CIO

WRAPSHEET

[ANALYTICS IN ACTION]

10 critical questions to ask analytics vendors *(before you buy)*

PAGE 2

PLUS

8 ways to make the most of **customer data**

PAGE 17

THINKSTOCK

10 critical questions to ask analytics vendors *(before you buy)*

Analytics purchase decisions are littered with opportunities for missteps. These 10 questions will help ensure you get off on the right foot.

BY THOR OLAVSRUD

Choosing the right analytics platform is essential. After all, the product you choose will be helping leaders make critical business decisions for years to come – that's if you get it right.

Here are 10 questions you should ask analytics vendors early in the buying process.

1 Can I answer questions?

It sounds very basic, but the capability to ask and answer new questions simply, not just a set of predefined questions, is table stakes.

"If you're confined to a preset view of a slice of your data and it requires going to your IT department or doing complex scripting to ask a new question, you're not going to get a lot of value out of your analytics," says Ellie Fields, vice president of Product Marketing, Tableau Software. "You need to live in a world where your questions are not predefined."

2 What is your experience in my industry or with my specific business issues?

Expertise in analytics is not enough. You need domain expertise to get the most out of your analytics products and services. Does your vendor have a set of industry-specific data models (e.g., for consumer products, banking, insurance companies, etc.)?

"To develop analytics or visualization or products specific to industry you need to be in tune with business issues," says Mark Shilling, principal, Deloitte Consulting and leader of its Analytics and Information Management Service Line. "You cannot deliver wireframes with KPIs and an underlying data model to support it if you do not know what you are measuring and why."

3 Will people want to use it?

Your analytics solutions need to be able to answer your questions, but using them has to be an enjoyable experience. As anyone who's deployed new products knows, getting the users to actually use it is a big part of the battle.

"Every enterprise system throughout history has been bedeviled by people not wanting to use the system," Fields says. "It has to be useful for answering questions. You want them to dig in and feel that asking that question is a really good experience. People will spend more time on it if it's easy, intuitive and they feel like they're getting something out of it."

4 What is your tool's capacity?

What number of concurrent users does it support? What about numbers of users, accounts and sites? What's the response time?

"This is tied to performance at scale," says Deloitte's Shilling. "You need to make sure that the tool can accommodate the size and pace of business growth."

5 Does it work with all my data?

Organizations have data in many places: cloud, on-premises, data warehouses, Excel spreadsheets, Web logs and more. Does this analytics tool work with everything you have? It may be structured data, unstructured data or semi-structured data.

"Some tools require you to move data before you can analyze it, or require you to get it into their proprietary system," Tableau's Fields says. "What data do I work with today and what data do I think I'll be working with in a couple of years? Does this product support all of that data?"

"Can you perform analysis for text, audio and voice? Not every vendor crosses into the domain of big data," Shilling adds. "Some are specifically building capabilities tied to a data set."

6 Do you have tools that can address analytical needs of different business functions?

For instance, can your tools provide a 360-degree view of a customer, supply-chain planning, distribution effectiveness and scenario planning and modeling? Shilling says the answer to this question shows the breadth of usage across domains.

7 Is your pricing model based on outcomes and value?

Accessing data is only the first step in an analytics journey. Successful analytics projects are about outcomes that deliver value.

"Have you entered into value-based billing agreements before? 'VBB' — what more is there to tell? Every software company today is being measured against open source or free. So to justify why someone would plunk down capital investment you need to have pricing tied to outcomes," Shilling says.

8 What customers are using your tools and how?

Can you provide references?

Trust, but verify. No executive buys a product to be a pioneer, Shilling says. "It needs to be proven technology and that is what a reference is about," he adds. "Validation."

9 What security features are embedded in your software?

How can you guarantee the security of personal information?

Data security and data governance are essential, especially in regulated industries, where it's mandated by law. Personally identifiable information (PII) and protected health information (PHI) in regulated industries like health care and life sciences need to be obfuscated, Shilling notes.

10 What is your product roadmap?

How do you assist your clients with version updates and migration to new products? How do you manage the migration and cutover processes? Do you have a program to help train users?

Your analytics solutions will hopefully serve you for a long time to come. As with other software and solutions, analytics solutions won't remain static. Shilling says you need to understand your vendor's software development life cycle, product enhancement and change management. ■

THINKSTOCK

CIOs have to learn the **new math** of analytics

Today's data-driven business runs on the almighty algorithm. But if you're not careful, those geeky formulas can stir up legal and ethical trouble.

BY KIM S. NASH

We're used to algorithms recommending books, movies, music and websites. Algorithms also trade stocks and predict crime, identify diabetics and monitor sleep apnea, find dates (and babysitters), calculate routes and assess your driving, and even build other algorithms. These math equations, which can reach thousands of pages of code and routinely crunch hundreds of variables, may someday run our lives. Companies increasingly use them to run the digital business and gain competitive advantage.

Unleashing an algorithm can lead to new customers and revenue, but it can also bring encounters with ethical and legal trouble. Already, consumer advocates and regulators are training their sights on the dark side of the algorithm revolution, such as creepy over-personalization and the potential for illegal price discrimination.

As CEOs look to chief digital officers and data scientists to conquer the next frontier, CIOs have sometimes been on the sidelines, whether by choice or default. But as business leaders, CIOs may now have to elbow into meetings where Ph.D.s, corporate lawyers and other colleagues are talking about the data-driven future. CIOs need to join those conversations to help steer company strategy, certainly, but also to contribute to decisions about what data to pour into an algorithm and what to keep out, and how to monitor what the algorithm does.

That includes devising a defensible policy for handling the information produced, says Frank Pasquale, a professor of law at the University of Maryland. "Algorithmic accountability" will become part of the IT leader's job, he says.

That realization can hurt. Athena Capital Research, a high-frequency stock trader, used a proprietary algorithm called Gravy to slip in big buy and sell orders milliseconds before the NASDAQ exchange closed for the day in order to push stock prices higher or lower, to Athena's advantage. The Securities and Exchange Commission viewed that as illegal manipulation and last year called out Athena's CTO for helping other managers plot the most effective use of Gravy during at least six months in 2009. Athena settled the case for $1 million.

No one says CIOs must delve into Ph.D.-level math. But a working knowledge of basic concepts behind algorithms can help avoid bad results and bad press. "Algorithms allow us to get rid of biases we thought were there in human decision-making," says Michael Luca, an assistant professor at Harvard Business School. "But pitfalls are equally important to think about."

"Algorithms allow us to get rid of biases we thought were there in human decision-making."

-MICHAEL LUCA, ASSISTANT PROFESSOR, HARVARD BUSINESS SCHOOL

Math magic

Algorithms can be used to make operations more efficient, answer "what if" questions and make new products and services possible. At United Parcel Service, the 1,000-page Orion algorithm does all of that. In 2003, UPS started building Orion (for On-Road Integrated Optimization and Navigation) to optimize delivery routes. You might have six errands to do on a given day. A UPS driver has about 120. The company wanted to save time and fuel by having drivers follow the most efficient routes possible while still making deliveries on time, says Jack Levis, director of process management. Levis oversees Orion and the team of 700 engineers, mathematicians and others who support it.

Cutting just one mile per driver per day saves $50 million per year, Levis says, and Orion has so far saved seven to eight miles per driver per day. UPS is on track to save $300 million to $400 million per year in gas and other costs by 2017.

The most important thing any manager can do when embarking on an algorithm project is to "work backwards," Levis says. That is, define carefully what business decisions the company struggles with, then identify what knowledge would help — what information you'd need to teach you the knowledge you lack. Then identify the raw data that — when combined and teased apart and interpreted — would provide that information.

UPS spent nine years working on Orion before putting it into production, adding and subtracting data, testing, then adding and subtracting again. For example, at first Orion used publicly available maps. But they weren't detailed enough. So UPS drew its own, showing features such as a customer's half-mile driveway or a back alley that shaves time getting to a receiving dock — data points that Orion needs in order to plan how to get a package delivered by 10:30 a.m.

But an algorithm created by data scientists in a laboratory can't anticipate every factor or account for every nuance. Suppose a business customer typically receives one package per day. If Orion knows the package isn't tied to a certain delivery time, the algorithm might suggest dropping it off in the morning one day but in the afternoon the next, depending on the day's tasks. That might be the most efficient approach for UPS, but customers wouldn't know what to expect if delivery times changed frequently.

People don't like that amount of uncertainty, and it might have cost UPS business. Companies often take deliveries in the morning, go about their business during the day and then call UPS back to request a late-afternoon pickup of an outgoing package. If UPS pushed deliveries to the end of the day for efficiency's sake, it might not get that later call, Levis says. "We started realizing the rules we told the algorithm weren't as good as they should have been," he says. "We've learned you need to balance optimality with consistency."

> **"We've learned you need to balance optimality with consistency."**
>
> -JACK LEVIS,
> DIRECTOR OF PROCESS MANAGEMENT. UPS

The Orion team is outside IT, but Levis says the IT group built the production version of Orion and CIO Dave Barnes understands what Orion can and can't do, which is critical when he helps UPS devise business strategy. UPS's My Choice service, which notifies customers of pending deliveries and lets them change delivery times or locations, wouldn't be workable without Orion, Levis says. Not only does My Choice reduce multiple delivery attempts, it also brings in new revenue: 7 million customers have signed up for the service and pay $5 per change or $40 per year for unlimited changes. Next, UPS wants to bring it to other countries.

To grow new business from algorithmic insights, companies must look for correlations that competitors haven't spotted.

Take H&R Block, for example. In December, executives at the provider of tax filing software and services talked in detail with financial analysts about the company's new algorithm, which tailors marketing email messages and in-software pop-up boxes to individual customers. The company rolled it out this tax season, after starting algorithmic tests to quantify and categorize the behavior of 8,700 tax filers in an effort to predict what customers will do.

CMO Kathy Collins discussed how, for example, H&R Block may know that, based on past behavior, you're typically a February filer who prefers to interact with the company via mobile device. If you haven't filed by Feb. 10, the algorithm will suggest that someone nudge you with an email reminder and a discount on help preparing your return. Other customers may receive an email offer the week they receive their W-2 forms.

Over time, H&R Block expects to improve its algorithm by analyzing not only the content of customer tax returns but also the very clicks a taxpayer makes while using its software, said Jason Houseworth, president of global digital and product management. "In our case," he said, "the personal data is as rich as it gets."

The personalization made possible through the algorithm, Houseworth said, "will make each user feel that the software was not only designed for them, but is always a step ahead."

Some customers may like that, but others won't, says Pasquale, who wrote *The Black Box Society: The Secret Algorithms That Control Money and Information*. "There's so much pressure to know more. That's the arms race I fear."

The idea of knowing more about people is a driving force at eHarmony. The dating service matches members by their self-identified characteristics, such as hobbies and sexual orientation. But eHarmony also extrapolates what it calls unstated "deep psychological traits," such as curiosity, by putting answers to questionnaires through various formulas. A neural network also produces a "satisfaction estimator"

Examples of algorithms in action

Companies use algorithms in all sorts of systems to make money or optimize operations, and sometimes both at once. Here's a sampling.

PUTTING THE SQUEEZE ON

Coca-Cola's Minute Maid division optimizes drink blends–especially orange juice–by understanding consumer preferences about factors such as mouth feel and pulp content. But operations variables are also figured in, including freight and labor costs. Coca-Cola wants to create not only juices that people will like, but also ones that work with the realities of its supply chain and production processes.

WHO'S WATCHING WHAT?

Netflix knows that several people might share one household account, so it wants to be able to correctly guess who is using its video service and make appropriate recommendations without having to ask customers to identify themselves or answer questions during every visit. Toward that end, the company has developed an algorithm that analyzes patterns of customer behavior and alters movie suggestions based on certain variables. One is time of day. In the afternoon, the viewer is more likely to be a child. Late at night, an adult. In the future, the system might make gender and age assumptions based on how hard the customer hits buttons on the device he's using.

SHOPPING AROUND

Retailers have turned to algorithms to help solve the omnichannel puzzle. American Eagle Outfitters uses an algorithm to figure out how best to fulfill online orders with products shipped from physical stores. One goal: Focus on slow-moving products at low-volume stores that are within two days' shipping distance of online customers.

MONETIZING SMALL TALK

MeetMe, an app vendor that matches members for mobile chats, has an "icebreaker algorithm" that uses quantitative data points, such as location, plus qualitative information, like self-declared interests in certain games, to suggest partners. This year, MeetMe plans to improve the relevance of its matches so people will spend more time chatting. Longer encounters mean more exposure to ads and more revenue for MeetMe.

FASTEN YOUR SEATBELTS

Auto insurers have long calculated premiums based on the number of miles someone drives, the customer's age, gender, marital status, vehicle type, driving record and home address. Now Allstate is fine-tuning that process with an algorithm that weights each risk factor differently. For example, the relative risk related to a 16-year-old driver could be triple that of someone age 30 to 50. But Allstate doesn't simply assume that all the miles a teenager drives are equally risky. Over the course of a year, for example, the miles driven at the start of a policy are probably riskier than the ones driven at the end because, of course, practice helps. So Allstate determines risk levels for both kinds of miles. However, there's a point at which learning slows down and each mile after that carries diminishing returns. The algorithm tries to find the point in the year at which that likely happens.

-K.S.N.

to rate potential pairings, and the system learns over time, as members report back about their satisfaction with matches eHarmony suggests.

The company doesn't have a CIO; COO Armen Avedissian handles that role. Decisions about whether to change the algorithm are made by a team that includes Avedissian, CTO Thod Nguyen, vice president of matching Steve Carter and corporate lawyers. "It's not just hardware and software but the tactics and strategy of data manipulation," Avedissian says.

The company looks at 29 dimensions of compatibility, such as "emotional energy" and "curiosity," each of which incorporates several variables collected through detailed questionnaires. More than

125TB of data is involved. The algorithm learns by assessing what a member does with each match that eHarmony suggests (contact right away? ignore?) as well as what feedback the members provide in questionnaires and open-ended responses. That data gets poured back into the equation and the cycle starts again, more informed, Avedissian says.

The more relevant the matches, the higher the rate at which members will communicate with each other. The more they engage, the more likely they are to buy annual subscriptions. All the algorithms at eHarmony are intended to convert registrants into subscribers.

The dating service tests ideas by running slightly different algorithms for different customers, then measuring the rate at which registered members convert to annual subscribers. Risk and compliance teams run their own algorithms to see how the company's other algorithms are using sensitive data.

One recent discovery: Whether someone smokes and drinks turns out to be more important in dating in Europe compared to the United States. Once eHarmony more heavily weighted the smoking and drinking variables in its matching algorithm in the U.K., "business just took off," Avedissian says. Meaning, suggested matches were more on-target, therefore satisfaction increased—and so did conversion rates.

However, not all outcomes are expected.

Problems are most likely to arise when algorithms make things happen automatically, without human intervention or oversight.

Unintended consequences

Uber is upending the taxi business with an app to connect passengers with rides and a proprietary algorithm that, in part, governs "surge pricing," which raises fares at times of heavy demand. Taxi associations from New York to Paris and back have protested Uber for cutting into their business, and government regulators have challenged the company on questions of fair pricing and safety. Even so, the darling of disruption has raked in an estimated $4.9 billion in investor funding.

In December, the cold, hard math collided with high emotion: Uber's algorithm automatically jacked up rates in Sydney, Australia, as people tried to get away from a downtown café where an armed man held 17 people hostage. Three people, including the gunman, died. Uber later apologized for raising fares, which reportedly were up to quadruple normal rates, and made refunds. "It's unfortunate that the perception is that Uber did something against the interests of the public," a local Uber manager said in a blog post. "We certainly did not intend to."

Problems are most likely to arise when algorithms make things happen automatically, without human intervention or oversight. Control is critical, says Alistair Croll, a consultant and author of —*Lean Analytics: Use Data to Build a Better Startup Faster*—. "If algorithms are how you run your business and you haven't figured out how to regulate your algorithms," he says, "then by definition you're losing control of your business."

Uber is working on a global policy to cap prices in times of disaster or emergency, a spokeswoman says.

Other unintended consequences involve the liability of knowing too much.

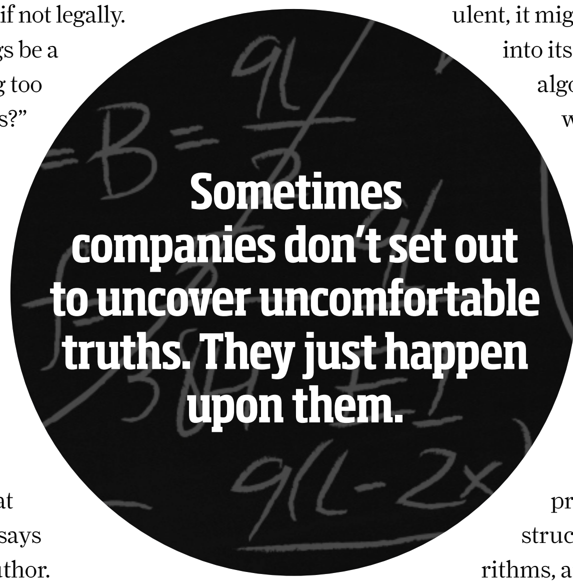
For example, say a hospital uses patient data to identify people who may be headed toward an illness, then calls them to schedule preventive care. If the math is imperfect, the hospital might overlook someone who later contracts an illness or dies. Or a whole group of people could get overlooked. "There's concern about who are the winners and losers and can the company stand by it later, when exposed," Pasquale says.

In another scenario, a company could open itself up to discrimination claims if it keeps too much data and insights about its employees, he says. Someone might be able to prove the company knew about, say, a health condition before letting him go.

Or if a car insurance company discovers there's a higher chance a customer will get into a crash after driving a certain number of miles, it may find itself in a "duty to warn" situation, Pasquale says. That's when a party is legally obligated to warn others of a potential hazard that they otherwise couldn't know about. It usually applies to manufacturers in product liability cases, or to mental health professionals in situations involving dangerous patients. And as the use of revelation-producing algorithms spreads, Pasquale says, people in other sectors could be subject to a similar standard—at least ethically, if not legally.

"At what point will things be a liability for you by knowing too much about your customers?" he asks.

Sometimes companies don't set out to uncover uncomfortable truths. They just happen upon them.



Insurance company executives, for example, should think carefully about results that could emerge from algorithms that help with policy decisions, says Croll, the consultant and author. That's true even when a formula looks at metadata — descriptions of customer data, not the data itself. For example, an algorithm could find that families of customers who had changed their first names were more likely to file claims for suicide, he speculates. Further analysis could conclude that it is likely those customers were transgender people who couldn't cope with their changes.

An algorithm that identified that pattern would have uncovered a financially valuable piece of information. But if it then suggested that an insurer turn down or charge higher premiums to applicants who had changed their first names, the company might appear to be guilty of discrimination if it did so, Croll says.

The CIO's best role

The best way a CIO can support data science is to choose technologies and processes that keep data clean, current and available, says Chris Pouliot, vice president of data science at Lyft, a competitor of Uber. Before joining Lyft in 2013, Pouliot was director of algorithms and analytics at Netflix for five years and a statistician at Google.

CIOs should also create systems to monitor changes in how data is handled or defined that could throw off the algorithm, he says. Another key: CIOs should understand how best to use algorithms, even if they can't build algorithms of their own.

For example, if a payment service needs to figure out whether pending transactions could be fraudulent, it might hard-code an algorithm into its payment software. Or the algorithm could be run offline, with the results of the calculations applied after the transaction, potentially preventing future transactions. The CIO has to understand enough about what the service is and how the algorithm works to make such decisions, Pouliot says.

CIOs should, of course, provide the technology infrastructure to run corporate algorithms, and the data they require, says Mark Katz, CIO of the American Society of Composers, Authors and Publishers, which licenses, tracks and distributes royalties to songwriters, composers and music publishers.

Katz meets regularly with ASCAP's legal department to make sure the results of the algorithms comply with the organization's charter and pertinent regulations.

"We're all information brokers at the end of the day," he says.

CIOs can expect increasing scrutiny of analytics programs. The Federal Trade Commission, in particular, is watching the use of algorithms by banks, retailers and other companies that may inadvertently discriminate against poor people. An algorithm to advise a bank about home loans, for example, might unfairly predict that an applicant will default because certain characteristics about that person place him in a group of consumers where defaults are high.

Or online shoppers might be shown different prices based on criteria such as the devices they use to access an e-commerce site, as has happened with Home Depot, Orbitz and Travelocity. While companies may think of it as personalization, customers may see it as an unfair practice, Luca says.

The Consumer Federation of America recently expressed concern that, in the auto insurance industry, pricing optimization algorithms could violate state insurance regulations that require premiums to be based solely on risk factors, not profit considerations.

Consumers, regulators and judges might start asking exactly what's in your algorithm, and that's why algorithms need to be defensible. In a paper published in 2013 in the Boston College Law Review, researchers Kate Crawford and Jason Schultz proposed a system of due process that would give consumers affected by data analytics the legal right to review and contest what algorithms decide.

The Obama administration recently called on civil rights and consumer protection agencies to expand their technical expertise so that they'll be able to identify "digital redlining" and go after it. In January, President Obama asked Congress to pass the Consumer Privacy Bill of Rights, which would give people more control over what companies can do with their personal data. The president proposed the same idea in 2012, but it hasn't moved forward.

Meanwhile, unrest among some consumers grows. "Customers don't like to think they are locked in some type of strategic game with stores," Pasquale says. CIOs should be wary when an algorithm suddenly produces outliers or patterns that deviate from the norm, he warns. Results that seem to disadvantage one group of people, he says, are also cause for concern. Even if regulators don't swoop in to audit the algorithms, customers may start to feel uneasy.

As Harvard's Luca puts it, "Almost every type of algorithm someone puts in place will have an ethical dimension to it. CIOs need to have those uncomfortable conversations." ■

THINKSTOCK

Big data makes a difference at Penn Medicine

Here's how one healthcare organization is making use of the massive amount of information – measurable in petabytes – it now has at its disposal to save lives.

BY KEN TERRY

The team of clinicians and medical informatics experts led by Mike Draugelis, chief data scientist at Penn Medicine in Philadelphia, is busy these days. Using insights from a massively parallel computer cluster that stores a huge volume of data, the team is building prototypes of new care pathways, testing them out with patients and feeding the results back into algorithms so that the computer can learn from its mistakes.

This big data approach to improving the quality of care has already produced one significant success: The Penn team has improved the ability of clinicians to predict which patients are at risk of developing sepsis, a highly dangerous condition, and it can now identify these patients 24 hours earlier

"We're creating machine learning predictive models based on thousands of variables." - DRAUGELIS, UNIVERSITY OF PENNSYLVANIA

than it could before the algorithm was introduced.

Draugelis and his colleagues work in the hospital of the University of Pennsylvania. On the academic research side, the university's medical school has launched an Institute of Biomedical Informatics (IBI) to do basic research using big data techniques. Announced in 2013, IBI is now coalescing a few months after naming Jason Moore, Ph.D., who founded a similar institute at Dartmouth, as its director. IBI will focus its efforts on precision medicine, a hot field that is starting to take off as genomic sequencing costs drop.

The effort to link genomic differences with "phenotypes" – the variations in patients' characteristics and diseases – has been underway for five years, says C. William Hanson III, M.D., chief medical information officer and vice president of Penn Medicine and a member of IBI. But he sees this kind of research quickly accelerating.

Steven Steinhubl, M.D., director of digital medicine at the Scripps Translational Science Institute in La Jolla, Calif., agrees. "We're still on the rising part of the curve of what we're going to learn from big data," he says. "It's rapidly growing, but it will accelerate even more as large medical centers like UPenn take advantage of the data they're already collecting and add genomics on top of that."

Changing clinical pathways

Draugelis' team at Penn Medicine is using algorithms to tweak the guidelines that doctors and nurses follow in diagnosing and treating particular conditions. When a protocol changes, he explains, the clinical team must develop a new care pathway that specifies each step in the workflow of clinicians. It is very intensive work, and so is coding the changes that must be made in the algorithm to adjust to the feedback from the frontline of patient care.

"We're working in two week sprints, where the clinicians adjust their pathways, and we adjust the algorithms to their needs," Draugelis notes.

The team builds a prototype of a new pathway for a particular condition about once every six months. Currently, it is focusing on finding a better way to predict which patients have congestive heart failure and which are likely to be readmitted after discharge from the hospital. In addition, the team is working on acute conditions such as maternal deterioration after delivery and severe sepsis.

"We're creating machine learning predictive models based on thousands of variables," Draugelis says. "We look at them in real time, but we train them up over millions of patient records."

In the case of sepsis, the team started with an expert model known as SIRS (systematic inflammatory response syndrome), which uses specific thresholds of temperature, heart rate, respiratory rate, and white blood count as key indicators of sepsis risk. After loading in all of the available data on a patient, including electronic health record (EHR) data, the computer uses the algorithm to determine how closely a patient's characteristics match those of patients who developed sepsis in the past. When a patient matches that profile, the clinician caring for the patient receives an alert, acts on it or doesn't, and feeds his or her reaction back to the algorithm to improve it.

Penn Medicine's bedside monitors continuously track vital signs and document them in the EHR. This automated documentation of vital signs didn't occur five years ago, Hanson notes. It is still not widespread outside of intensive care units, says Steinhubl, but when it does become routine, he adds, it will provide a major boost to the kind of work that Draugelis' team does.

Dean Sittig, Ph.D., a professor at the University of Texas Health School of Biomedical Informatics in Houston, likes the idea of continuous monitoring and feeding data into computer algorithms. In contrast to the average floor nurse, who can only watch a patient 20 percent of the time if she has five patients, "The computer can be looking at every minute, and the idea of continuous monitoring and surveillance is very powerful," he says. "If you can teach the computer what the nurse would be looking for, the computer can be much more vigilant [than the nurse]."

To make the decision support alerts useful, however, the staff has to be ready to spring into action,

> The sheer volume of genomic data is staggering.

especially with a condition like sepsis, Sittig says. In addition, the alerts that the algorithm triggers must be fairly accurate. "As a rule of thumb, if the computer is right more than half the time – especially with something serious like sepsis – clinicians will pay attention to it. But if it's only right 10 percent of the time, it starts to be a bother."

Precision medicine

Two important developments have come together to make possible the kind of precision medicine research that Penn Medicine's IBI is doing. First, EHRs have become widespread in the past few years: most hospitals and more than 80 percent of physicians now have these systems. Second, the cost of genomic sequencing has dropped to around $1,000 for a complete genome. The cost of partial genome or exome sequencing is less than that. As a result of these trends, the idea of correlating genotypic and phenotypic variants to discover individual responses to diseases and drugs is now feasible.

To perform this kind of research, Penn Medicine has created a specialized "bio-bank" that, so far, has stored about 20,000 genomic samples with patients' permission, says Brian Wells, associate vice president of health technology and academic computing for the healthcare system. A separate center for personalized diagnostics has sequenced tumor genomes for more than 5,000 patients, he notes.

The sheer volume of genomic data is staggering. For example, Penn Medicine has two petabytes of disk space in its high performing computer cluster, and it plans to expand that, says Wells.

"One researcher told us that in the next few years, he might go from five to 30 petabytes of space related to neuroscience sequencing. So we're prepared to add to that as we need to," he notes.

Challenges for CMIOs and CIOs

The biggest challenges that Hanson faces as Penn Medicine grapples with its big data projects, he says, is the lack of interoperability among EHRs and the need for good, clean, structured data. Currently, Penn has different EHRs in its hospital, ER, ICU and ambulatory practices, but it is moving to a single system. Structured clinical data is harder to deliver, however, because "clinicians tend to document in an unstructured way," he says.

Penn intends to use natural language processing (NLP) to mine unstructured data in EHRs and convert it into structured information, Wells notes. "That's for retrospective analysis rather than clinical decision support, because you can't rely on it 100 percent of the time," he adds.

Current big data methods are adequate for processing the huge flood of genomic data, but bioinformaticians who know how to work with this data are in short supply, Steinhubl says. He predicts that a bottleneck will develop in data processing and storage when healthcare providers begin to review the physiologic data that is expected to flow in from mobile devices and wearable sensors.

Nevertheless, Steinhubl is very excited about the promise of big data in fields like precision medicine and clinical quality improvement. "Eventually, it's going to completely change medicine and the way we treat common chronic conditions," he says.

For example, he notes, most cases of hypertension are defined as a single disease. "So we put them all in one basket and treat them the same way. With these tools, we'll be able to refine their phenotype and their genotype and better treat these individuals. Right now, it's mostly trial and error."

Hanson leavens the great expectations of big data with a few sober reflections. First, he notes, it will be some time before most providers are ready to pull in remote monitoring data, because it has to be prescreened to be usable in patient care. Second, while precision medicine is a great idea, most people haven't yet been sequenced, and "we don't have a consistent way of interpreting their genotypic data and making it actionable."

While oncologists are increasingly using information about the genetic differences among individual cancer patients, it will be a while, Hanson says, before this approach filters down to primary care physicians. However, precision medicine research is moving fast at Penn Medicine and other leading academic medical centers. "We're on the verge of an explosive development," he says. ■

Don't look for unicorns, build a data science team

Bob Rogers, chief data scientist with Intel's Big Data Solutions, says that rather than seeking out rare individuals who excel in all the areas that encompass data science, CIOs should build data science teams with complementary talents.

BY THOR OLAVSRUD

THINKSTOCK

When it comes to big data, one thing seemingly everyone can agree on is that organizations face a shortfall of data science talent. After all, the ideal data scientists aren't just wunderkinds in advanced mathematics and statistics, they're creative, non-linear thinkers with excellent communication skills. In popular parlance they're unicorns — magical creatures that don't exist.

Research firm McKinsey has predicted that by 2018, the U.S. alone may face a 50 percent to 60 percent gap between supply and demand of deep analytic talent.

Bob Rogers, chief data scientist, Big Data Solutions, Intel, might just fit the unicorn bill. Rogers, who holds a PhD in Physics from Harvard University, got his start studying supermassive black holes. He co-wrote a book on time series forecasting using artificial neural networks, which led him to co-found a quantitative futures hedge fund that leveraged large amounts of historical and streaming tick-by-tick data from markets. He's also helped a medical technology firm revolutionize glaucoma diagnostics and co-founded another business to help the healthcare industry extract data from electronic health records.

For the past year, he's been the chief data scientist at Intel's Big Data Solutions, which started as a project to better leverage the data inside Intel, but has grown to encompass helping Intel clients better understand analytics and data problems. For instance, he's been working closely with the Knight Cancer Institute at Oregon Health & Science University (OSHU) to help develop the Collaborative Cancer Cloud, which aims to make it possible to sequence an individual's cancer, analyze it and formulate a precision treatment plan within 24 hours.

You'll never find that data science unicorn

In the process, Rogers has helped to define what Intel looks for in its data scientists, and it's not unicorns who have a background in math, statistics, physical science or hard science; the ability to write production-level code; and the ability to talk to business people in their own language.

"You don't have to be a unicorn," he says. "We're looking for people who have one of the major skill sets and some comfort level with the others — the ability to be creative, handle ambiguity and communicate well. One of the key outputs of that kind of thinking is the ability to characterize what's important to others."

Think in terms of data science teams with diverse capabilities that can complement each other, rather than seeking to hire individuals that can do it all, Rogers says.

"It's true that having advanced knowledge of mathematics and programming is fantastic background for a data scientist," he says. "But, in any company, you won't find just one data scientist doing it all — just like Michael Jordan couldn't have scored so many points without Scotty Pippen at his side, data scientists all bring their own skills to the table that together build an ideal team."

"In fact, we're looking for all kinds of skills and backgrounds as we look to build our team at Intel — from programmers to those with creativity, curiosity and those with great communication skills," he adds. "It's rare to find a "data unicorn" that can do it all, and we're not spending our time recruiting for such a talent. We build out teams to reflect a variety of backgrounds and experience, which brings greater insight to our data analytics work."

Getting hands dirty with data

Because there's no one-size-fits-all path to becoming a data scientist, it can be difficult to identify good candidates. Rogers' advice is to look for individuals that can show their mettle by getting their hands on a data set — perhaps from Kaggle, DataKind or the government — building up a data analytics environment and telling a story with that data. And individuals interested in pursuing data science should take it upon themselves to seek out data sets to work with.

"Get your hands on data and do something with it," he says. "There are big data sets out there, some of them sufficiently ugly enough to give you some real experience working with data. Take a big data set, put it into an environment where you can really do something with it and answer a simple question. You don't need to do anything too technically crazy. When you work with data that's messy, you start to see where data sets go wrong. That is the moment you start to speak the data scientists' language." ■

THINKSTOCK

8 ways to make the most of customer data

Big data, business analytics and marketing experts discuss how organizations can best put to use all that consumer data they've been collecting.

BY JENNIFER LONOFF SCHIFF

Everyone talks about the importance of big data. But many organizations, although they collect and store customer data, do not put that data to good use – or they don't know how to.

So how can businesses leverage all that big data they've collected? Following are eight ways to make the most of customer and consumer data.

1 Use customer data to create a more personalized, pleasurable shopping experience. "According to an Infosys study, 70 percent of Americans are willing to spend an average of 13 percent more with companies they feel provide superior service," says Scot DeLancey, director, Department and Specialty Retail Solution Management, NCR.

By leveraging customer and consumer data, retailers can "make personalized recommendations, inform shoppers of special offers and promotions that are most relevant to them and fully maximize cross-sell and up-sell opportunities to realize increased revenues."

Video game retailer GameStop, for example, has found that "data from loyalty programs offer the best insight on each customer's interests, past purchases and engagement preferences," says Rob Lloyd, its CFO. "This data can be used to create a customized shopping experience down to the individual and even offer product recommendations."

"Companies with brick and mortar stores can use customer data to improve their customer's in-store experience," adds Greg Petro, CEO, First Insight.

"For example, using customer location data and heat maps, companies can better understand and improve traffic flows and recognize where there may be opportunity to optimize a store's layout or adjust merchandizing displays," says Chris Wareham, senior director of Product Management, Adobe Analytics. "This technique can even be used at other types of locations, such as stadiums, airports, museums and more."

"Also, customer data can identify differences in preference by region, enabling stores in each region to feature the right products in store windows and in key locations within the store," says Petro.

2 Use customer data to customize promotions and special offers. "Successfully analyzing and acting on customer, market and competitive data can help companies provide their customers with customized offers, appropriate marketing and ad campaigns, the right deals, or even when to back off," says Chris Selland, vice president, business development, big data at HP. "The most loyal customers are also the most profitable, and by analyzing and acting effectively, organizations can ensure long-term customer loyalty and significantly enhance their profitability."

"Real-time analysis of in-store purchases and website clicks can show a retailer what promotions are effective as loss leaders right now, what high-margin items they drag along with them, and the differences in purchasing habits and profitability for these items between online and in-store purchases," says Jake Freivald, vice president of product marketing, Information Builders, makers of business intelligence, data integrity and integration software. "That real-time insight based on customer data can ensure the right promotions go to the right channels, ensuring a win-win: happy customers and high profitability."

3 Use customer data to get helpful product feedback - and improve your products or services. "By applying analytics to customer data, companies can identify the specific products that each customer is likely to want, and the price they are willing to pay," says Petro.

"One online travel site with more than 2 million members wanted to reward frequent travelers while making it easy and intuitive to cash in on rewards," says Susan Ganeshan, CMO, Clarabridge, a customer intelligence platform provider. "Using survey feedback data and online reviews, the team was able to improve both the digital experience and the rewards program, propelling them to be named the No. 1 travel site."

You can also "use customer data to identify the most relevant users to ask for feedback on new features as they're the ones you really want using those features," says Aaron Forman, manager of Communications, Intercom, which helps businesses connect with customers.

4 Use customer data to improve your marketing. "Big data enables marketers to understand the cross-channel behavior of prospects that become customers, meaning you can see the prospect-to-customer journey and the campaigns that influenced them the most," says Azita Martin, CMO, Datameer, a big data analytics application for Hadoop.

"For example, by correlating data between digital and non-digital advertisements (e.g., purchasing history, profile information, behavior of customers on social media sites) companies can find patterns of behavior for high-value customers," she says. "With these insights, they can adjust their marketing strategy to target people of certain profiles with specific advertisements" and make their marketing dollars go farther. For example, by analyzing its customer data, "one company was able to decrease its customer acquisition costs by $3.5 million per year and at the same time increase conversion by 20 percent."

"We use very discrete information about how each customer is engaging with our brand to drive one-to-one marketing and sales efforts," says Tyler Lessard, CMO, Vidyard, which specializes in video marketing. "For example, we're not only tracking what videos they click on throughout the day, but how long they watch a piece of content, helping us better infer their interests and intent," he says. "Getting visibility into their actual engagement in our marketing assets gives us incredible context on each and every customer and helps us create very targeted nurture programs and drive sales efficiency."

5 **Use customer data to create new products or services.** For example, "industries that have access to information from device sensors can turn this information into new service offerings," says Jean-Michel Franco, director of product marketing, Talend, a provider of open-source integration software. "For instance, a sports equipment manufacturer could create a complimentary service providing guidance to athletes on how they optimize their workout regime, or a healthcare provider could use information from a fitness watch to provide personalized healthcare services to their clients," he explains.

6 **Use customer data to provide better customer service.** "Use historical customer data from multiple functions, like purchase or support history, [to] provide more personalized customer support," suggests John Fanelli, vice president of marketing at DataTorrent, which helps large organizations make sense of mass amounts of data in real time. "For example, when a customer calls, [agents could greet him] with a customized response along the lines of, 'Hello, Joe. Thank you for being a customer since 1995. I see that your last call was regarding product X. Is this the same product you're calling about?'"

7 **Use customer data to improve organizational effectiveness and reduce risk and fraud.** "Many companies see customer insights as opportunities for cost savings," notes Jeffrey Hunter, vice president, North America, Insights & Data, Capgemini, an international management consulting firm. "As a matter of fact, 65 percent of respondents from a recent Capgemini report [saw] big data as a key enabler for organizational effectiveness/competitiveness," he says.

For example, "by analyzing customer data, such as customer churn rates and behaviors, companies are empowered with the insights needed to maximize efficiency, reduce risk [and] better [detect and] deter fraud," which can result in substantial savings.

8 **Use customer data to create shareable content.** Use the "data you've gathered – from surveys, for example – and turn it into a well-designed, interesting infographic," suggests Termeh Mazhari, a PR, marketing & SEO consultant. Then share it on social media to generate buzz about your business. ■