August 2014

Knowledge@Wharton – SAP

Improving Government Performance, Anticipating Citizens' Needs

Sponsored by

Knowledge@Wharton – SAP

Improving Government Performance, Anticipating Citizens' Needs

Predictive analytics is becoming a vital tool for a wide array of government activities, from identifying tax evaders and possible terrorists to figuring the best way to deploy resources in a flu epidemic.

Every political campaign confronts an age-old problem: how to identify the voters who are still on the fence. There is no point spending time and money on diehards, be they supporters or opponents. Instead, the campaign wants to find the undecideds who can be turned into supporters with just the right message.

So, as the 2012 U.S. presidential re-election campaign got underway, President Obama's team turned to the rapidly advancing field of predictive analytics for what was undoubtedly the most sophisticated effort any campaign had made up to that point to identify voters who could go either way at the polls. The high-tech solution looked past voting patterns in precincts and zip codes — often the sharpest resolution campaigns could get in the past — to identify individuals worthy of the campaign's attention, even if they were isolated in Republican enclaves.

"It's all about running a more efficient campaign — which TV spot you buy, what door you knock on or which person you send email to," explains Peter Tanner, one of the campaign's top analysts.

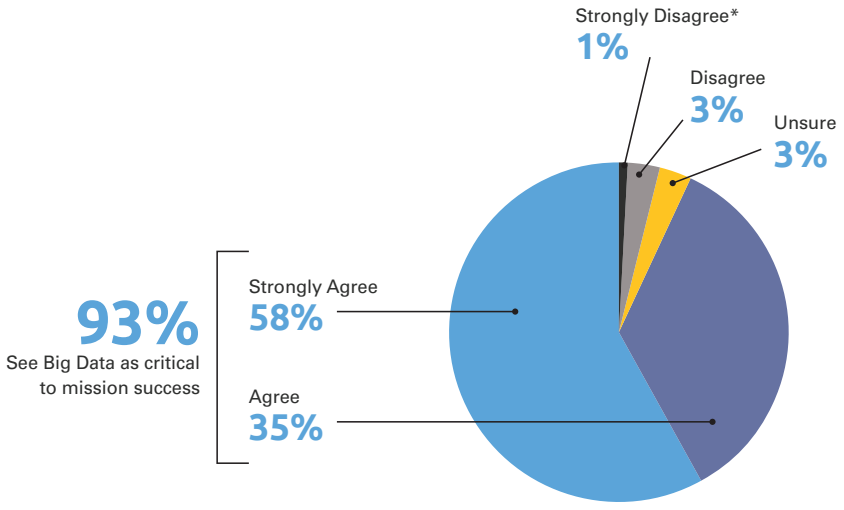
Capabilities that were impressive during that 2012 campaign have become even better since then, says Dante Ricci, senior director for public services at SAP.

"There are things that were impossible to do before that can be done today," he adds, describing how massive amounts of historical data can now be effectively combined with real-time data for timely insights when speed is of the essence.

"It's all about running a more efficient campaign."

— Peter Tanner, Obama Campaign Analyst

***Big Data is critical to mission success, say survey respondents**



Source: SAP Performance Benchmarking (https://valuemanagement.sap.com)

Predictive analytics is becoming a vital tool for a wide array of government activities, from identifying tax evaders and possible terrorists to figuring out the best way to deploy resources in a flu epidemic, or to counter challenges like the pine bark beetle or traffic congestion. It can even help with mundane matters like setting up the most efficient preventative maintenance schedule for road equipment or reducing theft of city-owned garbage cans.

"More and more local governments, as well as the federal government, are putting data online that has been scrubbed," says Shawndra Hill, a Wharton professor of operations and information management, who studies data mining and other computer analysis. "What that opens up is a way for a lot of people to think about how to use the data that is generated by the government.... Getting people to think about using data from different angles could result in huge payoffs."

Some of the newest efforts scan the billions of entries on social networking sites like Facebook, Twitter and YouTube to find out what people are saying about a given subject moment by moment. These capabilities will only get stronger over the next few years as computers get better at understanding full sentences instead of just spotting individual words; computing power becomes cheaper via the cloud; instantly accessible computer memory grows bigger, and predictive analytics functions are embedded in other software so that non-experts can produce reliable forecasts with little or no set-up time.

Intelligence agencies have led the federal government in using predictive analytics techniques like language analysis, but other governmental bodies will gradually adopt these capabilities as they become better and cheaper, notes Lyle H. Ungar, a professor of computer and information science at the University of Pennsylvania and operations and information management at Wharton. "Government agencies are only starting to explore using crowd-sourced feedback from citizens. What problems should we be addressing? From 'Where are the potholes?' to 'What could we do to save energy?'"

Agencies can begin to answer these questions in part because "computing has become cheaper, literally, overnight," Hill adds.

Crunching Huge Samples

For the 2012 Obama reelection campaign, the analytics team assembled about 50 people — up from a half dozen in the 2008 campaign, which had demonstrated predictive analytics' promise, Tanner recalls. In 2012, the process was used for everything from predicting sales of campaign merchandize, to organizing solicitations from major donors, to deciding which voters to target for e-mail messages. Late in the campaign, the team refined its get-out-the-vote effort by filtering out individuals who had voted early.

To do all this, the Obama campaign combined a range of databases including voter-registration files, details on donors past and present, Census data, lists of licensed drivers and databases from private vendors that had information like education levels by neighborhood. Much of the data from government sources like the Census Bureau was free and readily available on the Internet.

"Pushing the button to solve the model is actually trivial. It's really the cleaning of the data that's hard."

— Shawndra Hill, Wharton

"In terms of our database, on each person we would have close to 1,000 factors that you could go in to consider," Tanner said, describing variables from age to race, to income and education, and even things like magazine subscriptions. The high granularity allowed the campaign to do things like identify a likely Obama supporter who, after the 2008 campaign, had moved to a Republican enclave in another state.

One of the most valuable efforts, Tanner recalls, involved a "persuasion model" that started with a selection of about 20,000 voters. Half were contacted and given a pro-Obama economic message, the rest were not. A week later both groups were questioned and the program identified those who had received the message and decided to vote for Obama, as well as virtually identical counterparts who had not been exposed to the message and remained undecided or opposed.

"Since the sample size is large enough, you can actually [identify] the types of people who would support Obama if and only if they had gotten that message," Tanner explains. The campaign used the system to craft the most effective messages, then to direct them to the most persuadable voters in its vast database.

Fundamentally, predictive analytics involves looking for correlations between variables. This kind of analysis has been used for decades and experts have long been able to make sound forecasts given the right data, says Hill. But the field has advanced dramatically in just the past two or three years, she adds, because of growing computing power and the wider availability of big, accurate databases.

In the past, analyses like political polling were based on relatively small samples of a larger universe. The typical process would look at no more than a few thousand individuals, seeking correlations among a handful of variables chosen by the analyst, such as subjects' ages, genders, incomes, ethnicity, addresses and political leanings. The analyst would then look for patterns — women in a certain age and income range who tended to vote Democratic or Republican, who support more spending on schools or oppose it, for example.

Predictive analytics, in contrast, can look at the whole universe of data — hundreds of millions of individuals in a search for tax fraud, for instance. Rather than relying on variables chosen by the researcher, the program looks at all the variables available, even if they run into the thousands, to determine which really matter. The process can detect subtle correlations and patterns a researcher might never have imagined, or which would not be statistically valid in a small sample.

Predictive analytics, says Hill, "is mostly about attacking age-old problems with new and different, and maybe more interesting and complex, data."

A search for potential drug traffickers, for instance, could involve examining widening circles of relationships, notes Charles Gadalla, SAP's senior director of advanced analytics. A law enforcement agency could start by identifying men similar to the notorious Colombian drug lord Pablo Escobar. But then the search could widen to identify people who are similar to those who were part of Escobar's inner circle, since community ties are

"Government entities just getting started with predictive analytics might do well to start with a pilot project involving a fairly simple question. Layers of new data can be added later to deal with more complex issues."

— Dante Ricci, SAP

common among criminals. "If you look at it from a community perspective, maybe you look at his mother-in-law or his niece ... someone you would never have thought of," Gadalla adds. Having very large amounts of data available makes it possible to detect connections that would otherwise be lost in the fog.

Gadalla says that governments have used predictive analytics to help wrestle with a variety of problems. "Tax evasion has been a very popular one, both in the U.S. and outside the U.S.," he explains. Doing this requires very large data sets with thousands of fields involving, in addition to tax-return data, demographic data and information on things like where people work, how much they earn, what they buy and where. The software determines which fields, or variables, are found in patterns of known tax cheats, and then it searches the universe for people who fit the pattern.

Such systems were able to help one Latin American country increase its tax collections 11-fold, Gadalla notes.

Many Potential Applications

Government, like business, can use predictive analytics to make predictions in a range of areas, from identifying where crime is more likely to occur to figuring which applicants are more likely to default on student loans. But while a business typically uses predictive analytics to improve the bottom line by growing sales or cutting costs, governments use other measures of success.

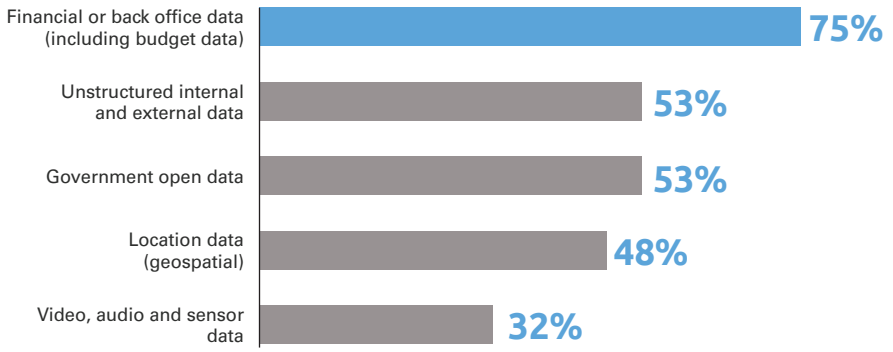
They might, for instance, want to identify people who could benefit from certain services, to reduce crime in given neighborhoods or simply to ease traffic congestion. A government agency using predictive analytics must therefore be especially diligent in framing the questions it asks and in understanding not only the answer, but also why that answer was what it was, Hill says.

Government users of predictive analytics must also be especially attentive to privacy issues, she adds: "While the public is comfortable about giving up their data to firms, they are not comfortable giving it up to government." Governments also must respect the line that separates legitimate prediction from inappropriate or illegal profiling. The public may want tax cheats to be caught, but not if the IRS takes to targeting certain groups for more audits. "Governments have to take better care [than business] to understand who the models are discriminating, and why," she says.

"There are frightening privacy implications to the fact that one's phone or Facebook account contain unprecedented amounts of personal information," Ungar adds. "The issues are fundamentally legal and moral,

> *"... Companies and governments will get increasingly good pictures of our intents, plans and wishes — the tasks we are working on, life events we experience ...."*
>
> — **Lyle H. Ungar, Wharton**

***The most sought-after insights involve financial data, say survey respondents**



Source: SAP Performance Benchmarking (https://valuemanagement.sap.com)

not technical. Government agencies need clear polices on what data they will or will not collect, and transparency as to what those policies are. The law hasn't caught up with the fact that although searching my house requires a warrant, searching my phone, which contains more private information than my house, may not."

In some cases, privacy and profiling issues can be avoided by "anonomizing" the database or replacing the individual's identity with something as simple as a number, says Ricci. The database remains vast and highly detailed — meaning it is still useful for deep insight — even if the end user does not know individual names.

Depending on the user's needs, systems can typically be directed to provide answers to a given certainty level, Gadalla notes. That can be a high-stakes decision if it involves something like terrorism. A low sensitivity could cause law enforcement to waste time tracking too many people and to invite a backlash over profiling. But a high setting could cause the system to miss people who should be watched. "What happens if they don't catch the guy and there's a 911 situation?"

Whether federal, state or local, a government entity thinking about using predictive analytics should start with a few basic steps, says Ricci. The first is to identify the kinds of questions that need to be answered, whether they involve fighting crime, delivering services or planning projects like new roads or schools. Government entities just getting started with predictive analytics might do well to begin with a pilot project involving a fairly simple question, he adds. Layers of new data can be added later to deal with more complex issues.

The goals should be realistic — some obstacles can be very difficult to surmount, Ricci says. The military, for example, could use predictive analytics for a real-time gauge of readiness by tracking units and keeping up to date on the location of key officers and supplies. But for security reasons, critical data might be kept in separate silos.

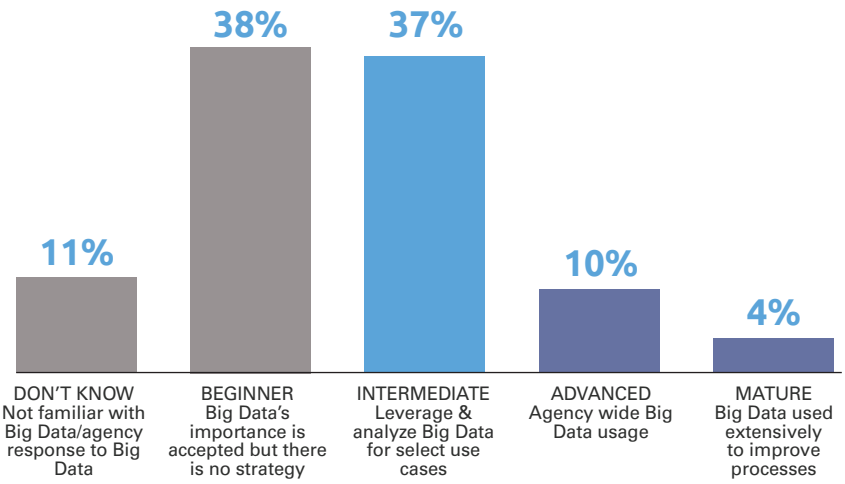
Preparation for predictive analytics also involves a look at the existing databases. What information do the agency and others already collect? What additional data could be collected in-house, and what is available outside the government for free or for purchase? Is there enough data to answer the questions the agency wants to ask?

Often, says Hill, one of the biggest hurdles is to make various databases compatible, and to scrub the bad entries. Many governments have a crazy quilt of legacy computer systems developed at different times to serve different goals. "Pushing

"You don't have to be a statistician anymore, you just have to say 'find fraud.' "You push a big red button."

— Charles Gadalla, SAP

***Just 14% of respondents report "Advanced" or "Mature" progress in their agencies' Big Data efforts**



Source: SAP Performance Benchmarking (https://valuemanagement.sap.com)

the button to solve the model is actually trivial," she notes. "It's really the cleaning of the data that's hard." Though vendors offer software that can help clean up databases, it is valuable for all the databases to use standardized definitions for each variable they contain.

Ideally, the data will be put into the form of a "flat file," Gadalla says. An example would be a table with each person listed on a separate row, followed by column after column of variables, factors or characteristics. That makes the data easy to scan.

While vendors can supply software aimed at doing specific tasks, buyers should look for packages that are flexible and scalable enough to deal with more complex issues in the future, Ricci notes. The newest systems are capable of upgrades that appear seamless to the user, saving the trouble and expense of installing an entirely new software release.

Flexibility is especially important because the field of predictive analytics is advancing so fast that it is hard to say what will be possible in two or three years. While today's systems, for example, are good at scanning a social networking stream for a key word such as "bomb," systems will get better and better at seeing the difference between the phrases "the movie was a bomb" and "I have finished building the bomb."

Over time, more and more data will be stored in memory that is instantly accessible — the equivalent of using more RAM to speed up a laptop, says Hill. And costs will come down as more data and software move to the cloud, which "makes computing extremely fast and extremely cheap." Many kinds of analysis can now be broken into pieces that are crunched simultaneously by different systems residing in the cloud, a far cheaper approach than using a big mainframe to start at one end of the data and work toward the other.

Still, Hill notes, there will continue to be a need for "enterprise systems" that reside in the user's machines, especially when the analysis involves a lot of proprietary data that evolves with the organization's day-to-day operations. In this area, too, costs are coming down, making enterprise systems useful for smaller organizations.

As capabilities grow, predictive analytics will advance further into "time-critical" analysis, Gadalla says. A system, for instance, might use existing databases to identify a potential terrorist, and then flash an alert to law enforcement when the GPS data on that person's phone shows he or she is approaching an oil refinery. Real-time analysis could involve things like the fastest ways to evacuate people from disasters, how best to tackle epidemics or even how to route traffic after a big sporting event.

Predictive analytics, Ricci adds, will gradually be integrated into mobile devices like smartphones, giving government workers in the field quick analytics for real-time problems, and allowing them to quickly feed data to the computers.

"Increasing amounts of our life are recorded electronically," Ungar says. "We send photos and buy items online. Skype 'sees' who we are talking to. Our phones 'hear' our conversations and know where we are having them and who we are talking to. As this information becomes collected and mined, companies and governments will get increasingly good pictures of our

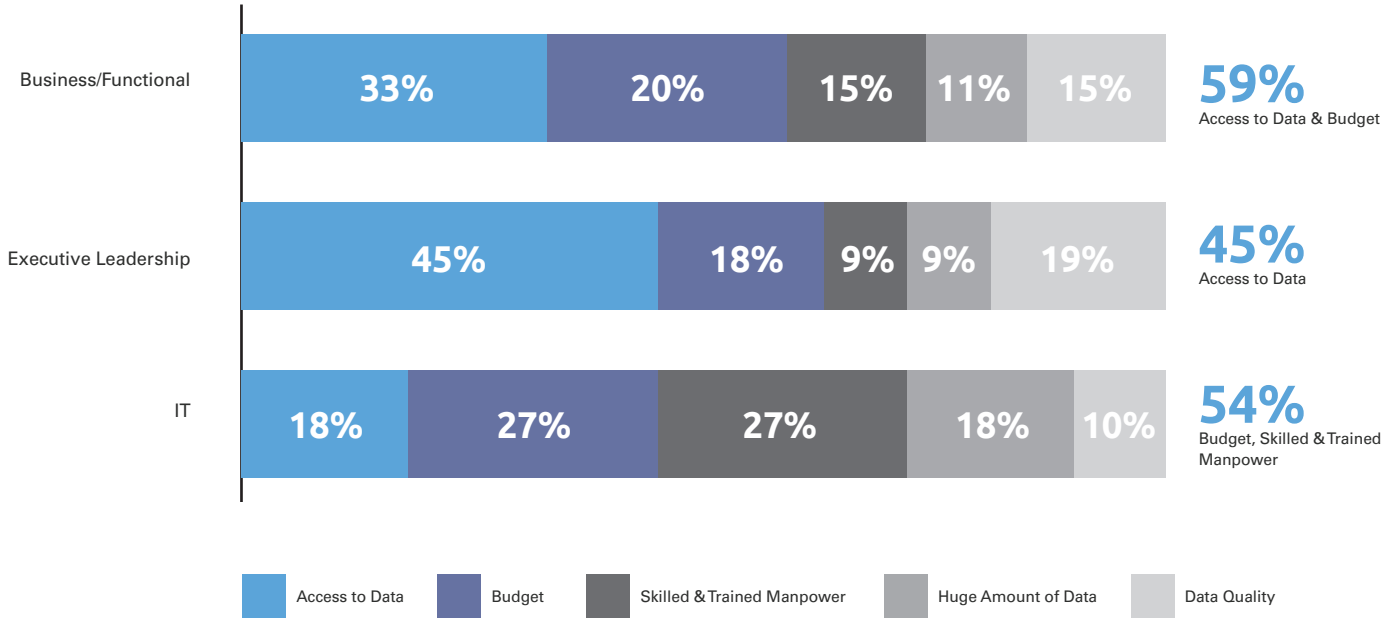
intents, plans and wishes — the tasks we are working on, life events we experience, perhaps our unstated desires."

For the public, the benefits can be great: more efficient use of tax dollars, better protection from the bad guys; smarter choices on issues like which roads or airports to improve; and fairer approaches to matters like who to audit and who to trust.

At its heart, government use of predictive analytics is all about being better informed before spending taxpayer dollars and delivering services. And the ease of use found in the newest systems allows more and more governmental agencies to benefit from those opportunities.

"You don't have to be a statistician any more, you just have to say 'find fraud,' " Gadalla says. "You push a big red button."

***Stakeholders' perception of challenges in Big Data vary, say survey respondents**



*The information for accompanying graphs in this white paper originated with a survey of more than 80 business and information technology executives from federal agencies, who attended the SAP Federal Forum in Washington, D.C. in April 2014.

This article was produced by Knowledge@Wharton, the online business journal of The Wharton School of the University of Pennsylvania. The project was sponsored by SAP.

www.SAP.com http://knowledge.wharton.upenn.edu

Founded in 1881 as the first collegiate business school, the Wharton School of the University of Pennsylvania is recognized globally for intellectual leadership and ongoing innovation across every major discipline of business education. With a broad global community and one of the most published business school faculties, Wharton creates economic and social value around the world. The School has 5,000 undergraduate, MBA, executive MBA, and doctoral students; more than 9,000 annual participants in executive education programs; and a powerful alumni network of 92,000 graduates.

About Knowledge@Wharton

Knowledge@Wharton is the online business analysis journal of the Wharton School of the University of Pennsylvania. The site, which is free, captures relevant knowledge generated at Wharton and beyond by offering articles and videos based on research, conferences, speakers, books and interviews with faculty and other experts on global business topics.

For more information: knowledge.wharton.upenn.edu

About SAP

As a market leader in enterprise software, SAP helps companies of all sizes and industries run better. From back office to boardroom, warehouse to storefront, on premise to cloud, desktop to mobile device — SAP empowers people and organizations to work together more efficiently and use business insight more effectively to stay ahead of the competition. SAP applications and services enable customers to operate efficiently, adapt continuously, and perform sustainably. SAP has more than 253,500 customers in 188 countries, more than 66,500 employees and locations in more than 130 countries and a 42-year history of innovation and growth as a true industry leader.

For more information, please visit www.SAP.com