

CHAPTER 1

Liftoff! Prediction Takes Action

How much guts does it take to deploy a predictive model into field operation, and what do you stand to gain? What happens when a man invests his entire life savings into his own predictive stock market trading system? Launching predictive analytics means to act on its predictions, applying what's been learned, what's been discovered within data. It's a leap many take—you can't win if you don't play.

In the mid-1990s, an ambitious postdoc researcher couldn't stand to wait any longer. After consulting with his wife, he loaded their entire life savings into a stock market prediction system of his own design—a contraption he had developed moonlighting on the side. Like Dr. Henry Jekyll imbibing his own untested potion in the moonlight, the young Dr. John Elder unflinchingly pressed “go.”

There is a scary moment every time new technology is launched. A spaceship lifting off may be the quintessential portrait of technological greatness and national prestige, but the image leaves out a small group of spouses terrified to the very point of psychological trauma. Astronauts are in essence stunt pilots, voluntarily strapping themselves in to serve as guinea pigs for a giant experiment, willing to sacrifice themselves in order to be part of history.

From grand challenges are born great achievements. We've taken strolls on our moon, and in more recent years a \$10 million Grand Challenge prize was awarded to the first nongovernmental organization to develop a reusable manned spacecraft. Driverless cars have been unleashed—“Look, Ma, no hands!” Fueled as well by millions of dollars in prize money, they navigate autonomously around the campuses of Google and BMW.

Replace the roar of rockets with the crunch of data, and the ambitions are no less far-reaching, “boldly going” not to space but to a new final

frontier: predicting the future. This frontier is just as exciting to explore, yet less dangerous and uncomfortable (outer space is a vacuum, and vacuums totally suck). Millions in grand challenge prize money go toward averting the unnecessary hospitalization of each patient and predicting the idiosyncratic preferences of each individual consumer. The TV quiz show *Jeopardy!* awarded \$1.5 million in prize money for a face-off between man and machine that demonstrated dramatic progress in predicting the answers to questions (IBM invested a lot more than that to achieve this win, as detailed in Chapter 6). Organizations are literally keeping kids in school, keeping the lights on, and keeping crime down with predictive analytics (PA). And success is its own reward when analytics wins a political election, a baseball championship, or . . . did I mention managing a financial portfolio?

Black-box trading—driving financial trading decisions automatically with a machine—is the holy grail of data-driven decision making. It’s a black box into which current financial environmental conditions are fed, with buy/hold/sell decisions spit out the other end. It’s black (i.e., opaque) because you don’t care what’s on the inside, as long as it makes good decisions. When working, it trumps any other conceivable business proposal in the world: Your computer is now a box that turns electricity into money.

And so with the launch of his stock trading system, John Elder took on his own personal grand challenge. Even if stock market prediction would represent a giant leap for mankind, this was no small step for John himself. It’s an occasion worthy of mixing metaphors. By putting all his eggs into one analytical basket, John was taking a healthy dose of his own medicine.

Before continuing with the story of John’s blast-off, let’s establish how launching a predictive system works, not only for black-box trading but across a multitude of applications.

GOING LIVE

Learning from data is virtually universally useful. Master it and you’ll be welcomed nearly everywhere!

—John Elder

New groundbreaking stories of PA in action are pouring in. A few key ingredients have opened these floodgates:

- wildly increasing loads of data;
- cultural shifts as organizations learn to appreciate, embrace, and integrate predictive technology;
- improved software solutions to deliver PA to organizations.

But this flood built up its potential in the first place simply because predictive technology boasts an inherent generality—there are just so many conceivable ways to make use of it. Want to come up with your own new innovative use for PA? You need only two ingredients.

EACH APPLICATION OF PA IS DEFINED BY:

1. **What's predicted:** the kind of behavior (i.e., action, event, or happening) to predict for each individual, stock, or other kind of element.
2. **What's done about it:** the decisions driven by prediction; the action taken by the organization in response to or informed by each prediction.

Given its open-ended nature, the list of application areas is so broad and the list of example stories is so long that it presents a minor data-management challenge in and of itself! So I placed this big list (182 examples total) into nine tables in the center of this book. Take a flip through to get a feel for just how much is going on. That's the sexy part—it's the "centerfold" of this book. The Central Tables divulge cases of predicting: stock prices, risk, delinquencies, accidents, sales, donations, clicks, cancellations, health problems, hospital admissions, fraud, tax evasion, crime, malfunctions, oil flow, electricity outages, approvals for government benefits, thoughts, intention, answers, opinions, lies, grades, dropouts, friendship, romance, pregnancy, divorce, jobs, quitting, wins, votes, and more. The application areas are growing at a breakneck pace.

Within this long list, the quintessential application for business is the one covered in the Introduction for mass marketing:

PA APPLICATION: TARGETING DIRECT MARKETING

1. **What's predicted:** Which customers will respond to marketing contact.
2. **What's done about it:** Contact customers more likely to respond.

As we saw, this use of PA illustrates *The Prediction Effect*.

The Prediction Effect: *A little prediction goes a long way.*

Let's take a moment to see how straightforward it is to calculate the sheer value resulting from The Prediction Effect. Imagine you have a company with a mailing list of a million prospects. It costs \$2 to mail to each one, and you have observed that one out of 100 of them will buy your product (i.e., 10,000 responses). You take your chances and mail to the entire list.

If you profit \$220 for each rare positive response, then you pocket:

$$\begin{aligned} \text{Overall profit} &= \text{Revenue} - \text{Cost} \\ &= (\$220 \times 10,000 \text{ responses}) - (\$2 \times 1 \text{ million}) \end{aligned}$$

Whip out your calculator—that's \$200,000 profit. Are you happy yet? I didn't think so.

If you are new to the arena of direct marketing (welcome!), you'll notice we're playing a kind of wild numbers game, amassing great waste, like one million monkeys chucking darts across a chasm in the general direction of a dartboard. As turn-of-the-century marketing pioneer John Wanamaker famously put it, "Half the money I spend on advertising is wasted; the trouble is I don't know which half." The bad news is that it's actually more than half; the good news is that PA can learn to do better.

A FAULTY ORACLE EVERYONE LOVES

The first step toward predicting the future is admitting you can't.

—Stephen Dubner, Freakonomics Radio, March 30, 2011

The “prediction paradox”: The more humility we have about our ability to make predictions, the more successful we can be in planning for the future.

—Nate Silver, *The Signal and the Noise: Why So Many Predictions Fail—but Some Don't*

Your resident “oracle,” PA, tells you which customers are most likely to respond. It earmarks a quarter of the entire list and says, “These folks are three times more likely to respond than average!” So now you have a short list of 250,000 customers of whom 3 percent will respond—7,500 responses.

Oracle, shmoracle! These predictions are seriously inaccurate—we still don't have strong confidence when contacting any one customer, given this measly 3 percent response rate. However, the overall IQ of your dart-throwing monkeys has taken a real boost. If you send mail to only this short list then you profit:

$$\begin{aligned}\text{Overall profit} &= \text{Revenue} - \text{Cost} \\ &= (\$220 \times 7,500 \text{ responses}) - (\$2 \times 250,000)\end{aligned}$$

That's \$1,150,000 profit. You just improved your profit 5.75 times over by mailing to *fewer* people (and, in so doing, expending fewer trees). In particular, you predicted who wasn't worth contacting and simply left them alone. Thus you cut your costs by three-quarters in exchange for losing only one-quarter of sales. That's a deal I'd take any day.

It's not hard to put a value on prediction. As you can see, even if predictions themselves are generated from sophisticated mathematics, it takes only simple arithmetic to roll up the plethora of predictions—some accurate, and others not so much—and reveal the aggregate bottom-line effect. This isn't just some abstract notion; The Prediction Effect means business.

PREDICTIVE PROTECTION

Thus, value has emerged from just a little predictive insight, a small prognostic nudge in the right direction. It's easy to draw an analogy to science fiction, where just a bit of supernatural foresight can go a long way. Nicolas Cage kicks some serious bad-guy butt in the movie *Next*, based on a story by Philip K. Dick. His weapon? Pure prognostication. He can see the future, but only two minutes ahead. It's enough prescience to do some damage. An unarmed civilian with a soft heart and the best of intentions, he winds up marching through something of a war zone, surrounded by a posse of heavily armed FBI agents who obey his every gesture. He sees the damage of every booby trap, sniper, and mean-faced grunt before it happens and so can command just the right moves for this Superhuman Risk-Aversion Team, avoiding one calamity after another.

In a way, deploying PA makes a Superhuman Risk-Aversion Team of the organization just the same. Every decision an organization makes, each step it takes, incurs risk. Imagine the protective benefit of foreseeing each pitfall so that it may be avoided—each criminal act, stock value decline, hospitalization, bad debt, traffic jam, high school dropout . . . and each ignored marketing brochure that was a waste to mail. *Organizational risk management*, traditionally the act of defending against singular, macrolevel incidents like the crash of an aircraft or an economy, now broadens to fight a myriad of microlevel risks.

Hey, it's not all bad news. We win by foreseeing good behavior as well, since it often signals an opportunity to gain. The name of the game is "Predict 'n' Pounce" when it pops up on the radar that a customer is likely to buy, a stock value is likely to increase, a voter is likely to swing, or the apple of one's online dating eye is likely to reciprocate.

A little glimpse into the future gives you power because it gives you options. In some cases the obvious decision is to act in order to avert what may not be inevitable, be it crime, loss, or sickness. On the positive side, in the case of foreseeing demand, you act to exploit it. Either way, prediction serves to drive decisions.

Let's turn to a real case, a \$1 million example.

A SILENT REVOLUTION WORTH A MILLION

When an organization goes live with PA, it unleashes a massive army, but it's an army of ants. These ants march out to the front lines of an organization's operations, the places where there's contact with the likes of customers, students, or patients—the people served by the organization. Within these interactions, the ant army, guided by predictions, improves millions of small decisions. The process goes largely unnoticed, under the radar . . . until someone bothers to look at how it's adding up. The improved decisions may each be ant-sized, relatively speaking, but there are so many that they come to a powerful net effect.

In 2005, I was digging in the trenches, neck deep in data for a client who wanted more clicks on their website. To be precise, they wanted more clicks on their sponsors' ads. This was about the money—more clicks, more money. The site had gained tens of millions of users over the years, and within just several months' worth of tracking data that they handed me, there were 50 million rows of learning data—no small treasure trove from which to learn to predict . . . *clicks*.

Advertising is an inevitable part of media, be it print, television, or your online experience. Benjamin Franklin forgot to include it when he proclaimed, “In this world nothing can be said to be certain, except death and taxes.” The flagship Internet behemoth Google credits ads as its greatest source of revenue. It's the same with Facebook.

But on this website, ads told a slightly different story than usual, which further amplified the potential win of predicting user clicks. The client was a leading student grant and scholarship search service, with one in three college-bound high school seniors using it: an arcane niche, but just the one over which certain universities and military recruiters were drooling. One ad for a university included a strong pitch, naming itself “America's leader in creative education” and culminating with a button that begged to be clicked: “Yes, please have someone from the Art Institute's Admissions Office contact me!” And you won't be surprised to hear that creditors were also placing ads, at the ready to provide these students another source of funds: loans. The sponsors would pay up to \$25 per lead—for each

would-be recruit. That's good compensation for one little click of the mouse. What's more, since the ads were largely relevant to the users, closely related to their purpose on the website, the response rates climbed up to an unusually high 5 percent. So this little business, owned by a well-known online job-hunting firm, was earning well. Any small improvement meant real revenue.

But improving ad selection is a serious challenge. At certain intervals, users were exposed to a full-page ad, selected from a pool of 291 options. The trick is selecting the best one for each user. The website currently selected which ad to show based simply on the revenue it generated on average, with no regard to the particular user. The universally strongest ad was always shown first. Although this tactic forsakes the possibility of matching ads to individual users, it's a formidable champion to unseat. Some sponsor ads, such as certain universities, paid such a high bounty per click, and were clicked so often, that showing any user a less powerful ad seemed like a crazy thing to consider, since doing so would risk losing currently established value.

THE PERILS OF PERSONALIZATION

By trusting predictions in order to customize for the individual, you take on risk. A predictive system boldly proclaims, "Even though ad A is so strong overall, for this particular user it is worth the risk of going with ad B." For this reason, most online ads are not personalized for the individual user—even Google's AdWords, which allows you to place textual ads alongside search results as well as on other Web pages, determines which ad to display by Web page context, the ad's click rate, and the advertiser's bid (what it is willing to pay for a click). It is not determined by anything known or predicted about the particular viewer who is going to actually see the ad.

But weathering this risk carries us to a new frontier of customization. For business, it promises to "personalize!," "increase relevance!," and "engage one-to-one marketing!" The benefits reach beyond personalizing marketing treatment to customizing the individual treatment of patients and suspected criminals as well. During a speech about satisfying our widely varying preferences in choice of spaghetti sauce—chunky? sweet? spicy?—Malcolm

Gladwell said, “People . . . were looking for . . . universals, they were looking for one way to treat all of us[;] . . . all of science through the nineteenth century and much of the twentieth was obsessed with universals. Psychologists, medical scientists, economists were all interested in finding out the rules that govern the way all of us behave. But that changed, right? What is the great revolution of science in the last 10, 15 years? It is the movement from the search for universals to the understanding of variability. Now in medical science we don’t want to know . . . just how cancer works; we want to know how your cancer is different from my cancer.”

From medical issues to consumer preferences, individualization trumps universals. And so it goes with ads:

PA APPLICATION: PREDICTIVE ADVERTISEMENT TARGETING

- 1. What’s predicted:** Which ad each customer is most likely to click.
- 2. What’s done about it:** Display the best ad (based on the likelihood of a click as well as the bounty paid by its sponsor).

I set up PA to perform ad targeting for my client, and the company launched it in a head-to-head, champion/challenger competition to the death against their existing system. The loser would surely be relegated to the bin of second-class ideas that just don’t make as much cash. To prepare for this battle, we armed PA with powerful weaponry. The predictions were generated from machine learning across 50 million learning cases, each depicting a microlesson from history of the form, “User Mary was shown ad A and she did click it” (a positive case) or “User John was shown ad B and he did not click it” (a negative case).

The learning technology employed to pick the best ad for each user was a Naïve Bayes model. Rev. Thomas Bayes was an eighteenth-century mathematician, and the “Naïve” part means that we take a very smart man’s ideas and compromise them in a way that simplifies yet makes their application feasible, resulting in a practical method that’s often considered good enough at prediction and scales to the task at hand. I went with this method for its relative simplicity, since in fact I needed to generate 291 such models, one for each ad. Together, these models predict which ad a user is most likely to click on.

DEPLOYMENT'S DETOURS AND DELAYS

As with a rocket ship, launching PA looks great on paper. You design and construct the technology, place it on the launchpad, and wait for the green light. But just when you're about to hit "go," the launch is scrubbed. Then delayed. Then scrubbed again. The Wright brothers and others, galvanized by the awesome promise of a newly discovered wing design that generates lift, endured an uncharted, rocky road, faltering, floundering, and risking life and limb until all the kinks were out.

For ad targeting and other real-time PA deployments, predictions have got to zoom in at warp speed in order to provide value. Our online world tolerates no delay when it's time to choose which ad to display, determine whether to buy a stock, decide whether to authorize a credit card charge, recommend a movie, filter an e-mail for viruses, or answer a question on *Jeopardy!* A real-time PA solution must be directly integrated into operational systems, such as websites or credit card processing facilities. If you are newly integrating PA within an organization, this can be a significant project for the software engineers, who often have their hands full with maintenance tasks just to keep the business operating normally. Thus, the *deployment* phase of a PA project takes much more than simply receiving a nod from senior management to go live: It demands major construction. By the time the programmers deployed my predictive ad selection system, the data over which I had tuned it was already about 11 months old. Were the facets of what had been learned still relevant almost one year later, or would prediction's power peter out?

IN FLIGHT

*This is Major Tom to Ground Control
I'm stepping through the door
And I'm floating in a most peculiar way . . .*

—"Space Oddity" by David Bowie

Once launched, PA enters an eerie, silent waiting period, like you're floating in orbit and nothing is moving. But the fact is, in a low orbit around Earth you're actually screaming along at over 14,000 miles per hour. Unlike the drama of launching a rocket or erecting a skyscraper, the launch of PA is a

relatively stealthy maneuver. It goes live, but daily activities exhibit no immediately apparent change. After the ad-targeting project's launch, if you checked out the website, it would show you an ad as usual, and you could wonder whether the system made any difference in this one choice. This is what computers do best. They hold the power to silently enact massive procedural changes that often go uncredited, since most aren't directly witnessed by any one person.

But, under the surface, a sea change is in play, as if the entire ocean has been reconfigured. You actually notice the impact only when you examine an aggregated report.

In my client's deployment, predictive ad selection triumphed. The client conducted a head-to-head comparison, selecting ads for half the users with the existing champion system and the other half with the new predictive system, and reported that the new system generated at least 3.6 percent more revenue, which amounts to \$1 million every 19 months, given the rate at which revenue was already coming in. This was for the website's full-page ads only; many more (smaller) ads are embedded within functional Web pages, which could potentially also be boosted with a similar PA project.

No new customers, no new sponsors, no changes to business contracts, no materials or computer hardware needed, no new full-time employees or ongoing effort—solely an improvement to decision making was needed to generate cold, hard cash. In a well-oiled, established system like the one my client had, even a small improvement of 3.6 percent amounts to something substantial. The gains of an incremental tweak can be even more dramatic: In the insurance business, one company reports that PA saves almost \$50 million annually by decreasing its loss ratio by *half a percentage point*.

So how did these models predict each click?

ELEMENTARY, MY DEAR: THE POWER OF OBSERVATION

Just like Sherlock Holmes drawing conclusions by sizing up a suspect, prediction comes of astute observation: What's known about each individual provides a set of clues about what he or she may do next. The chance a user will click on a certain ad depends on all sorts of elements, including the individual's current school year, gender, and e-mail domain

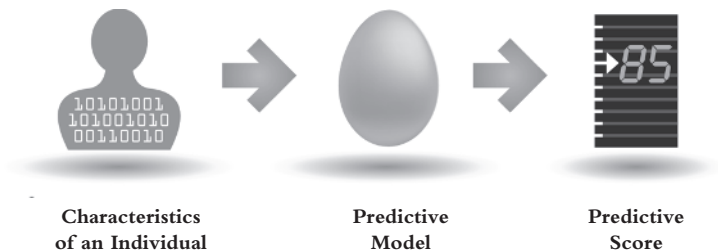
(Hotmail, Yahoo, Gmail, etc.); the ratio of the individual’s SAT written-to-math scores (is the user more a verbal person or more a math person?), and on and on.

In fact, this website collected a wealth of information about its users. To find out which grants and scholarships they’re eligible for, users answer dozens of questions about their school performance, academic interests, extracurricular activities, prospective college majors, parents’ degrees, and more. So the table of learning data was long (at 50 million examples) and was also wide, with each row holding all the information known about the user at the moment the person viewed an ad.

It can sound like a tall order: *harnessing millions of examples in order to learn how to incorporate the various factoids known about each individual so that prediction is possible*. But we can break this down into a couple of parts, and suddenly it gets much simpler. Let’s start with the contraption that makes the predictions, the electronic Sherlock Holmes that knows how to consider all these factors and roll them up into a single prediction for the individual.

Predictive model—*a mechanism that predicts a behavior of an individual, such as click, buy, lie, or die. It takes characteristics of the individual as input and provides a predictive score as output. The higher the score, the more likely it is that the individual will exhibit the predicted behavior.*

A predictive model (depicted throughout this book as a “golden” egg, albeit in black and white) scores an individual:



A predictive model is the means by which the attributes of an individual are factored together for prediction. There are many ways to do this. One is to weigh each characteristic and then add them up—perhaps females boost their score by 33.4, Hotmail users decrease their score by 15.7, and so on.

Each element counts toward or against the final score for that individual. This is called a *linear model*, generally considered quite simple and limited, although usually much better than nothing.

Other models are composed of *rules*, like this real example:

IF the individual
is still in high school
AND
expects to graduate college within three years
AND
indicates certain military interest
AND
has not been shown this ad yet
THEN the probability of clicking on the ad for the Art Institute is
13.5 percent.

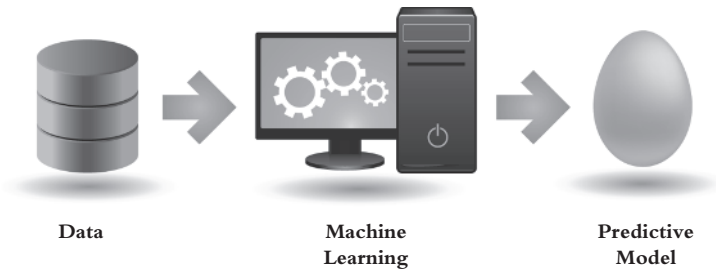
This rule is a valuable find, since the overall probability of responding to the Art Institute’s ad is only 2.7 percent, so we’ve identified a pocket of avid clickers, relatively speaking.

It is interesting that those who have indicated a military interest are more likely to show interest in the Art Institute. We can speculate, but it’s important not to assume there is a *causal* relationship. For example, it may be that people who complete more of their profile are just more likely to click in general, across all kinds of ads.

Various types of models compete to make the most accurate predictions. Models that combine a bunch of rules like the one just shown are—relatively speaking—on the simpler side. Alternatively, we can go more “supermath” on the prediction problem, employing complex formulas that predict more effectively but are almost impossible to understand by human eyes.

But all predictive models share the same objective: They consider the various factors of an individual in order to derive a single predictive score for that individual. This score is then used to drive an organizational decision, guiding which action to take.

Before using a model, we've got to build it. Machine learning builds the predictive model:



Machine learning crunches data to build the model, a brand-new prediction machine. The model is the product of this learning technology—it is itself the very thing that has been learned. For this reason, machine learning is also called *predictive modeling*, which is a more common term in the commercial world. If deferring to the older metaphorical term *data mining*, the predictive model is the unearthed gem.

Predictive modeling generates the entire model from scratch. All the model's math, weights, or rules are created automatically by the computer. The machine learning process is designed to accomplish this task, to mechanically develop new capabilities from data. This automation is the means by which PA builds its predictive power.

The hunter returns back to the tribe, proudly displaying his kill. So, too, a data scientist posts her model on the bulletin board near the company ping-pong table. The hunter hands over the kill to the cook, and the data scientist cooks up her model, translates it to a standard computer language, and e-mails it to an engineer for integration. A well-fed tribe shows the love; a psyched executive issues a bonus.

TO ACT IS TO DECIDE

Knowing is not enough; we must act.

—Johann Wolfgang von Goethe

Once you develop a model, don't pat yourself on the back just yet. Predictions don't help unless you do something about them. They're just thoughts, just

ideas. They may be astute, brilliant gems that glimmer like the most polished of crystal balls, but displaying them on a shelf gains you nothing—they just sit there and look smart.

Unlike a report sitting dormant on the desk, PA leaps out of the lab and takes action. In this way, it stands above other forms of analysis, data science, and data mining. It desires deployment and loves to be launched—because, in what it foretells, it mandates movement.

The predictive score for each individual directly informs the decision of what action to take with that individual. Doctors take a second look at patients predicted to be readmitted, and service agents contact customers predicted to cancel. Predictive scores issue imperatives to *mail, call, offer a discount, recommend a product, show an ad, expend sales resources, audit, investigate, inspect for flaws, approve a loan, or buy a stock*. By acting on the predictions produced by machine learning, the organization is now applying what's been learned, modifying its everyday operations for the better.

To make this point, we have mangled the English language. Proponents like to say that PA is *actionable*. Its output directly informs actions, commanding the organization about what to do next. But with this use of vocabulary, industry insiders have stolen the word *actionable*, which originally meant *worthy of legal action* (i.e., “sue-able”), and morphed it. They did so because they're tired of seeing sharp-looking reports that provide only a vague, unsure sense of direction.

With this word's new meaning established, “your fly is unzipped” is *actionable* (it is clear what to do—you can and should take action to remedy), but “you're going bald” is not (there's no cure; nothing to be done). Better yet, “I predict you will buy these button-fly jeans and this snazzy hat” is actionable to a salesperson.

Launching PA into action delivers a critical new edge in the competitive world of business. One sees massive commoditization taking place today as the faces of corporations appear to blend together. They all seem to sell pretty much the same thing and act in pretty much the same ways. To stand above the crowd, where can a company turn?

As Thomas Davenport and Jeanne Harris put it in *Competing on Analytics: The New Science of Winning*, “At a time when companies in many industries offer similar products and use comparable technology, high-performance business

processes are among the last remaining points of differentiation.” Enter PA. Survey results have in fact shown that “a tougher competitive environment” is by far the strongest reason why organizations adopt this technology.

But while the launch of PA brings real change, it can also wreak havoc by introducing new risk. With this in mind, we now return to John’s story.

A PERILOUS LAUNCH

Dr. John Elder bet it all on a predictive model. He concocted it in the lab, packed it into a black box, and unleashed it on the stock market. Some people make their own bed in which they must then passively lie. But John had climbed way up high to take a leap of faith. Diving off a mountaintop with newly constructed, experimental wings, he wondered how long it might take before he could be sure he was flying rather than crashing.

The risks stared John in the face. His and his wife’s full retirement savings were in the hands of an experimental device, launched into oblivion and destined for one of the same two outcomes achieved by every rocket: glory or mission failure. Discovering profitable market patterns that sustain is the mission of thousands of traders operating in what John points out is a brutally competitive environment; doing so automatically with machine learning is the most challenging of ambitions, considered impossible by many. It doesn’t help that a stock market scientist is completely on his own, since work in this area is shrouded in secrecy, leaving virtually no potential to learn from the successes and failures of others. Academics publish, marketers discuss, but quants hide away in their Batcaves. What can look great on paper might be stricken with a weakness that destroys or an error that bankrupts. John puts it plainly: “Wall Street is the hardest data mining problem.”

The evidence of danger was palpable, as John had recently uncovered a crippling flaw in an existing predictive trading system and personally escorted it to its grave. Opportunity had come knocking on the door of a small firm called Delta Financial in the form of a black-box trading system purported to predict movements of the Standard & Poor’s (S&P) 500 with 70 percent accuracy. Built by a proud scientist, the system promised to make millions, so stakeholders were flying around all dressed up in suits, actively lining up

investors prepared to place a huge bet. Among potential early investors, Delta was leading the way for others, taking a central, influential role. The firm was known for investigating and championing cutting-edge approaches, weathering the risk inherent to innovation. As a necessary precaution, Delta sought to empirically validate this system. The firm turned to John, who was consulting for them on the side while pursuing his doctorate at the University of Virginia in Charlottesville. John's work for Delta often involved inspecting, and sometimes debunking, black-box trading systems.

How do you prove a machine is broken if you're not allowed to look inside it? Healthy skepticism bolstered John's resolve, since the claimed 70 percent accuracy raised red flags as quite possibly too darn good to be true. But he was not granted access to the predictive model. With secrecy reigning supreme, the protocol for this type of audit dictated that John receive only the numerical results, along with a few adjectives that described its design: *new, unique, powerful!* With meager evidence, John sought to prove a crime he couldn't even be sure had been committed.

Before each launch, organizations establish confidence in PA by "predicting the past" (aka backtesting). The predictive model must prove itself on historical data before its deployment. Conducting a kind of simulated prediction, the model evaluates across data from last week, last month, or last year. Feeding on input that could only have been known at a given time, the model spits out its prediction, which then matches against what we now already know took place thereafter. Would the S&P 500 go down or up on March 21, 1991? If the model gets this retrospective question right, based only on data available by March 20, 1991 (the day just before), we have evidence the model works. These retrospective predictions—without the manner in which they had been derived—were all John had to work with.

HOUSTON, WE HAVE A PROBLEM

Even the most elite of engineers commit the most mundane and costly of errors. In late 1998, NASA launched the Mars Climate Orbiter on a daunting nine-month trip to Mars, a mission that fewer than half the world's launched probes headed for that destination have completed successfully. This \$327.6 million

calamity crashed and burned, due not to the flip of fate's coin, but rather a simple snafu. The spacecraft came too close to Mars and disintegrated in its atmosphere. The source of the navigational bungle? One system expected to receive information in metric units (newton-seconds), but a computer programmer for another system had it speak in English imperial units (pound-seconds). Oops.

John stared at a screen of numbers, wondering if anything was wrong and, if so, whether he could find it. From the long list of impressive—yet retrospective—predictions, he plainly saw the promise of huge profits that had everyone involved so excited. If he proved there was a flaw, vindication; if not, lingering uncertainty. The task at hand was to reverse engineer: Given the predictions the system generated, could he infer how it worked under the hood, essentially eking out the method in its madness? This was ironic, since all predictive modeling is a kind of reverse engineering to begin with. Machine learning starts with the data, an encoding of things that have happened, and attempts to uncover patterns that generated or explained the data in the first place. John was attempting to deduce what the other team had deduced. His guide? Informal hunches and ill-informed inferences, each of which could be pursued only by way of trial and error, testing each hypothetical mess-up he could dream up by programming it by hand and comparing it to the retrospective predictions he had been given.

His perseverance finally paid off: John uncovered a true flaw, thereby flinging back the curtain to expose a flustered Wizard of Oz. It turned out that the prediction engine committed the most sacrilegious of cheats by looking at the one thing it must not be permitted to see. It had looked at the future. The battery of impressive retrospective predictions weren't true predictions at all. Rather, they were based in part on a three-day average calculated across yesterday, today . . . and tomorrow. The scientists had probably intended to incorporate a three-day average leading up to today, but had inadvertently shifted the window by a day. Oops. This crippling bug delivered the dead-certain prognosis that this predictive model would not perform well if deployed into the field. Any prediction it would generate today could not incorporate the very thing it was designed to foresee—tomorrow's stock price—since, well, it isn't known yet. So, if foolishly deployed, its accuracy could never match the exaggerated performance falsely demonstrated across

the historical data. John revealed this bug by reverse engineering it. On a hunch, he handcrafted a method with the same type of bug and showed that its predictions closely matched those of the trading system.

A predictive model will sink faster than the *Titanic* if you don't seal all its "time leaks" before launch. But this kind of "leak from the future" is common, if mundane. Although core to the very integrity of prediction, it's an easy mistake to make, given that each model is backtested over historical data for which prediction is not, strictly speaking, possible. The relative future is always readily available in the testing data, easy to inadvertently incorporate into the very model trying to predict it. Such temporal leaks achieve status as a commonly known gotcha among PA practitioners. If this were an episode of *Star Trek*, our beloved, hypomaniac engineer Scotty would be screaming, "Captain, we're losing our temporal integrity!"

It was with no pleasure that John delivered the disappointing news to his client, Delta Financial: He had debunked the system, essentially exposing it as inadvertent fraud. High hopes were dashed as another fairy tale bit the dust, but gratitude quickly ensued as would-be investors realized they'd just dodged a bullet. The wannabe inventor of the system suffered dismay but was better off knowing now; it would have hit the fan much harder postlaunch, possibly including prosecution for fraud, even if inadvertently committed. The project was aborted.

THE LITTLE MODEL THAT COULD

Even the young practitioner that he was, John was a go-to data man for entrepreneurs in black-box trading. One such investor moved to Charlottesville, but only after John Elder, PhD, new doctorate degree in hand, had just relocated to Houston in order to continue his academic rite of passage with a postdoc research position at Rice University. He'd left quite an impression back in Charlottesville, though; people in both the academic and commercial sectors alike referred the investor to John. Despite John's distance, the investor hired him to prepare, launch, and monitor a new black-box mission remotely from Houston. It seemed as good a place as any for the project's Mission Control.

And so it was time for John to move beyond the low-risk role of evaluating other people's predictive systems and dare to build one of his own. Over several months, he and a small team of colleagues built upon core insights from the investor and produced a new, promising black-box trading model. John was champing at the bit to launch it and put it to the test. All the stars were aligned for liftoff except one: The money people didn't trust it yet.

There was good reason to believe in John. Having recently completed his doctorate degree, he was armed with a fresh, talented mind, yet had already gained an impressively wide range of data-crunching problem-solving experience. On the academic side, his PhD thesis had broken records among researchers as the most efficient way to optimize for a certain broad class of system engineering problems (machine learning is itself a kind of optimization problem). He had also taken on predicting the species of a bat from its echolocation signals (the chirps bats make for their radar). And in the commercial world, John's pregrad positions had dropped him right into the thick of machine learning systems that steer for aerospace flight and that detect cooling pipe cracks in nuclear reactors, not to mention projects for Delta Financial looking over the shoulders of other black-box quants.

And now John's latest creation absolutely itched to be deployed. Backtesting against historical data, all indications whispered confident promises for what this thing could do once set in motion. As John puts it, "A slight pattern emerged from the overwhelming noise; we had stumbled across a persistent pricing inefficiency in a corner of the market, a small edge over the average investor, which appeared repeatable." Inefficiencies are what traders live for. A perfectly efficient market can't be played, but if you can identify the right imperfection, it's payday.

PA APPLICATION: BLACK-BOX TRADING

- 1. What's predicted:** Whether a stock will go up or down.
- 2. What's done about it:** Buy stocks that will go up; sell those that will go down.

John could not get the green light. As he strove to convince the investor, cold feet prevailed. It appeared they were stuck in a stalemate. After all, this

guy might not get past his jitters until he could see the system succeed, yet it couldn't succeed while stuck on the launchpad. The time was now, as each day marked lost opportunity.

After a disconcerting meeting that seemed to go nowhere, John went home and had a sit-down with his wife, Elizabeth. What supportive spouse could possibly resist the seduction of her beloved's ardent excitement and strong belief in his own abilities? She gave him the go-ahead to risk it all, a move that could threaten their very home. But he still needed buy-in from one more party.

Delivering his appeal to the client investor raised questions, concerns, and eyebrows. John wanted to launch with his own personal funds, which meant no risk whatsoever to the client and would resolve any doubts by field-testing John's model. But this unorthodox step would be akin to the dubious choice to act as one's own defense attorney. When an individual is without great personal means, this kind of thing is often frowned upon. It conveys overconfident, foolish brashness. Even if the client wanted to truly believe, it would be another thing to expect the same from coinvestors who hadn't gotten to know and trust John. But with every launch, proponents gamble something fierce. John had set the rules for the game he'd chosen to play.

He received his answer from the investor: "Go for it!" This meant there was nothing to prevent moving forward. It could have also meant the investor was prepared to write off the project entirely, feeling there was nothing left to lose.

HOUSTON, WE HAVE LIFTOFF

Practitioners of PA often put their own professional lives a bit on the line to push forward, but this case was extreme. Like baseball's Billy Beane of the Oakland A's, who literally risked his entire career to deploy and field-test an analytical approach to team management, John risked everything he had. It was early 1994, and John's individual retirement account (IRA) amounted to little more than \$40,000. He put it all in.

“Going live with black-box trading is really exciting and really scary,” says John. “It’s a roller coaster that never stops. The coaster takes on all these thrilling ups and downs, but with a very real chance it could go off the rails.”

As with baseball, he points out, slumps aren’t slumps at all—they’re inevitable statistical certainties. Each one leaves you wondering, “Is this falling feeling part of a safe ride, or is something broken?” A key component to his system was a cleverly designed means to detect real quality, a measure of system integrity that revealed whether recent success had been truly deserved or had come about just due to dumb luck.

From the get-go, the predictive engine rocked. It increased John’s assets at a rate of 40 percent per year, which meant that after two years his money had doubled.

The client investor was quickly impressed and soon put in a couple of million dollars himself. A year later, the predictive model was managing a \$20 million fund across a group of investors, and eventually the investment pool increased to a few hundred million dollars. With this much on tap, every win of the system was multiplicatively magnified.

No question about it: All involved relished this fiesta, and the party raged on and on, continuing almost nine years, consistently outperforming the overall market all along. The system chugged, autonomously trading among a dozen market sectors such as technology, transportation, and healthcare. John says the system “beat the market each year and exhibited only two-thirds its standard deviation—a home run as measured by risk-adjusted return.”

But all good things must come to an end, and just as John had talked his client up, he later had to talk him down. After nearly a decade, the key measure of system integrity began to decline. John was adamant that they were running on fumes, so with little ceremony the entire fund was wound down. The system was halted in time, before catastrophe could strike. In the end, all the investors came out ahead.

A PASSIONATE SCIENTIST

The early success of this streak had quickly altered John’s life. Once the project was cruising, he had begun supporting his rapidly growing family

with ease. The project was taking only a couple of John's hours each day to monitor, tweak, and refresh what was a fundamentally stable, unchanging method within the black box. What's a man to do? Do you put your feet up and sip wine indefinitely, with the possible interruption of family trips to Disney World? After all, John had thus far always burned the candle at both ends out of financial necessity, with summer jobs during college, part-time work during graduate school, and this black-box project, which itself had begun as a moonlighting gig during his postdoc. Or do you follow the logical business imperative: Pounce on your successes, using all your free bandwidth to find ways to do more of the same?

John's passion for the craft transcended these self-serving responses to his good fortune. That is to say, he contains the spirit of the geek. He jokes about the endless insatiability of his own appetite for the stimulation of fresh scientific challenges. He's addicted to tackling something new. There is but one antidote: a growing list of diverse projects. So, two years into the stock market project, he wrapped up his postdoc, packed up his family, and moved back to Charlottesville to start his own data mining company.

And so John launched Elder Research, now the largest predictive analytics services firm (pure play) in North America. A narrow focus is key to the success of many businesses, but Elder Research's advantage is quite the opposite: its diversity. The company's portfolio reaches far beyond finance to include all major commercial sectors and many branches of government. John has also earned a top-echelon position in the industry. He coauthors massive textbooks, frequently chairs or keynotes at Predictive Analytics World conferences, takes cameos as a university professor, and served five years as a presidential appointee on a national security technology panel.

LAUNCHING PREDICTION INTO INNER SPACE

With stories like John's coming to light, organizations are jumping on the PA bandwagon. One such firm, a mammoth international organization, focuses the power of prediction introspectively, casting PA's keen gaze on its own employees. Read on to witness the windfall and the fallout when scientists dare to ask: Do people like being predicted?