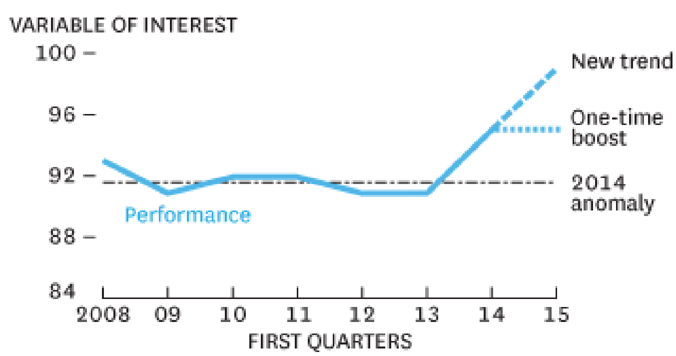


SOURCE THOMAS C. REDMAN HBR.ORG

This plot suggests that the first quarter is pretty mundane; except for 2014, performance is tightly contained in a 91 to 93 band.

So what's the prediction for 2015's first quarter? As the figure, "Potential Predictions for First Quarter, 2015," depicts, I can argue for at least three:

POTENTIAL FIRST QUARTER PREDICTIONS, 2015



SOURCE THOMAS C. REDMAN HBR.ORG

1. "We should expect 1Q15 to be like most first quarters. There were several huge snowstorms last year, so 2014 was an anomaly." Perhaps the explanation of a seasoned veteran who's learned it's best to under-promise and over-deliver.
2. "2014 is the new normal. We got a one-time boost because we turbocharged the knurdle valve." Perhaps the prediction and

explanation of an engineer who is proud to have improved a piece of the variable in question.

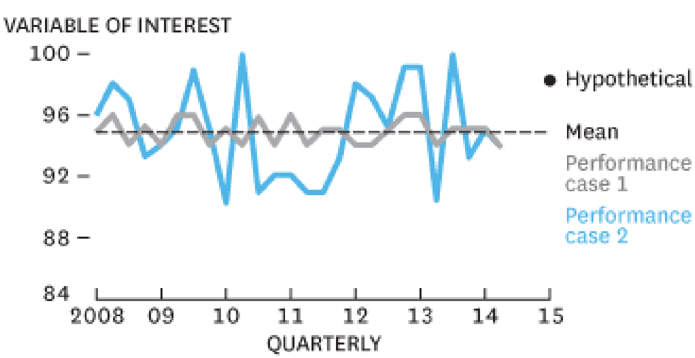
3. "We started a new trend in 2014 and should expect to see similar gains in 1Q15." Perhaps the prediction and explanation of a new product manager, aiming to score points with the boss, who is demanding across-the-board improvements.

The quandary here underscores the importance of algorithms. I have no doubt that each of these managers is smart, well-meaning, and doing his or her best. But at most, only one of them is "right." One in three is an excellent batting average in baseball, but hardly up to the demands of competitive business. Algorithms offer distinct advantages. Good ones are unemotional and (largely) apolitical. They don't care that it is best to under-promise and over-deliver or that the new boss is particularly demanding.

At the same time, they're capable of digging deeper. They can help evaluate whether the weather really played a factor in 2014 and take weather forecasts into account in predicting 2015. Similarly, they can seek evidence for the "new trend" in the second quarter and in similar variables. They can also search for possible causes. (Note: Algorithms can only detect correlation, though. Individuals must work out causation.)

In a related vein, good predictions should feature ranges, such as 94.9 ± 2.4. To see why this is, take a look at the figure below. The plot features two cases, one exhibiting low variation (in gray—note that all past values are between 94 and 96), the second relatively higher variation (in blue—values here range from 90 to 100). In both cases the mean is 94.9.

VARIATION IMPACTS THE PREDICTION RANGE



SOURCE THOMAS C. REDMAN HBR.ORG

Now suppose 1Q15 comes in at 98 (the hypothetical dot in the figure). This should come as a surprise in case 1—98 is far above anything that's happened in the past. Not so in case 2—several past values are greater. Thus prediction ranges (94.9 ± 2.4 for case 1 and 94.9 ± 9.5 for case 2) help managers understand just how close they should expect actual performance to be to the point prediction (94.9).

Calculating these ranges is quite technical. Few can do it by eyeball alone. But for good computerized algorithms, it is a snap.

These three abilities—to take emotion and politics out of the prediction, to seek deeper insights, and to quantify variation—are powerful, and leaders should seek to leverage them. That being said, managers should not be seduced into thinking that predictive algorithms are all-knowing. They are not.

Algorithms only operate on the inputs they're provided. Snowstorms affect many things and may lie at the heart of the boost in the first quarter of 2014, as mentioned above. But if weather is not part of the algorithm, the suspected explanation cannot be taken into account.

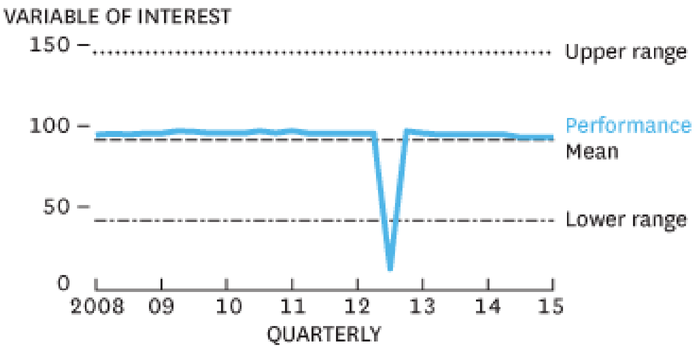
Algorithms can also be remarkably sensitive to bad data. Consider the result if you were to change one data value by dropping a decimal place (e.g., a 95 became 9.5). The resulting prediction interval changes from 94.9 ± 2.4 to 91.4 ± 50, setting a trap for the unwary. At first glance, one might not challenge the 91.4 ± 50 and use it without too much thought. The impact, from preordering too much stock to missing an opportunity to reserving too little resources for completing an important project, may go unnoticed as well. But the costs can add up. Bad data is all too common and the impact on predictions can be subtle and vicious. At the root of the financial crisis, bad data on mortgage applications led banks to underestimate the probability of default—an issue that cascaded as those mortgages were packaged into complex products.

In addition, algorithms are also based on assumptions that may effectively be hidden in arcane technical language. For example, you may have heard, "We've assumed that variables are independent, homoscedastic, and follow normal distributions." Such language can camouflage an assumption that is simply not true, since the terminology can scare people off from digging deeper. For example, the assumption that mortgage defaults are independent of one another held true enough (or didn't matter) for a long time, until pressed in the run-up to the financial crisis. As Nate Silver describes in *The Signal and The Noise*, this led those who held the mortgages to underestimate risk by orders of magnitude (and exacerbating the data quality issues noted above).

Thus, you should never trust an algorithm that you don't understand. The same applies for the input data. The only way to truly understand the algorithm is to ask (either yourself or data scientists) a lot of questions. You need to understand the physical reality that drives the variables you're interested in and the explanatory factors you're using to predict them. You need to understand the real-world implications of the assumptions.

More than anything, you need to know when the algorithm breaks down. Plots like the one below help. The figure presents the time-series for the "one misplaced decimal" situation I referenced above. I've also added the upper and lower prediction ranges (technically, this is a "control chart," and the ranges are lower and upper control limits respectively). It is easy enough to see that 3Q12 was very strange indeed. There might be an explanation (i.e., bad data), or it may be that the underlying process is unstable. This is the key insight smart managers really seek. Until they know, smart managers don't trust any prediction.

CONTROL CHART FOR PREDICTED PERFORMANCE WITH ONE MISPLACED DECIMAL



This picture also underscores the need to invest in data quality. Over the long run, nothing builds better predictions more than knowing you can trust the data. Conversely, there is nothing worse than having a meeting about the implications of 1Q15's predictions degrade into a shouting match about whether bad data stymies everything.

Finally, you must develop a keen sense of smell for predictive analytics, the data, and your own intuition. Trust your intuition and use it to challenge the analytics and the data, and conversely, use them to train your intuition. If something just doesn't "smell right," become very, very skeptical.

Good algorithms make better predictions than people most of the time—except when they don't. If you're fighting the first half of this claim, you need to get over it. Stop thinking of the algorithm as your enemy. And if you doubt the second half, prepare for some very harsh surprises.

A PROCESS FOR HUMAN-ALGORITHM DECISION MAKING

BY MICHAEL C. MANKINS AND LORI SHERER

Think for a moment about how an organization makes a decision. First come the facts, the data that will inform the decision. Using these facts, someone formulates *alternative courses of action* and evaluates them according to agreed-on criteria. The decision maker then *chooses* the best alternative, and the organization *commits* itself to action.

Advanced analytics can automate parts of this sequence; it offers the prospect of faster, better-informed decisions and substantially lower costs. But unless you're prepared to transform how people work together throughout the decision-making process, you're likely to be disappointed.

Take a simple example: a company's collections function. In years past, dozens of collection agents would receive hundreds of randomly allocated delinquent accounts every day, each one with a few facts about the customer. Each agent then reviewed a standard list of alternatives and decided how he or she would try to collect what was owed.

Today, an algorithm can assemble many more facts about the accounts than any human being could easily process: lengthy payment histories, extensive demographic data, and so on. Using these facts, it can separate the accounts into simple categories, say red-yellow-green.

Now the alternative courses of action are simpler. Red ones—low value, unlikely to pay—go straight to a collection agency. Green ones—high value, likely to pay—go to specially trained callers for white-glove service. The yellow ones require a careful review of alternatives and much more human intervention before a decision is reached.

Within the yellow and green categories, sophisticated test-and-learn experiments can inform the decisions that remain. Agents can discover from these experiments which channels and messages generate the greatest financial return while minimizing costs and customer dissatisfaction. They can thus optimize their choices about how to pursue delinquent accounts.

The new way of doing things is better and more efficient. But look at how it changes the process itself—and what's expected of the people involved:

- Software now assists with the collection and analysis of critical information, eliminating tasks once done by human beings. But people have to determine which facts to collect and how to weight them.
- Red-yellow-green or other simple categorization schemes can speed up the formulation of alternatives. Advanced analytic models can incorporate the experience of an organization's best decision makers, helping to eliminate alternatives that are less viable than others and focusing the evaluation on the most promising courses of action. People will require training in how to use the insights from the new decision-support tools.
- Within the yellow and green groups, test-and-learn results can dramatically improve the quality of decisions an organization makes. People will still need to figure out what experiments to run and then interpret the results.

The new decision procedures are likely to require investments in technology—for example, software that embeds rules and new decision logic into the workflow systems. They'll also require redesigning people's roles to fit with the new process. The possible need for new skills could mean extensive retraining and may require hiring new talent altogether.

The use of analytics can hugely improve the quality of your decisions and can increase decision process efficiency by as much as 25%. When executed well, it leads to higher customer and employee satisfaction. But analytics alone won't achieve these results; the decision process needs to change, with people learning new skills and taking on new roles. The transformation is organizational as well as technological, and is more extensive than many companies imagine.

FINDING ENTREPRENEURS BEFORE THEY'VE FOUNDED ANYTHING

BY WALTER FRICK

Venture capital is slowly but surely becoming a more data-driven business. Although data on private companies can sometimes be scarce, an increasing number of firms are relying on quantitative analysis to help determine which start-ups to back. But Bloomberg's venture capital arm, Bloomberg Beta, is going one step further: it's using an algorithm to try to select would-be entrepreneurs before they've even decided to start a company.

I asked Roy Bahat, head of the fund, to tell me a little more about it, and just how good an algorithm can be at picking out entrepreneurs.

HBR: Tell me a little bit about the fund.

Bahat: Our fund is backed by Bloomberg LP, the financial data and news company. We were created a little bit more than a year ago because Bloomberg recognized that there was something special happening in the world with start-ups. And really the only way to have a productive relationship with what I call a "day zero start-up" is to invest in them, because many of them are too early to take on big corporate partnerships, or they're still figuring out what they're doing. And what's unique about start-ups now is that in past decades, you could wait a while and watch a start-up develop before you decided how important it was. Today, in a blink of an eye something can go from two people nobody ever heard of to a significant force affecting business; hence, you have to get involved earlier. The fund invests for financial return not for quote-unquote "strategic value."

Tell me about the program with Mattermark.

We started to think, was there a way to get to know people even earlier? And we'd seen what companies were doing with predictive analytics to predict and select their customers using data. And so we just wondered: before a founder explicitly became a founder, could we predict that and develop a relationship with them? And so together with Mattermark, we built this model based on data from past and present venture-backed founders and we used it to try and predict, from a pool of 1.5 million people, the top 350 people in Silicon Valley and New York, which is where we're focused, who had not yet started a venture-backed company but we believed would do so. And so that's what we did and we reached out to them.

What factors are you drawing on that you believe are predictive?

It's drawn from a variety of public sources. It's mostly people's professional background. So the factors are things like: Did you work for venture-backed company? What role were you in that company? Educational background definitely plays a role.

But what's interesting about what it predicted is the predictions absolutely were not the caricature of a typical start-up founder. For one, the groups skewed older than the caricature of the typical start-up founder. For example, we found that being in the same job for a long time—even a decade or more in the same company—was not a disqualifier.

Second, it was an incredibly diverse group. Even though we collected zero demographic data, the output of the model was an incredibly diverse group and when we held the first event in San Francisco, it was one of the more diverse rooms that I had ever been in at an event in the technology industry. And that was just really gratifying.

And then the last thing I'd say is it was actually less engineering concentrated, less technical than we expected. We expected it to be virtually everybody having CS degrees and that kind of thing. And while many people worked at technology companies, the proportion of people who were business people was actually quite high. Having a business background actually turned out to be highly correlated with starting a venture-backed start-up.

Once you had this model, what did you do next?

We held a kick-off event in San Francisco and another in New York. The funny thing was a bunch of people who received our email saying, "You've been selected as a future founder" thought it was a scam. And so a bunch of people just simply didn't believe it, but then eventually they started to realize that actually we were completely serious.

We realized in those first few conversations that the most valuable thing in the program is the relationship they can form with each other and with actual start-up founders. And so we started hosting lunch once every other week with a small group of these future founders and some of our portfolio companies and friends in the industry and it's been great. The response has been terrific.

Our goal with them is to simply support them in achieving what they want to achieve in their careers because whether or not they end up starting a company, these people all have enormously high potential and some of them might end up being executives who we partner with at other companies. Some of them end up being recruits for our portfolio companies. Or some of them might end up inspiring us with ideas and being friends.

Is there a tension between looking for existing patterns of founder success using data and looking beyond the traditional paths? You don't want to just reflect back whatever biases might already exist in the data.

Yeah. That was one of our huge worries. Of course, you can't be exclusively data-driven. This is a business of creativity and invention. One of our worries about this future founder group was that if you use the data from past founders to predict future founders, they're all going to look exactly the same. They're going to have the same background. They're going to be identical. And it just turned out not to be true. It's interesting. When you look at the backgrounds of those founders and applied the model to new people, you ended up with a surprisingly diverse group because the data doesn't discriminate.

How will you gauge whether this works?

It's already worked. We're getting to know wonderful, unusual people with a wide range of backgrounds. They'll go places.

DO YOU KNOW WHO OWNS ANALYTICS AT YOUR COMPANY?

BY BILL FRANKS

At a corporate level, who has ultimate responsibility for analytics within your organization? The answers I most often get are "Nobody" or "I don't know." When I do get a name, it often differs depending on who I asked—a marketing executive points to one person, while finance identifies someone else. That isn't good. How can analytics become a strategic, core component of an organization if there is no clear owner and leader for analytics at the corporate level?

As predictive analytics becomes more commonplace, companies are grappling with how to better organize and grow their analytics teams. Analytics requirements can span business units, database and analysis systems, and reporting hierarchies. Without someone in a position to navigate across that complex landscape, an organization will struggle to move beyond targeted analytics that addresses just one part of the business at a time. It is also impossible to maintain consistency and efficiency when independent groups all pursue analytics in their own way. Who will champion enterprise-level analysis as opposed to business unit–level analysis?

Today, most companies have multiple pockets of analytics professionals spread throughout the organization. Years ago, one group, often marketing, decided it needed analytics support and so that group hired some analytics professionals. Over time, other groups did the same. As a result, different parts of the organization have independently had success with analytics. However, those pockets are still often completely standalone and disjointed. When I meet with analytics professionals in an organization, I've seen analysts from different parts of an organization begin the session by introduce themselves to each other—because our meeting is the first time they've ever met. It is time to connect these groups, elevate analytics to a strategic corporate practice, and assign executive leadership to oversee it.

The title isn't the important part—the role is. In some cases, it might be a Chief Analytics Officer (CAO) or a VP of Analytics. The point is that someone has to have corporate-level ownership of analytics and access to the C-suite to drive analytics initiatives and tie them to the right corporate priorities.

Where should the CAO report? In most cases today, the CAO doesn't report directly to the CEO, but to another member of the C-suite. This, too, might change over time. However, the key is that the CAO has the support of, and access to, the C-suite to drive analytics deeper into and more broadly across the organization. But wherever he or she lands, the CAO should be viewed neutrally—a Switzerland of the executive suite. The CAO should be under an executive that naturally spans all of the business units that have analytical needs, such as the Chief Strategy Officer, the CFO, and the COO.

It is often easier to see where a CAO role should not report. For example, marketing analytics is quite important to many organizations. However, if the CAO reports to the CMO, then other business units such as product development or customer service might not feel that they get equitable treatment.

I am not suggesting that the CAO come in and consolidate all analytics professionals within one central team. I have written in the past that what works best is a hybrid organization with a small centralized team supporting the distributed, embedded teams. This is sometimes, but not always, called a Center of Excellence model. Leaving successful teams in place within the units where they currently sit is fine. The key is for the CAO and his or her corporate-level team to begin to provide extra support for the distributed teams, to ensure efficiency of spend and effort across the teams, to ensure the impact of analytics is being measured consistently, and to champion the cause for new, innovative analytics possibilities that are identified.

As predictive analytics specifically and analytics in general continue to permeate organizations and change how business is done, it is imperative to put the proper emphasis and leadership in place to ensure success. An analytics revolution is coming. Creating a role such as a CAO is one way to demonstrate a firm commitment to joining the revolution.

If you can't say who owns analytics in your organization, I suggest you consider fixing that today.

COLLECT YOUR EMPLOYEES' DATA WITHOUT INVADING THEIR PRIVACY

BY CHANTRELLE NIELSEN

Research shows that businesses using data-driven decision-making, predictive analytics, and big data are more competitive and have higher returns than businesses that don't. Because of this, the most ambitious companies are engaged in an arms race of sorts to obtain more data, from both customers and their own employees. But gathering information from the latter group in particular can be tricky. So how should companies collect valuable data about time use, activities, and relationships at work, while also respecting their employees' boundaries and personal information?

In helping our customers adopt people analytics at their own companies, we've worked directly with legal teams from large companies around the world, including over a dozen in the Fortune 500. We've seen a wide range of cultures, processes, and attitudes about employee privacy, and learned that in every case there are seven key points that need to be addressed for any internal predictive analytics initiative to be successful:

Find a sponsor. The team that's proposing the data analysis needs to have real power and motivation to change the business based on the findings. Most need a sponsor in a senior-level position for this kind of institutional support. First, this person can help balance opportunistic quick wins with a long view of how predictive analytics fits into strategic plans. He or she should also explain why the data collection and analysis is so important to employees across the organization, and can serve as the person ultimately accountable for ensuring that the data stays private. In many cases, if a company's legal team doesn't see strong sponsorship and support, they are likely to de-prioritize approval of the initiative—to the point where it may be forgotten entirely.

Have a hypothesis. Before you start collecting data, decide why it's needed in the first place. For one, legal departments can't often approve a project without an objective. But in addition, the team proposing the project needs to be clear and transparent about what they're trying to accomplish. This includes having a tangible plan for what data is being sought, what changes will be made based on the findings, how the results of these changes will be measured, and the return on investment that justifies the time and energy put into the project.

The hypothesis can be as specific as "underperforming customer accounts are not getting as much time investment as high-performing accounts," or as general as "correlations will be found between people analytics metrics and business outcome x," but the outcome needs to *matter*. Projects without a purpose confuse people and incite skepticism, setting a bad precedent for future analytics efforts.

Default to anonymity and aggregation. There is more to be learned by examining the relationship between sales and marketing as a whole than there is by examining the relationship between James in sales and Elliott in marketing. Analytics initiatives are not the place for satisfying personal curiosity. In our work, we use metadata only, usually beginning with email and calendar. By default, we anonymize the sender and recipients' email addresses to their departments. To further protect anonymity, we aggregate reporting to a minimum grouping size so that it's not possible to drill down to a single person's data and try to guess who they are. This removes the possibility of even innocent snooping.

If you can't let employees be anonymous, let them choose how you use their data. In a few cases, business objectives can't be met with anonymous data. Some of our customers, for example, conduct social network analyses to identify the people who make important connections happen across disparate departments or geographies. After identifying these key "nodes" in the social graph, managers will interview them and then help them influence others. In a case like this, the best approach is to ask permission before gathering the data in one of two ways:

1. Using an opt-out mechanism is the simplest. Employees are sent one or more email notifications that they will be included in a study, with details on the study plan and scope. They have to take an action (usually clicking a link) to be excluded from the study.
2. Opt-in earns a bit lower participation, because recipients have to take the action in order to be included in the study. More sensitive legal teams may require an opt-in.

Whether it's opt-out or opt-in, the worker should know what's in it for them. We find that the most relevant reward is access to data—after all, most people are curious how they compare with their peers across various dimensions. We provide people with personal, confidential reports that compare their own data to organizational benchmarks, and this helps give them an incentive to participate. Real, personalized data also helps to make the message about the study interesting, cutting through the inbox noise so the opt-in gets attention. And if you don't have the ability to give people back their own personal data, you can promise future access to some form of aggregated study results to reward them for participating.

Screen for confidential information. Then screen again. Certain teams, such as legal, HR, or mergers and acquisitions, will be dealing with more sensitive matters than normal, and their data may need greater protection. Whether data will be gathered from humans, electronic sources, or both, sensitive information should be screened out in two ways:

1. Don't gather it in the first place by configuring the instrument to exclude keywords, characteristics, or participants that would indicate sensitivity.
2. Re-validate and remove any data that wasn't screened by the initial configuration, because both people and software can miss the meaning of textual information. Perform a second validation before sharing the data with the final audience.

Don't dig for personal information. Every person experiences interruptions in their workdays for personal reason—dentist appointments, children's activities, etc. At the same time, by policy, some companies protect their employees' privileges to use company systems for personal reasons. Regardless of policy, there really isn't much business value in looking analytically or programmatically at data about peoples' personal lives, and we automatically exclude it from our dataset. The bottom line is that employees have a human right to personal privacy, as well as significant legal rights that vary in different countries. Personal matters should be handled by managers, not by analytics initiatives.

For additional protection, consider using a third party. It is common in some applications for a third-party vendor to perform the data cleansing, anonymization and aggregation, so that the risk of privacy violations by employees of the enterprise is removed. This work can be performed by third parties even within the firm's firewall, if desired. But there's an important caveat: Companies that handle sensitive data should follow security practices, like background checks for their employees who have access to the data, and should not, in general, use subcontractors to perform their work.

The opportunity in data and predictive analytics, particularly people analytics, is huge, which makes it especially important that companies take a responsible and proactive approach to privacy. By collecting and using data in a way that respects and rewards employees, leaders remove friction points in the adoption of increasingly valuable analytical capabilities. The seven practices outlined will help clear the path for pioneering programs and build an organizational culture that prizes and rewards analytical thinking at all levels.

WHAT THE COMPANIES THAT PREDICT THE FUTURE DO DIFFERENTLY

BY JEANNE HARRIS AND MARK McDONALD

If knowledge is power, then predictive analytics promises the ultimate knowledge—that of the future. Such knowledge does not come easily, but the increasing density of digital information, deeper automated connections across companies, and increased storage and computing power create new options for enterprise leaders. For the first time in history, the predictive future—the increasing awareness and likelihood of potential future actions and outcomes—is within reach. No wonder, then, that executives have placed predictive analytics at the top of the executive agenda since 2012, according to a recent Accenture survey.

But to know more about potential future actions and outcomes and their probability—and to act on that knowledge—organizations are engaging in new kinds of relationships. We have found that the most forward-looking organizations do these three things:

1. Look to the outside: The main focus of analytics has until recently been internal, directed toward high-frequency, standardized, repeatable processes that connect variance with intervention. Using analytics, organizations have deployed bigger data sets, cheaper cloud computing power, and more aggressive algorithms to successfully standardize previously non-standard processes such as sales and service, making them more repeatable, predictable, and amenable to analytics.

To apply analytics to the future, though, self-knowledge is insufficient. The information most likely to influence the future comes from looking out the window, not in the mirror. Sheer computing power isn't the key differentiator either, because the predictive future relies less on additional statistical mastication than on a greater diversity of inputs. Consider the example of a manufacturer of production equipment that collects sensor-based telemetry about its machines' operations, the status of their parts, their performance, their resource consumption, and other data. This monitoring turns up an anomaly at a key customer that indicates a failure is imminent. Such a failure would cause a significant cost and damage the customer's brand. The manufacturer notifies the customer, which pulls the machine off line and repairs it, saving millions of dollars in lost production and damage to its brand. Business continues as usual and the equipment manufacturer has a very grateful customer.

In this example, information that was critical to the customer came from outside its walls. But while such information exchanges have become technically feasible, they are not yet financially beneficial to the information provider and difficult for the customer to value and incorporate into their management systems. Turning information exchange into value and revenue involves changing the nature of information relationships as well as management's abilities to act on that information. The most forward-thinking companies are developing new business models to create value from these kinds of information exchanges.

2. Develop open multiple multi-sided relationships: Altruism or openness alone will not give rise to ready access to the diversity of data required to understand the predictive future. The availability and veracity of the data involved in the predictive future requires creating multiple multi-sided relationships with customers, suppliers, trading partners, and just about anyone else with potentially beneficial information. It is no longer enough to share information one-to-one with partners. Increasing predictive power rests in positioning yourself at the center of multiple information flows.

Current information-based services, such as Bloomberg, involve an information provider selling a single set of information with segmented services to multiple customers. Such models play a part in the predictive future, but the industrial Internet and expanded communications capabilities change the nature of information products. From one product distributed to many customers, the move is underway toward products that feed information from many sources to a single party, which rearranges and redistributes the information to many customers. In short, from many to one to many.

There's a demand for this type of information, and thus product and market opportunities, but in an information services marketplace where people want everything for nothing, it is not easy to monetize information products. We expect, though, that a viable market will emerge as commercial terms evolve to support the multiple multi-sided relationships that give subscribers unique access to information and therefore value. Whether the information source is commercial brokers or existing commercial relationships, diverse information sources fuel the predictive future.

3. Update management and leadership practices: An extended analytics engine fueled by multiple information sources, however, can accomplish little without the ability to act on future predictions. The practice of management itself must evolve for this capability to emerge.

It is hard enough to act on solid information about the past. The level of difficulty rises when management is asked to deal with a set of predictive futures rather than projections based on past performance. Effective use of predictive analytics involves mastering a new set of management, operational, and financial techniques and disciplines.

Managerially, organizations need to revise management practices, including: increasing the use of experiments and pilots to enhance risk-taking based on external and incomplete data; incorporating test-and-learn experiences into decisions and action; enhancing awareness of the differences between causation, correlation, and coincidence; and placing tangible value on avoiding adverse effects and missed opportunities.

Operationally, organizations need to establish their own trust and execution mechanisms for multi-sided, information-based relationships. These mechanisms entail creating new analytics capabilities, securing access to third-party information and capabilities, continuously refreshing sources, and determining which data need to remain private to retain their value.

Financially, organizations require new models to account for information assets beyond treating them as intangibles. Financial arrangements have to evolve to handle pricing and payments for value based on possible futures. The ultimate goal is to treat information as a tangible flow rather than an intangible asset stuck on the balance sheet.

"The future is already here, it's just unevenly distributed." William Gibson's dictum, though overused and abused, remains true. The predictive future is valuable precisely because it's unevenly distributed and therefore in demand. Finding this future in the deluge of information available requires doing a better job of boiling the ocean. It requires investing in management, information-intensive relationships, and a broader view of analytics in the enterprise.

WHAT TO DO WHEN YOU CAN'T PREDICT YOUR TALENT NEEDS

BY JOHN BOUDREAU

Predictive analytics are often used in strategic workforce planning (SWP), to forecast and close the gap between the future talent you'll have versus the future talent you'll need. Now, powerful analytical tools are driving that organizational calculus. Those tools predict who will leave and when, where talent will be plentiful and scarce, and how talent will move between roles. But there's a catch: Very precisely matching talent to "the future" is of little value if that future doesn't happen. For example, it can take five years or more to develop today's high potentials into leadership roles. Can you know today the five-year future for which you should prepare them? Increasingly, you cannot. Yet, because HR strategy typically reacts to organization strategy, SWP often assumes a single future as its goal.

Does this mean predictive analytics don't work for talent? No. Powerful analytics have value in preparing for a VUCA (volatile, uncertain, complex, and ambiguous) world, but optimizing your talent decisions will often mean balancing *less predictive power* applied to *many* futures, against *more predictive power* applied to *one* future. Options will often trump predictions.

Where's the right balance? "Work diligently, but don't fixate on one outcome." In the yoga *Sutras*, this is *Abhyasa* (diligence) with *Vairagya* (non-attachment). It may be key to effective predictive analytics, especially for your talent.

It's easy to think expertise can solve this problem through more accurate predictions, but Philip Tetlock's book, "Expert Political Judgment" reports results from over 20 years of evidence spanning over 80,000 expert predictions. He found that "people who make prediction their business… are no better than the rest of us." In fact, the deeper the expertise, the more chance of missing something important. Tetlock found that "hedgehogs," who know a lot about one big thing, predict less accurately than "foxes" who know less about any one thing, but a moderate amount about each of many things. *Forbes* said, "Experts who had one big idea they were certain would reveal what was to come were handily beaten by those who used diverse information and analytical models, were comfortable with complexity and uncertainty and kept their confidence in check."

Do you approach strategy and talent like a hedgehog or a fox? With the power that predictive analytics bring, it's even more important for you to answer that question—are you driving toward one deeply-analyzed future or keeping your confidence in check by preparing for many futures? A hedgehog would start with a confident position such as, "the middle class in emerging regions will be the main source of consumer growth over the next 20 years," and deeply focus predictive analytics on how to meet that future. A fox would start with many positions (such as different likely regional growth predictions) and use predictive analytics to optimize a collection of tactics for different futures.

In finance, the "fox" strategy is similar to using real options, and it can help you make talent decisions just as it helps in your decisions about R&D, manufacturing and finance. Consider your talent resource like an investment portfolio. As with financial investments, you could "bet on the most likely future" (build talent to fit the one highest-probability scenario and win big if you're right but lose big if you're wrong), the typical approach noted above. Sometimes, organizations admit they can't predict the future and "go generic" by building talent attributes like intelligence, engagement and learning agility that are generally useful in most future situations, but not a complete match for any one.

Or, you might "diversify" talent, building several different talent arrays, each one well-suited to a different future scenario, similar to holding diversified financial assets, each well-suited to a particular future. Only a small portion of the portfolio will actually "fit" the eventual future, but skillful mixing in advance can optimize risk and return. Of course, people aren't financial instruments. You can adjust a financial portfolio by selling assets, but removing or retraining talent requires careful consideration. Yet, in those arenas where VUCA-like uncertainty is pivotal to your strategic success, using predictive analytics to diversify your talent options may be wiser than using predictive analytics to bet big on one future.

A "hedgehog" approach to organization and talent strategy can be a trap, even when supported by powerful predictive analytics. Perhaps your strategists should be more like foxes, optimizing prediction and options, by knowing when analytics should predict many futures moderately, rather than one future perfectly.

BEWARE THE ANALYTICS BOTTLENECK

BY BRIAN McCARTHY

Within the next three years there will be over 20 billion connected devices (e.g. oil pipelines, smart cities, connected homes and businesses, etc.) which can empower the digital enterprise—or intimidate them. With the pace of digital "always on" streaming devices and technology innovation accelerating, one might think technology would continue to pose a challenge for businesses. Historically, new technologies from the mainframe to client server and ERP—while enabling organizations to pursue new business goals—became a bottleneck to progress. This is due to constraints like lengthy implementation processes and inflexibility to adapt as business conditions changed. Turns out that isn't the case today. There is a new, even more elusive, bottleneck: the organization itself and its ability to adopt and adapt big data and analytics capabilities.

Based on our work with clients in a variety of industries from financial services to energy, here are three ways we've seen organizations embrace the analytics opportunities of today and transform from being the constraint into being the change agent for their company's future.

Don't be overwhelmed—start slower to go faster: Given the ferocious pace of streaming data, it can be challenging for many organizations to glean data insights at the same speed and determine the right data-driven decisions and actions to take. To avoid getting overwhelmed by all the data and the possible opportunities it could uncover, companies should slow down and just focus on the things that matter—it's much easier to focus on resolving five issues that could truly make a difference instead of 500 issues that might help the business.

Once the shortlist of focus areas is determined, organizations can then more effectively chase their desired outcomes by doubling down their analytics efforts in data automation and embedding insights in decision processes to help achieve their wanted results, quicker. This should also be done in tandem with continuing to drive analytics adoption in the business for an even bigger benefit.

An upstream energy equipment manufacturer, for example, used this approach to better understand the amount of time production equipment sat idling. The company knew there was huge value in solving the idle problem, but it could not do so leveraging traditional technologies as the data volumes were too large (i.e. 300,000 locations, approximately 20 machines per location, 2-300 data points per machine, and 45 millisecond sensor sample rates). Using a Big Data Discovery platform and methodology, within 10 weeks the team was able to show more than $70M in savings from analysis from a subset of the locations and could analyze the data at high speeds (e.g. 13,500 sites, 20 TB, 15 seconds to render).

Technology doesn't have to be exposed (Keep the complexity behind the curtain): Organizations shouldn't be reticent to explore new technologies and experiment with their data to improve the effectiveness of their analytics insights for key decision processes. Machine learning, or the growing set of data discovery and analysis tools used to uncover hidden insights in the data, is a sophisticated technology that can do just this. Its data exploration capabilities and simplicity are also becoming necessities to ensuring competitiveness in the connected world.

Machine learning techniques can aid a company to: learn from past behavior and predict behavior of new customers (e.g. risk models to predict consumer risk to default), segment consumer behavior in an optimized, market friendly fashion (e.g. customer lifestyles modeled from geo-location data on cellphones), or conduct crowd simulation models where each customer's response to a reward is modeled. This is just a snapshot of possibilities; many more types of outcomes from machine learning are also possible.

For example, one retail bank applied machine learning to its customer analytics and achieved a 300% uplift on sales campaigns compared to a control group. Despite this lift, the bank was experiencing relatively slow adoption in the retail channel with many branch managers still using traditional methods of relationship selling. To improve the adoption rate the bank focused on a change program that dumbed down what qualified leads meant and also showed the managers the WIIFM ("What's in it for me?") approach to show how this would help them achieve their goals.

Make faster decisions for faster rewards: It's important for businesses to sense, analyze, interpret and act fast on the data insights as competitive advantages will likely be more fleeting than long lasting in the hypercompetitive world. With this, we are seeing a fundamental shift in strategic decision making that is powered by big data discovery, a capability that accelerates the time to insight.

As an example, a large bank used a data discovery capability to gain deeper insight into their customer experience strategy and understand why there was a drop off in customer satisfaction. The data discovery analysis took weeks instead of months, where a team of data scientists, functional experts and business analysts worked to tag, filter and find correlations in the data, and how it differed by customer segments. The analytics team discovered that the bank's most affluent customer segments were the most digitally savvy, and they were dissatisfied with their digital experience, online and on their mobile devices. The bank thought service fees were the issue, and while they were a strong issue overall across all customers, it wasn't the most important issue for their most profitable customers. As a result, the bank changed their customer experience strategy by altering their approach to service fee refunds and enabling wealth advisers to connect with customers digitally.

It's a reality: Data is going to keep growing and technology options will follow the same trajectory. Organizations shouldn't run from this new digital reality, but learn to embrace it by adopting and adapting their analytics strategies to remain competitive. By applying the power of data and analytics techniques such as machine learning, a firm can make smarter, faster decisions for their business and its customers, and actively disrupt their industry.

XBOX POLLING AND THE FUTURE OF ELECTION PREDICTION

BY DAVID ROTHSCHILD

For generations, pollsters have used probability polling (think of the Gallup polls quoted on the nightly news) as their go-to method to forecast the outcomes of elections. But cost increases and concerns about accuracy have called the method into question. A new form of polling called non-probability sampling—opt-in surveys on the internet, prediction markets, and even polls on gaming systems—has emerged as an improvement, and a viable replacement.

First, let's take a look at probability polling, which works like this: ask a random sample of likely voters who they would vote for if the election were held that day, and the answer is almost as accurate as asking everyone. This method has worked relatively well in countless election cycles, but it's growing more difficult to receive accurate results. One reason: the rise of cell phones. For a period in the 1980s nearly all likely voters owned a land-line; now the catalog of likely voters is spread over landlines and cell phones, or both, which makes it hard to figure out what the sample really is. In other words, where are all of the likely voters? The next problem is non-response error. Not all likely voters are equally likely to answer the poll, if contacted. This error is due to simple things like differences between demographics (e.g., some groups are more likely to answer calls from unknown numbers) and more complex things like household size. In other words, which likely voters are responding to polls, and do they differ from likely voters who don't?

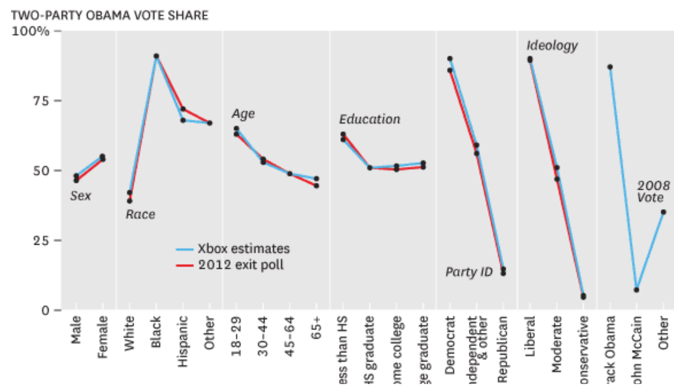
There are serious selection issues with non-probability samples as well—just like probability samples, they are prone to coverage and non-response errors—but the data is so much faster and cheaper to acquire. For example, in 2012, my colleagues and I collected opinions on the U.S. Presidential election from Xbox users by conducting a series of daily voter intention polls on the Xbox gaming platform. We pulled the sample from a non-representative group of users who had opted-in to our polls. In total, over 350,000 people answered 750,000 polls in 45 days, with 15,000 people responding each day and over 30,000 people responding 5 or more times. At a small fraction of the cost, we increased time granularity, quantity of response, and created a panel of repeated interactions.

But the raw data still needed to be turned into an accurate forecast. With our Xbox data, we first needed to create a model that incorporated the key variables of the respondents. We did this by determining the likelihood that a random person, from any given state, would poll for Obama or Romney on any given day, based on state, gender, age, education, race, party identification, ideology, and previous presidential vote. Then, we post-stratified all possible demographic combinations, thousands per state, over their percentage of the estimated voting population; for transparency we used exit poll data from previous elections. Finally, we transformed the Xbox data into an expected vote share — by detailed demographics, and probability of victory, for all states.

In the below figure, you can see the accuracy of the Xbox pre-election estimates compared to exit-poll data.

XBOX ESTIMATES VERSUS EXIT POLLS

Xbox data—which predicted the percentage of Obama vote shares across various demographic subgroups—was remarkably accurate compared to actual voting behavior.



As you can see, our forecasts were accurate—even compared with the best aggregations of the traditional polls—and they provided detailed demographic insight as well. Further, we were able to gain a new understanding of the movement (or lack thereof) of swing voters, because we had so many repeated users. The accurate forecasts, new relevant insights, and ability to easily update daily, all came at a much lower cost than traditional probability polling.

Yet there are meaningful groups of researchers that cling to the past, even as more papers confirm our findings. Their argument is that declining response rates don't affect results in a major way, so why worry and innovate?

Yes, it is possible that our Xbox polls would be slightly less accurate in other domains or with smaller samples. But within a few years, there's no doubt that traditional polls will lose their statistical power and become less accurate.

Xbox polling, and other forms of non-probability polling, will be an increasingly crucial tool for campaigns and advertisers in future elections. Campaigns have the capacity to target detailed demographic groups, or individuals, with messages specifically designed for them. And, because non-probability polling allows for continually updated forecasts for specific demographic groups, they can be even more efficient at targeting and delivering those messages.

HOW AIG MOVED TOWARD EVIDENCE-BASED DECISION MAKING

BY MURLI BULUSWAR AND MARTIN REEVES

New developments in data science offer a tremendous opportunity to improve decision-making. Machine learning, pattern recognition, and other predictive analytics tools can constitute a source of competitive advantage for those companies that adopt them early on; but like any new capability, there is an enormous gulf between awareness, intent and early engagement, and achieving significant business impact.

How can companies better manage the process of converting the potential of data science to real business outcomes? How can companies go beyond merely generating new insights to changing behaviors—not only of their employees, but customers too? We would like to offer some lessons from AIG's early experiences with deploying new analytical tools to leaders across industries who may be considering embarking on a similar journey.

In January 2012, AIG launched the "Science Team." One might be surprised to find a Science Team in an insurance company. However, Peter Hancock, President and CEO of the global insurance giant, saw a huge opportunity to apply evidence-based decision making in an industry which was still very reliant on individual expert judgment and in so doing to create not only tactical but also competitive advantage. By early 2014, 130 people from diverse scientific and managerial backgrounds were devoting themselves to realizing the team's mission: *To be a catalyst for evidence-based decision making across* AIG.

The Science Team intentionally refrains from using the words "data" or "analytics," as the team's capabilities stretch far beyond these two disciplines: behavioral economists, psychologists, engineers, and change management experts work hand-in-hand with data scientists, mathematicians, and statisticians. And for good reason: **this multidisciplinary approach is essential to go beyond merely generating new insights from data but also to systematically enhance individual human judgment in real business contexts**. Ninety percent of the team was recruited from beyond the insurance industry to enable it to challenge the status quo approach to decision-making. The Science Team not only prepares data and builds models, but also emphasizes the identification of business opportunities and education, change management and implementation—the complete value chain from framing questions through to changing behaviors.

Key factors in the success of the Science Team's efforts to date include the following:

Start by focusing on questions and problems that matter. A small proportion of worker's compensation claims account for a large proportion of complexity, contention, delay and losses for AIG: 10% of claims account for almost 60% of costs. Claims severity predictors therefore play a huge role in improving outcomes by enabling earlier and more accurate targeting of intervention measures like physician review and special investigations. This is a good example of the power of fully embedding the technical solution in the business: the result is not only better predictions and lower costs, but also better outcomes for customers.

Ensure that the mandate stretches beyond producing insights—supporting the change and learning process across the organization. AIG not only supports embedding solutions and managing change to realize specific opportunities, but has also launched a company-wide initiative to improve quantitative and decision-making skills using both physical summits and on-demand, modular online learning tools.

Work with early adopters to demonstrate significant wins which are visible to the whole organization. Much of AIG's business relies on agents and brokers. Relationships are assessed and prioritized based on volume, value, potential, and their overall effectiveness. The decision platform which AIG built is able to accurately predict the retention and "submission" (proposal) efficiency of single brokers—a level of micro-segmentation and prediction which few others in the industry have been able to achieve. Every day, aggregated and deep-dive performance analytics, presented in a user-friendly visual format, are pushed to the fingertips of sales managers to support decisions on how to manage the network of intermediaries.

Don't make the effort dependent on one or two initiatives: adopt a portfolio approach. In pioneering new approaches to decision-making not every effort can be a success and companies should therefore not bet only on the success of one project. In addition to the examples above, AIG currently has around a dozen decision making related projects at various stages of development.

An iterative, rapid cycle adaptive approach is much more effective than a planned, single step change—much of the learning occurs by taking action. Preventing fraudulent claims is an important area for AIG due to its significant financial impact. AIG has developed proprietary tools and models that identify predictive patterns in claims data using machine learning, predictive modeling, link analysis, pattern analysis and other techniques. After starting from scratch, the second generation of AIG-developed tools already identify almost twice as many cases of fraud than leading vendors' offerings. First applied to worker's compensation, the same approaches are being now being rolled-out across multiple businesses. This example illustrates the importance and power of an iterative, learning-based approach to solution development. Ironically, this involves a bias to action rather than planning or analysis—even in the area of analytics!

Plan for impact on multiple time-horizons, combining immediate evidence of value, some medium term big wins as well as a transformational long term perspective. In addition to the short- and medium-term solutions mentioned above, AIG is also contemplating some bolder, longer term initiatives which could potentially change the business model and the scope of the business. For example, it is looking at possibilities like assessing damage claims for auto accidents using image analysis of photographs, or measuring and modulating risk assessments using sensors and telematics.

The constantly evolving tools of data science will both enable and require companies to continue to improve how they make decisions. It's self-limiting to only improve existing decision-making, however—companies need also to be alert to the opportunity of creating fundamentally new ways of making decisions, and even to reconsider new the business models and the firm's activity footprint, as a result of the opportunities unleashed.

WHEN A SIMPLE RULE OF THUMB BEATS A FANCY ALGORITHM

BY JUSTIN FOX

For a retailer, it's extremely useful to know whether a customer will be back or has abandoned you for good. Starting in the late 1980s, academic researchers began to develop sophisticated predictive techniques to answer that question. The best-known is the Pareto/NBD (for negative binomial distribution) model, which takes a customer's order history and sometimes other data points, then simulates whether and how much she will buy again.

Actual retailers, though, have tended to stick with simpler techniques, such as simply looking at how long it has been since a customer last bought anything, and picking a cutoff period (nine months, say) after which that customer is considered inactive.

This resistance to state-of-the-art statistical models has frustrated the academics. So, a decade ago, marketing professor Florian von Wangenheim (now at the ETH Zurich technical university in Switzerland) and his then-student Markus Wübben (now an executive at a tech incubator in Berlin) set out, in Wangenheim's words, to "convince companies to use these models."

To do this, Wübben and Wangenheim tested the predictive accuracy of Pareto/NBD and the related BG/NBD model against simpler methods like the "hiatus heuristic"—the academic term for looking at how long it's been since a customer last bought anything—using data from an apparel retailer, a global airline, and the online CD retailer CDNow (from before it was acquired by Amazon in 2001). What they found surprised them. As they reported in a paper published in 2008, rule-of-thumb methods were generally as good or even slightly better at predicting individual customer behavior than sophisticated models.

This result wasn't a fluke. "I've seen much more research in this area, many variables have been added to these models," says Wangenheim. "The performance is slightly better, but it's still not much."

One way to look at this is that it's just a matter of time. Sure, human beings, with "their limited computational abilities and their incomplete information," as the great social scientist Herbert Simon put it, need to rely on the mental shortcuts and rules of thumb known as heuristics. But as the amount of data that retailers are able to collect grows and the predictive models keep improving, the models will inevitably become markedly better at predicting customer behavior than simple rules. Even Simon acknowledged that, as computers became more powerful and predictive models more sophisticated, heuristics might lose ground in business.

But there's at least a possibility that, for some predictive tasks at least, less information will continue to be better than more. Gerd Gigerenzer, director at the Max Planck Institute for Human Development in Berlin, has been making the case for decades that heuristics often outperform statistical models. Lately he and others have been trying to define when exactly such outperformance is most likely to occur. This work is still ongoing, but in 2011 Gigerenzer and his colleague Wolfgang Gassmaier wrote that heuristics are likely to do well in an environment with moderate to high uncertainty and moderate to high redundancy (that is, the different data series available are correlated with each another).

Citing the Wübben/Wangenheim findings, Gigerenzer and Gassmaier (why so many of the people involved in this research are German is a question for another day), posited that there's a lot of uncertainty over if and when a customer will buy again, while the time since last purchase tends to be closely correlated with every other available metric of past customer behavior. Ergo: heuristics win.

There are other areas where the heuristic advantage might be even greater. Financial markets are rife with uncertainty and correlation—and the correlations are strongest when the uncertainty is greatest (think of the parallel downward trajectories of lots of different asset classes during the financial crisis of 2008). Sure enough, while sophisticated financial models performed poorly during the recent financial crisis, simple market heuristics (buying stocks with low price-to-book-value ratios, for example) have withstood the test of time. Along those lines, Gigerenzer has been working with the Bank of England to come up with simpler rules for forecasting and regulating financial markets.

"In general, if you are in an uncertain world, make it simple," Gigerenzer said when I interviewed him earlier this year. "If you are in a world that's highly predictable, make it complex." In other words, your fancy predictive analytics are probably going to work best on things that are already pretty predictable.

INTEGRATE ANALYTICS ACROSS YOUR ENTIRE BUSINESS

BY BRIAN McCARTHY

IAn Accenture survey conducted last year found that only one in five companies said that they were "very satisfied" with the returns they've received from analytics to date. One of the reasons analytics is working for the companies in this select group is because they tend to deploy analytics technologies and expertise across the breadth of the enterprise. But the survey also found that only 33% of businesses in the U.S. and Western Europe are aggressively adopting analytics across the entire enterprise. This percentage marks an almost four times increase in the trend of enterprise-wise adoption compared to a survey conducted three years earlier, but the question must still be asked—how can we improve this number?

Cross-functional analytics can be a challenge to implement for a variety of reasons including functional silos and a shortage in analytics talent. Yes, these obstacles can seem daunting at first, but our experiences tell us that they are not insurmountable. Following are tips organizations can follow to drive a horizontal focus on analytics and achieve their desired business outcomes, such as customer retention, product availability, or risk mitigation.

Identify the right metrics that "move the needle". First, senior management should decide on the business goal for an analytics initiative and the key performance indicators to track that will put them on the right path toward success. For a high-performing retailer, we found that customer retention, product availability, labor scheduling, product assortment, and employee engagement were all leading indicators to driving growth and profitability for the company. Selecting the right critical metrics is a cornerstone of success as it brings focus and clarity on what matters most to the business.

Establish a center of gravity for analytics. Next, create an Analytics Center of Excellence (CoE) that spans the enterprise. A CoE is a team of data scientists, business analysts and domain experts from various business functions—sales, marketing, finance, and R&D, for example—that are brought together to facilitate a cross-pollination of experiences and ideas to find solutions to a variety of business goals. The CoE itself is organized into pods—generally made up of four to six people, with each person offering a different skillset—that are deployed across the business to solve problems that span multiple functions.

Develop a robust root cause analysis capability. Once CoE is created, the pod teams should perform root cause analyses to support the performance management process. The retailer example mentioned above used root cause analysis to answer the question around what factors contributed to an unsuccessful marketing promotion. They tested hypotheses by asking questions such as: were results poor because of the marketing message, pricing and bundling, product availability, labor awareness of the promotion or did a competitor have an attention-grabbing marketing campaign happening at the same time? A successful CoE model provides a company with the capability to not only answer these questions with validated cross-functional insight, but also to determine the best decision around what to do next.

Make collaborative decisions. Using a CoE affords functional managers the ability to make collaborative and informed decisions. They are not left alone to develop root cause analysis insights in a vacuum. Rather, as a team, the managers and the CoE are able to make decisions and take actions based on the insights garnered together. To accomplish this, it is critical to establish a forum with the cross-functional business leaders to share and visualize the data and interpret the insights for the purpose of decision making.

As an example, a consumer products company used a weekly executive management meeting as the forum to discuss the CoE's insights and make decisions based on the outputs. In this instance, the head of the Analytics CoE was the facilitator of the meeting and focused the executives' time on the decisions that needed to be made based on the important insights the data identified versus the noise that should be ignored (e.g. to better understand the effectiveness of a new product launch). The combination of data science, advanced visualization, and active decision making—along with an impartial facilitator with deep content expertise—was key to collaborative and effective decision making.

It's important to note that once data-driven decisions are made and actions are set in motion, companies should track their progress against the metrics that were established at the start of their analytics journey. If goals are not being realized, they should repeat the process to understand the root causes of an issue that will help them achieve their business goals. In one instance, a bank's Analytics CoE delivered such consistently positive results that the company formally branded all analysis coming out of the CoE so the business leaders could be aware of its quality and credibility outright. The branding encouraged business leaders to trust the insights and act on them faster.

When a company expands its analytics purview from functional to horizontal, it opens the door to greater opportunities and successes. While removing silos and taking a teaming approach to analytics is part of an internal virtuous cycle, another cycle is also created—the attained results are experienced by the customers and will keep them coming back for more.

CREATE A STRATEGY THAT ANTICIPATES AND LEARNS

BY JEFF ELTON AND SIMON ARKELL

The buzz around using predictive tools to analyze big data in discrete areas of a business is loud and deserved. In health care, these tools are changing the way doctors identify people at risk of developing certain diseases; in fashion, they crunch purchasing data to anticipate trends; sales and marketing experts use them to tailor ad campaigns. The restaurant chain, Olive Garden, uses predictive analytics to guide its food buying and retail staffing plans.

But maybe the thrill of accomplishment in these pockets is diverting senior managers' attention from another, even more critical opportunity: Digital technologies are also rapidly changing how managers can acquire and assess the information they use to develop and execute on enterprise-wide strategy. Strategy-making can now happen in real time. Strategy can anticipate and learn.

Traditionally, the discipline of strategy has emphasized a deep understanding of market economics and potential disruptors, the evolution of demand and value expectations, the competencies of the organization, and the role of talent and performance management. Long-dominant frameworks like the Five Forces or SWOT analysis have been based, accordingly, on a fundamental, often static or relatively long-duration, set of market and firm characteristics.

Today, though, many of those characteristics are in flux much of the time. And so the power of incumbency, firm competencies, and market share is giving way to the ability to engage across companies and industries, innovate, individualize, and deliver. The definition of a market, customer, partner, or even competitor is now a moving target. Consider, for example, the work that Apple is doing with Epic (an electronic health record provider for hospitals and large medical groups). Together, these two companies are bridging the divide between personal health data that's collected in a clinical setting, and data that's collected by the patient. Not only can patients gather more comprehensive "home-based" data with Apple's HealthKit platform, but also potentially stream that data (with permission) to their doctors via Epic's systems. Is Apple suddenly a healthcare company? To what extent? Neither Apple, nor Epic, is a cog in a linear value chain (as is, say, a company that provides a variety of components with applications in different industries, like semiconductors for aircraft, appliances, or vehicles). Instead, together, they are sketching the outlines of a new market.

In this new environment, where markets can be created by ecosystems of partners, and innovations can originate anywhere in an ecosystem and grow at great speed, the ability of business leaders to *predict and influence* what's around the corner—rather than act on what they see—becomes central to the ability to commit to a direction and allocate resources.

At the same time, powerful new tools are becoming increasingly available to enable real-time strategic decision making. Now we have an opportunity to crunch the insights of key talent, data assets, and technologies from multiple internal and external sources, as they arise. We can connect insights and execution at a pace never before possible. That strategy in real-time, or even more aptly, strategy that anticipates and learns. Using machine-learning tools, for example, data that currently exists in different enterprise systems and diverse external sources (production, supply chain, market, customer trend, financial and economic data) can be ingested and mashed together to reveal meaningful patterns and highlight gaps in markets. These analyses can identify opportunities for maverick business partnerships, and balance the biases of individual decision makers quickly and effectively.

This isn't a retread of scientific management, nor is it an updated take on scenario planning. It's an entirely different animal. To call it a new version of either, in fact, would be to overlook entirely the volume and scope of information that big data can provide and predictive analytics can crunch—in real time—for dynamic strategic purposes at the enterprise level. Scientific management and scenario planning, while forward-thinking, rely on information that's in the rear view mirror.

No company is yet an exemplar of setting and activating strategy in the way we envision it. In many companies, setting and enabling strategy is still a regularly scheduled process or a defined annual deliverable. But a number of businesses have put more pieces of the practice than others into place. Amazon, for example, analyzes enormous pools of data to predict who their customers are and what their customers will buy next. The company uses these insights to drive—and adapt—strategic plans for new device and service offerings (e.g., Fire TV, Amazon Prime) and to stock its warehouses.

One can easily imagine Amazon and other like-minded companies building out more and more tech-enabled strategic and operating capability—linking the pieces. Financial services companies have made a promising start, and we are also seeing signs that other sectors—healthcare, life sciences, media, and entertainment—are waking up to the possibilities. The snag is that using predictive analytics in this way will be difficult for global companies with traditional compliance-centric and business intelligence reporting capabilities. Rigid, rules-based enterprise systems, installed in most companies 15 or 20 years ago, can't easily be re-jiggered to integrate data and “mine” for patterns. Current enterprise technologies, and the business processes they support, are so hard-wired in most big companies that shifting to a more fluid, fast-paced, way of operating will be a major transformation.

The onus is on the senior leaders at these firms to demand predictive insights at the executive table and within core management processes by:

- Investing in enabling data infrastructure and advanced analytics just as they would top talent, a new product innovation, or a strategic relationship
- Embedding predictive analytic approaches throughout the organization—from the front line to the C-suite
- Advocating their use both formally (as performance requirements) and by example, to move the organization’s focus from planning and coordination to analytics-driven and anticipatory
- Holding these analytics to the same standard of precision, performance and improvement as management and key processes.

Predictive analytics is bringing new levels of speed, relevance, and precision to decision making. Prediction as a mode of engagement and insight will increasingly be a requirement for setting strategy. The companies and executive teams advancing, mastering, and integrating prediction as core to how they evolve strategies and manage will be the distinctive performers and leaders of the future.

OCTOBER 3, 2014

PREPARE YOUR ORGANIZATION TO CAPITALIZE ON PREDICTIVE ANALYTICS

FEATURING BRIAN McCARTHY

Contributors
Brian McCarthy, Managing Director, Information & Analytics Strategy, Accenture, Analytics

Walter Frick, Associate Editor, *Harvard Business Review*

Overview

Predictive analytics has the power to change what organizations do and how they do it. Yet many companies are equipped with the right technology, but most lack the organizational capacity to take full advantage of predictive analytics. In addition, many organizational processes aren't built to make use of analytics and make it a competitive advantage.

High-performing organizations leverage the power of analytics by channeling their efforts in four areas: focus, adopt, adapt, and activate. These companies have embraced a new paradigm that promotes agility, fast execution, and lasting organizational change.

Context

Brian McCarthy discussed how organizations can capitalize on analytics to solve key business problems and propel their business forward.

Key Takeaways

The promise of analytics is alluring, but many organizations fail to capture the full value.

While organizations in all industries are investing in analytics, executives are often disappointed by the return on those investments. Barriers to ROI are commonly associated with constraints related to the following areas:

1. **Technology trends.** It is difficult for organizations to keep up with the pace of technology innovation, as well as the data explosion associated with today's highly connected world.
2. **Business trends.** Companies often struggle with the hypercompetitive business environment. Business volatility is on the rise, as are competitive pressures and customer expectations.
3. **Organizational issues.** The ability to keep up with the pace of change is limited by organizations' infrastructure and capacity to adopt. Key factors include organizational constraints and culture.

Research reinforces the idea that many organizations are unable to capture the full value of analytics. Accenture research found that analytics adoption has increased in recent years, but the average ROI still lags behind expectations. Findings include:

- Analytics adoption has increased threefold in the last three years.
- One third of companies now use predictive analytics to run their business.
- Two thirds of organizations have appointed chief data officers.
- However, only one out of five organizations is "very satisfied" with its ability to derive value from analytics.

Obstacles to capitalizing on analytics include siloed organizational structures and shortages of analytical talent.

When high-performing organizations focus, adopt, adapt, and activate, it unleashes the power of analytics.

By channeling their efforts in four areas, high performers leverage the power of analytics:

1. **Focus.** It is essential to focus on the right metrics. Given the current information explosion, there are more things that can be counted than ever before. However, relatively few will move the needle in terms of performance. A retailer, for example, measured 52 KPIs at the store level. Yet, few of these measures were clearly linked to positive business results. High-performing organizations select a small number of critical metrics that affect the business.
2. **Adopt.** Analytics technologies and expertise are most productive when deployed horizontally across the enterprise. McCarthy recommended three best practices for improving cross-functional adoption of analytics:
    - *Establish a center of gravity for analytics.* It is helpful to establish pods of employees who have a portfolio of skills related to analytics, such as a technology architect, data

“Pods of excellence think big, start small, scale fast, and create organizational momentum around analytics.” —Brian McCarthy

scientist, business analyst, visualization artist, and data ninja. This cross-pollination of skills drives high-performance teams and is a key enabler.

- *Develop a strong root-cause analysis capability.* An analytics center of excellence (COE) is a great resource to answer questions about key metrics, generate and validate insights, and identify the best actions to drive value.
- *Make collaborative decisions.* In high-performing organizations, business leaders meet with members of the analytics COE to interpret insights and determine appropriate actions. In collaborative decision-making meetings, it's a good idea to ask an impartial facilitator to manage the discussions.

1. **Adapt.** In the past, technology was the constraint to change. Today, however, organizational ability to change has become the bottleneck. To embrace analytics opportunities, organizations must transform themselves into change agents. Adapting decision-making processes is the key to successful adoption. Three techniques organizations can use to streamline adaptation are:
   - *Go slower to go faster.* With so much information available, it can be difficult to glean insights. By focusing effort, however, it is possible to accelerate impact. One organization had 300,000 locations and 20 machines in each location that were generating data. It used its big data discovery platform to analyze a subset of that information, resulting in $70 million in savings.
   - *Keep complexity behind the curtain.* In the world of aviation, airplane cockpits used to be highly complicated. Although that complexity still exists, the user interface has been streamlined to shield pilots from distractions. In the world of analytics, machine learning serves a similar role. Machine learning is a set of data discovery and analysis tools that uncover hidden features and patterns in the data. Machine learning algorithms can be used to create risk models, customer lifetime value models, and crowd simulation models.
   - *Make faster decisions for faster rewards.* Agile decisions are a characteristic of high-performing organizations. The military's OODA Loop (Observe, Orient, Decide, Act), for example, enables commanders to compress the time between observing a situation and taking an action. Similarly, analytics can help organizations make more rapid decisions. A large North American bank saw a decrease in its Net Promoter scores and thought it was due to banking fees. Further analysis, however, revealed that the primary issue was that the bank's affluent customers were dissatisfied with its online and mobile banking capabilities. In response, the company stopped refunding service fees to its less desirable customer segments, took personal advisers out of the branches (since affluent customers were doing more business online), and repurposed those resources to improve its online banking experience.

4. **Activate.** High performers activate virtuous cycles via double-loop learning. The first learning loop focuses on learning from output. For example, if an organization doesn't meet its KPI targets, it makes adjustments and course corrects. The second learning loop focuses on learning from doing. The result of the second learning loop is process improvements over time.

A successful analytics journey requires organizations to embrace a new paradigm.

Based on his experience with numerous companies, McCarthy has found that organizations derive the greatest benefit from analytics when they do these three things well. They are:

1. **Agile in discovery.** High-performing organizations use a value-led approach to analytics. They identify key metrics that narrow their focus and enable more efficient access to data discovery. For example, a life insurance company realized that most data about older policies was in pdfs and written notes. The team used technology to scan the information into an unstructured database. They then structured the data, analyzed key signals, and developed a model that helps them better underwrite risk. An iterative approach to projects, which enables teams to fail early, also improves agility. Another best practice is to build relationships between the business and the analysts.
2. **Industrialized in execution.** Organizations that are “industrialized” in execution understand how to drive value quickly and get returns from their efforts. They often partner for innovation and use crowdsourcing. Relationships may be cultivated with third-party companies and academia. By looking for proof points, it is possible for organizations to quickly scale the insights gathered from analytics.
3. **Sustaining the change.** Organizations that sustain the new paradigm needed for successful analytics answer the “what” and “why,” before the “who” and “how.” Focusing on the value proposition is much more important than the organizational constructs. Change is also sustained through closed loop communication and building bridges between the business and the analytics teams. It is essential to celebrate success and learn by doing.

"When it comes to sustaining the change related to analytics, answer the 'what' and 'why' before the 'who' and 'how.' Focus on the value proposition, rather than the organizational constructs." —Brian McCarthy

Other Important Points

- **Best practices for COEs.** In some organizations, the CEO sponsors the Analytics COE. However, this top-down approach is less common than the "middle-up" approach. When COEs are developed from the middle up, pods of excellence conduct multiple analytics pilots that generate learning and organizational momentum. They think big, start small, and scale fast. Pods typically include four to six employees.
- **Front-line data literacy.** There are two aspects to employee data literacy. Business teams must strengthen their analytics competencies, but analytics teams must also strengthen their business acumen.
- **Analytics and small businesses.** When smaller organizations implement analytics programs, they are often more successful changing processes than larger companies.

Learn More

Predictive Analytics
Anticipate future demand, identify unseen trends or evaluate unknown outcomes for better decisions.

sas.com/en_us/insights/analytics/predictive-analytics.html

Need more insight from your data? Advanced analytics software from SAS can show you how.

sas.com/en_us/software/analytics.html

SAS Visual Statistics

Why it is predictive analytics redefined.

sas.com/en_us/software/analytics/visual-statistics.html

Biographies

Brian McCarthy
Managing Director, Information & Analytics Strategy, Accenture Analytics

Within Accenture Analytics, Brian McCarthy is responsible for the Information & Analytics Strategy market offerings and the global analytics innovation agenda. He also has responsibility for driving a broad range of analytics offerings through Accenture's Financial Services operating group North America.

Brian has extensive experience in value-based Performance Management, Finance Transformation and Analytics engagements with clients across multiple industries. His primary focus is to help client organizations address key performance management and analytics challenges and align around increasing shareholder and stakeholder value. He is currently helping a variety of national and global clients define their analytics strategies with a focus on using analytics to drive competitive advantage and improved business outcomes.

Walter Frick (Moderator)
Associate Editor, *Harvard Business Review*

Walter Frick is an associate editor at *Harvard Business Review*. He writes and edits on a wide range of topics, with a particular focus on data and technology, as well as on new business research. Before HBR, he covered startups and venture capital in Boston. He has written about technology and business for *The Atlantic*, BBC, and *MIT Technology Review*, among other publications.

This summary was created for *Harvard Business Review* by BullsEye Resources.
www.bullseyeresources.com