TDWI RESEARCH

FIRST QUARTER 2014

TDWI BEST PRACTICES REPORT

Predictive Analytics for Business Advantage

By Fern Halper

tdwi.org

tdwi

Research Sponsors

Actuate

Alteryx

Pentaho

SAP

Tableau Software

BEST PRACTICES REPORT

FIRST QUARTER 2014

Predictive Analytics for Business Advantages

By Fern Halper

Table of Contents

About the Author

FERN HALPER is director of TDWI Research for advanced analytics, focusing on predictive analytics, social media analysis, text analytics, cloud computing, and other "big data" analytics approaches. She has more than 20 years of experience in data and business analysis, and has published numerous articles on data mining and information technology. Halper is co-author of "Dummies" books on cloud computing, hybrid cloud, service-oriented architecture, and service management, and *Big Data for Dummies*. She has been a partner at industry analyst firm Hurwitz & Associates and a lead analyst for Bell Labs. Her Ph.D. is from Texas A&M University. You can reach her at fhalper@tdwi.org.

About TDWI

TDWI, a division of 1105 Media, Inc., is the premier provider of in-depth, high-quality education and research in the business intelligence and data warehousing industry. TDWI is dedicated to educating business and information technology professionals about the best practices, strategies, techniques, and tools required to successfully design, build, maintain, and enhance business intelligence and data warehousing solutions. TDWI also fosters the advancement of business intelligence and data warehousing research and contributes to knowledge transfer and the professional development of its members. TDWI offers a worldwide membership program, five major educational conferences, topical educational seminars, role-based training, onsite courses, certification, solution provider partnerships, an awards program for best practices, live Webinars, resourceful publications, an in-depth research program, and a comprehensive Web site, tdwi.org.

About the TDWI Best Practices Reports Series

This series is designed to educate technical and business professionals about new business intelligence technologies, concepts, or approaches that address a significant problem or issue. Research for the reports is conducted via interviews with industry experts and leading-edge user companies and is supplemented by surveys of business intelligence professionals.

To support the program, TDWI seeks vendors that collectively wish to evangelize a new approach to solving business intelligence problems or an emerging technology discipline. By banding together, sponsors can validate a new market niche and educate organizations about alternative solutions to critical business intelligence issues. Please contact TDWI Research Director Fern Halper (fhalper@tdwi.org) to suggest a topic that meets these requirements.

Acknowledgments

TDWI would like to thank many people who contributed to this report. First, we appreciate the many users who responded to our survey, especially those who responded to our requests for phone interviews. Second, our report sponsors, who diligently reviewed outlines, survey questions, and report drafts. Finally, we would like to recognize TDWI's production team: Jennifer Agee, Bill Grimmer, and Denelle Hanlon.

Sponsors

Actuate, Alteryx, Pentaho, SAP, and Tableau Software sponsored the research for this report.

Research Methodology and Demographics

Report Scope. Predictive analytics is fast becoming a decisive advantage for achieving a range of desired business outcomes, including higher customer profitability, stickier websites, and more efficient and effective operations. Predictive analytics involves methods and technologies for organizations to spot patterns and trends in data, test large numbers of variables, develop and score models, and mine data for unexpected insights. This report examines users' drivers, experiences, and best practices for improving business advantage with predictive analytics.

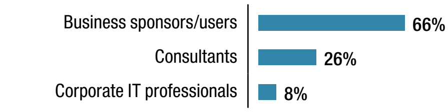
Survey Methodology. In August 2013, TDWI sent an invitation via e-mail to business and IT executives; VPs and directors of BI, analytics, and data warehousing; business and data analysts; data scientists; IT application managers; and other BI/DW professionals, asking them to complete an Internet-based survey. The invitation was also delivered via websites, newsletters, and publications from TDWI. The survey drew 580 responses. From these, we excluded incomplete responses as well as some respondents who identified themselves as vendors or academics. The resulting 373 responses form the core data sample for this report. Of these, 52% were investigating the technology (20% of these were engaged in a predictive activity), 34% were actively using it, and 14% had no plans for it.

Survey Demographics. The vast majority of survey respondents are business sponsors or users (66%). Included in this group are executives as well as business analysts, data scientists, and others involved in data analysis. The remainder consists of IT professionals (8%) and those who identified as consultants (26%). We asked consultants to fill out the survey with a recent client in mind.

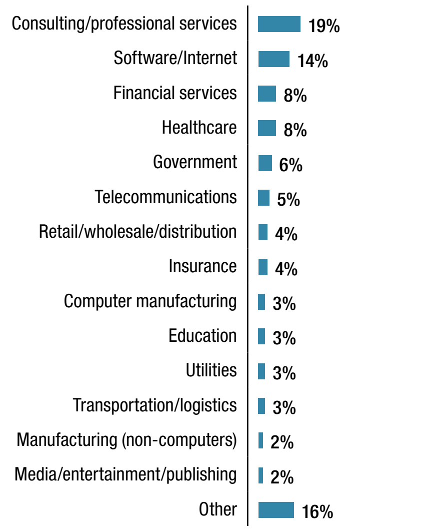
Respondents from consulting and professional services organizations made up the largest industry segment (19%), with software/Internet services (14%) and financial services and healthcare (8%) next highest. Most of the respondents reside in the United States (47%), followed by Europe (18%) and Asia (11%).

Other Research Methods. TDWI conducted telephone interviews with business and IT executives, VPs and directors of BI, and experts in predictive analytics. TDWI also received briefings from vendors that offer related products and services.

Position

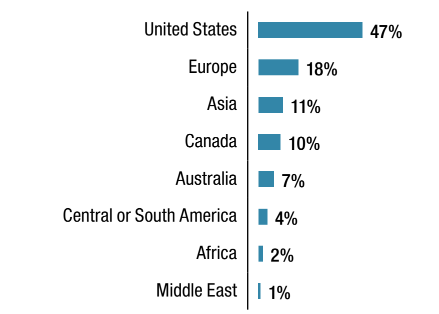


Industry

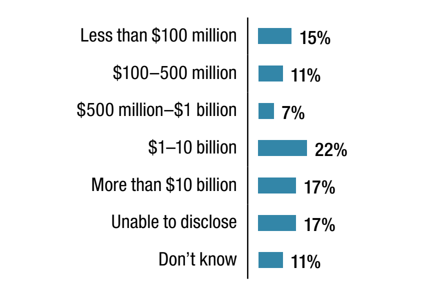


("Other" consists of multiple industries, each represented by 2% or less of respondents.)

Geography



Company Size by Revenue



Based on 373 survey respondents.

Executive Summary

Predictive analytics has finally become mainstream. It is being used by organizations from marketing and sales to finance and operations to achieve better business performance.

To compete effectively in an era in which advantages are ephemeral, companies need to move beyond historical, rear-view understandings of business performance and customer behavior and become more proactive. Organizations today want to be predictive; they want to gain information and insight from data that enables them to detect patterns and trends, anticipate events, spot anomalies, forecast using what-if simulations, and learn of changes in customer behavior so that staff can take actions that lead to desired business outcomes. Success in being predictive and proactive can be a game changer for many business functions and operations, including marketing and sales, operations management, finance, and risk management.

Although it has been around for decades, predictive analytics is a technology whose time has finally come. A variety of market forces have joined to make this possible, including an increase in computing power, a better understanding of the value of the technology, the rise of certain economic forces, and the advent of big data. Companies are looking to use the technology to predict trends and understand behavior for better business performance. Forward-looking companies are using predictive analytics across a range of disparate data types to achieve greater value. Companies are looking to also deploy predictive analytics against their big data. Predictive analytics is also being operationalized more frequently as part of a business process. Predictive analytics complements business intelligence and data discovery, and can enable organizations to go beyond the analytic complexity limits of many online analytical processing (OLAP) implementations. It is evolving from a specialized activity once utilized only among elite firms and users to one that could become mainstream across industries and market sectors.

This TDWI Best Practices Report focuses on how organizations can and are using predictive analytics to derive business value. It provides in-depth survey analysis of current strategies and future trends for predictive analytics across both organizational and technical dimensions including organizational culture, infrastructure, data, and processes. It looks at the features and functionalities companies are using for predictive analytics and the infrastructure trends in this space. The report offers recommendations and best practices for successfully implementing predictive analytics in the organization.

There is a definite shift in the builder and consumer of predictive analytics models.

TDWI Research finds a shift occurring in the predictive analytics user base. No longer is predictive analytics the realm of statisticians and mathematicians. There is a definite trend toward business analysts and other business users making use of this technology. Marketing and sales are big current users of predictive analytics and market analysts are making use of the technology. Therefore, the report also looks at the skills necessary to perform predictive analytics and how the technology can be utilized and operationalized across the organization. It explores cultural and business issues involved with making predictive analytics possible.

A unique feature of this report is its examination of the characteristics of companies that have actually measured either top-line or bottom-line impact with predictive analytics. In other words, it explores how those companies compare against those that haven't measured value.

Predictive Analytics: A Technology Whose Time has Come

Predictive analytics—a statistical or data mining solution consisting of algorithms and techniques that can be used on both structured and unstructured data to determine outcomes—is certainly not a new technology. In fact, it has been used on structured data for decades. However, market adoption and visibility of the technology is increasing for a number of reasons:

Computing power increases. Processing speed and memory have been increasing at an exponential rate. This has been publicized by the popular press. For instance, *TIME* magazine noted that the average smartphone in 2012 had more computing power than Apollo 11 did when it went to the moon in 1969.[1] Computing power is also cheaper. What does higher computing power at a lower cost mean for predictive analytics? In the past, it might have taken hours or days to run a predictive model that now takes minutes. Historically, it was often difficult to afford the computing power needed to interpret data that might be changing in real time. The lack of affordable computing power also made it difficult to integrate the output of a model into a business process, i.e., to operationalize it. With computing power increasing and the price per CPU dropping, predictive analytics is now much more practical for organizations to use.

Value is better understood. As companies put a solid BI foundation in place, they begin to look for other ways to derive value from their data. Many companies want to take BI to the next level. In fact, more than 90% of organizations that responded to this best practices survey and had a predictive analytics initiative either under investigation or under way feel their enterprise has a solid BI foundation in place. These organizations want to understand what actions their customers will take. They want to better predict failures in their infrastructure. They understand the value of predictive analytics.

Economic considerations. The recession has affected how businesses operate. Increasingly, organizations realize that data is a competitive asset and that predictive analytics can be an important tool in the analytics arsenal to help achieve business advantage. Adopters realize that it is not enough to look in the rearview mirror to gain insight and remain competitive. To be successful in a competitive environment, companies must utilize data and analytics to its fullest advantage. In fact, improving business performance was cited as a top driver for predictive analytics by survey respondents (see Figure 1, page 7).

Big data fuels the fire. As the amount of data continues to explode, enterprises are looking for ways to more effectively manage and analyze it for competitive advantage. Predictive analytics has been cited as a key form of analytics for big data. This has helped to drive the popularity of the technology. For example, 73% of companies surveyed are utilizing predictive analytics on a big data initiative.

Ease of use. As the market becomes increasingly aware of the power of predictive analytics, vendors are trying to make predictive analytics easier to use and offer it in a way that is consumable by a variety of end users. Many vendors have tried to make predictive analytics more "user friendly" by automating some model-building capabilities. They are including better visualization capabilities to aid in pattern detection. They have also introduced ways to operationalize predictive analytics in business processes, which has opened up the technology to more end users. For example, results of a model to predict churn can be operationalized as part of a business process that includes the call center. The call center agent sees the results of the model and acts on it during a call—without even necessarily knowing that a predictive model was at work behind the scenes. Such capabilities have helped to drive the adoption of predictive analytics.

Although it has taken some time, predictive analytics is finally becoming a mainstream technology.

[1] Richard Stengel [2012]. "Making Sense of our Wireless World," *TIME* online, August 27.

Business Intelligence versus Predictive Analytics

Predictive analytics is deeper and more proactive than traditional BI.

Potential users getting started with predictive analytics want to know: What's the difference between predictive analytics and business intelligence (BI)?

Predictive analytics differs from traditional or descriptive BI in a number of ways. BI does a good job of slicing and dicing data to help answer questions such as what happened or what is happening, and perhaps even why it happened. However, BI generally provides static reports or dashboards and can be inflexible. With predictive analytics, however, users can estimate outcomes (often called targets) of interest. Outcomes might include: Who will disconnect a service? How much will something increase in value? Predictive analytics is deeper, more proactive, and doesn't require a predefined cube data structure.

Some people get caught up in terminology and arguments such as whether predictive analytics is a subset of BI. One way to think about the relationship between BI and predictive analytics is to consider the spectrum of analysis techniques: from static, historical reporting through more advanced techniques that move from reactive to proactive, and from historical to future. This is where predictive analytics comes into the picture. It is one of a number of analytics techniques that can be much more sophisticated than descriptive techniques (such as reporting or dashboards).

Drivers for Predictive Analytics

There are numerous reasons why the market for predictive analytics is increasing, but what are the drivers for *actual* user adoption of the technology? We asked respondents who were either utilizing the technology now or actively investigating it to score the importance of several drivers for predictive analytics. On the five-point scale, 1 was extremely unimportant and 5 was extremely important. Drivers centered on various aspects of the business such as customer understanding, operations efficiency, product development, and innovation (see Figure 1).

Drivers for predictive analytics include understanding behavior and trends.

Understanding trends and behaviors ranks high. At the top of the list of drivers was predicting trends (3.95), followed closely by understanding customers (3.93) and predicting behavior (3.85). Clearly, respondents are interested in predictive analytics' ability to discern trends and patterns in data for a variety of reasons, one of which is understanding customers and customer behavior. In fact, customer-related analytics such as retention analysis and direct marketing are a top use case for predictive analytics.

Business process reasons are also important. In addition to understanding trends and behaviors, respondents were also interested in using predictive analytics to drive better business performance (3.89), strategic decisions (3.85), and operational efficiency (3.78).

Although there was little difference in the rating for the top drivers between those using predictive analytics now and those investigating it, it is interesting to compare the top drivers to the lower-rated drivers. For instance, respondents from both groups seemed less driven by more forward-looking uses of the technology (such as responding faster to change or using predictive analytics as a competitive differentiator) than they were about helping to drive better business performance. This was the case regardless of how long the respondent had been using the technology and indicates that predictive analytics is still relatively new for most organizations using it.

Please rate the drivers for predictive analytics in your organization or company on a scale of 1–5, where 1 is extremely unimportant, 2 is somewhat unimportant, 3 is neither important nor unimportant, 4 is important, and 5 is extremely important.

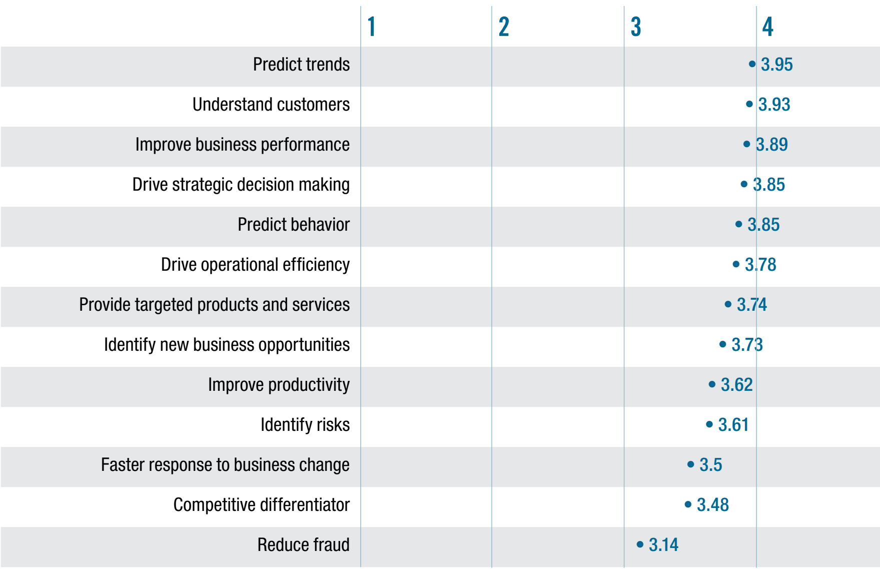


Figure 1. *Drivers for predictive analytics. Based on 329 respondents.*

The State of Predictive Analytics

Predictive analytics may be a technology whose time has finally come, but this doesn't mean it is widespread and part of the corporate culture. To understand the current state of predictive analytics, we asked respondents about where they are in their predictive analytics efforts as well as how and where predictive analytics is being used in their companies. For those who aren't using it, we asked, "why not?"

Predictive analytics is slowly gaining traction in organizations.

Predictive Analytics Adoption

Predictive analytics is making its way into organizations. Although predictive analytics is becoming a mainstream technology, it is still relatively new in most organizations. About half of the respondents to our survey were actively investigating the technology now, indicating increased interest in the technology. About 20% of these respondents have some predictive analytics activity already under way, perhaps a proof of concept (POC) or other experiment. When this group is called out specifically in this report, they will be referred to as the "investigating" group. Slightly more than 34% of the respondents are actively using predictive analytics. In this report, we'll refer to this group as the "active" group. Interestingly, when TDWI published a predictive analytics Best Practices Report in 2007, only 21% of the respondents had fully or partially implemented the technology. This suggests that predictive analytics is growing in market adoption.

Predictive analytics appears to be growing in adoption.

The remainder of the respondents (about 14%) are not using predictive analytics yet, most often citing their current focus on basic BI deployments.

Analytics is becoming part of the decision-making process. Many companies still do not use analytics to drive business decision making. We asked respondents who were either investigating the technology or utilizing it: "Would you say that analytics underpins your organization's business strategy and drives day-to-day decisions?" Twenty-five percent of the respondents answered "Definitely," although another 48% replied "Somewhat." Therefore, although organizations are just starting their analytics journeys, some inroads are being made in terms of utilizing analytics for decision making. We will see later in this report, however, that when analytics is standardized in a company, it does seem to drive top- and bottom-line impact.

Use Cases for Predictive Analytics

Predictive analytics is being used today primarily in marketing and sales.

Companies are using or want to use predictive analytics for a range of applications, from predicting consumer behavior to predicting machine failure to finding patterns in medical data. However, the top use cases among the active group currently center on sales and marketing. We asked active-group respondents to tell us what use cases they were currently using, planning to use, or had no plans to use for predictive analytics (see Figure 2).

What is predictive analytics being used for in your company? Now? Three years from now?

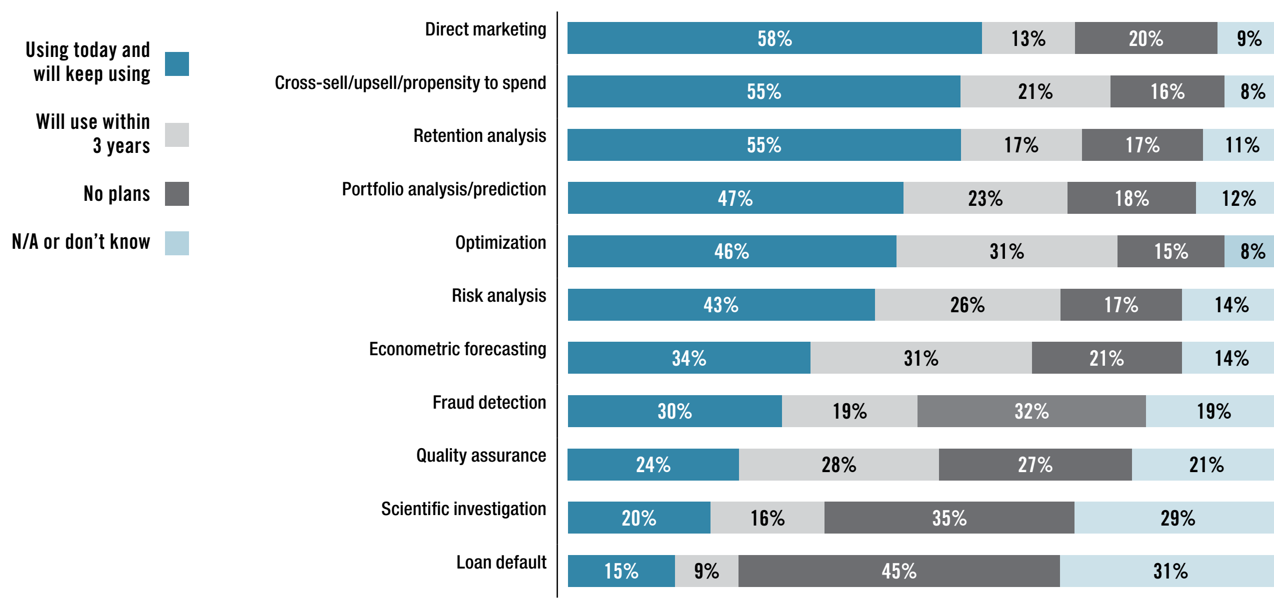


Figure 2. *Based on 126 active respondents.*

Retention analysis is a top use case among those currently deploying predictive analytics.

Marketing and sales analysis currently lead the way. The top use cases for predictive analytics among the active group include direct marketing (58%), cross-sell and upsell (55%), and retention analysis (55%). In fact, predictive analytics is currently being used primarily in marketing (64%) and sales (54%) by respondents in the active group (see Figure 3). Companies clearly want to predict customer response to direct marketing campaigns and be able to upsell or cross-sell a customer. They also want to be able to stem customer attrition. These are all high-impact activities that can improve top-line revenue. Retention analysis is also very important to those investigating the technology. Roughly 72% (not shown) of these respondents plan to use predictive analytics for retention analysis over the next few years.

Other analysis (such as portfolio and risk analysis) start to gain steam. Looking at the next three years, another set of applications will start to make more headway in organizations, including optimization, risk analysis, and portfolio analysis. If respondents stick to their plans, optimization-related predictive analytics will be used by close to 80% of the active-group respondents within three years—a higher percentage than the direct marketing use case during the same time period.

Optimization and risk analysis involve reducing the probability of negative outcomes. Risks come in many shapes and sizes, including financial risks, turnover risks, or even risks associated with loss of life. The increase in these other kinds of analysis suggests that other use cases are finding their way into the organization. In fact, such analysis leads the way for those currently investigating predictive analytics. For instance, about 66% of respondents investigating predictive analytics plan to use it for risk analysis, and 70% plan to use it for portfolio analysis over the next three years (not shown).

Organizations outside of marketing and sales join in. In Figure 3, marketing and sales are the two most popular areas of the company that use predictive analytics today among the active group. However, predictive analytics is also finding its way into other areas of the business, such as customer service, finance, and operations management. It will continue to grow in these areas over the next few years. Typically, once an enterprise starts to enjoy success with an emerging technology in one area, it will see the technology organically spread to other areas. Success breeds success. That appears to be the case here.

Where is predictive analytics used in your company? Now? 3 years from now?

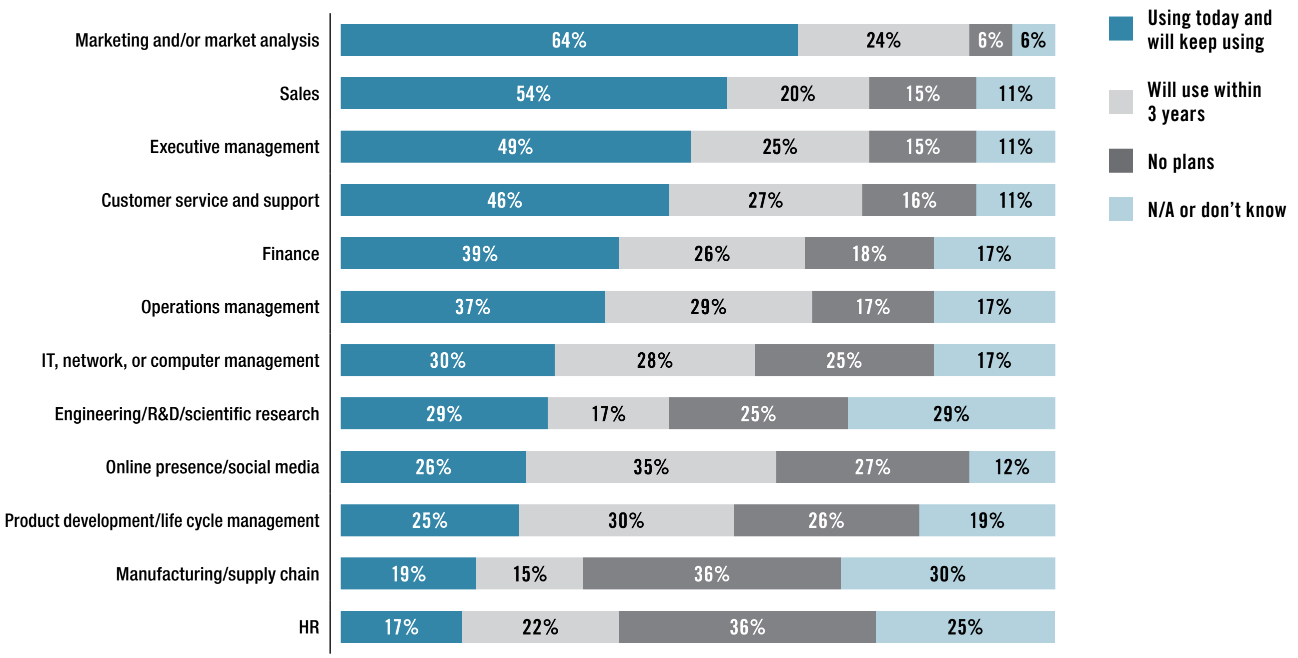


***Figure 3.** Based on 126 active respondents.*

Data, Data, Data

Although companies are primarily using structured data found in data warehouses or other data stores in their predictive analytics efforts, they are looking to broaden the range of data sources used. Not surprisingly, almost 100% of respondents using predictive analytics today are using it on

Companies are starting to make use of other sources of data for predictive analytics (aside from structured data).

structured data (see Figure 4). The second most popular data source is demographic information (77%) followed by time series data (65%). Demographic data includes census information or information about business size, revenue, and so on. Such data can be useful in helping to predict particular outcomes. For instance, consumer demographic data can be useful in determining who will respond to an offer. Business demographic data can be used by service companies to determine how the size, location, or length of time in business affects purchase patterns.

What kind of data do you use for predictive analytics? Now? Three years from now?

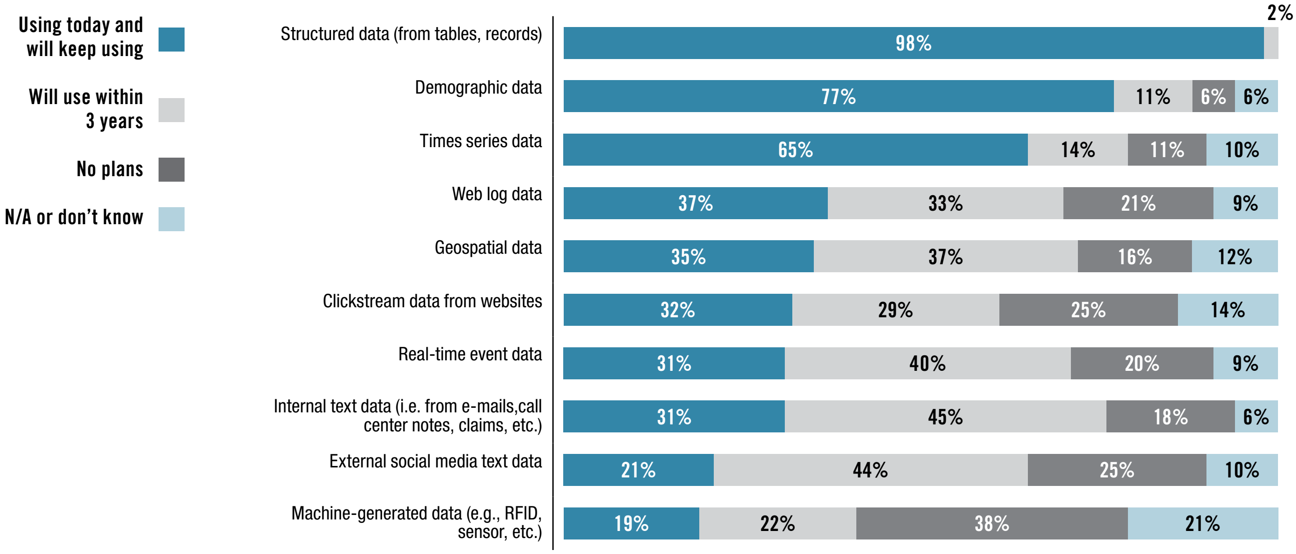


Figure 4. Based on 126 active respondents.

Web log, clickstream, and geospatial data are finally starting to gain traction. Newer sources of data are also becoming part of the mix. Respondents cited using Web log data (37%), clickstream data (32%), and geospatial data (35%) today for analysis, and it looks like these percentages will grow significantly over the next three years. For example, a media company might use clickstream data to look at the behavior of its anonymous site visitors versus its paid subscribers or to forecast impressions for its advertising campaign. Geospatial data (sometimes referred to as location data) is also growing in popularity. Companies are using geospatial data and geospatial analytics in applications ranging from marketing to operations. They are also using it in conjunction with other data. Analytics is moving past mapping to more sophisticated use cases such as visualization and predictive analytics. For example, geospatial analysis can help companies identify problem spots and turn off network traffic in certain places.

Companies are starting to use disparate data types in predictive analytics.

Text data will gain momentum. Survey results suggest that companies will start to incorporate even more text data (often referred to as "unstructured data," although that term might include other forms of non-traditional data such as video and audio) for use in analytics and predictive analytics. Thirty-one percent of the active group reported using internal text data today, and the percentage will double over the next three years if users keep to their plans.

Text data comes from internal sources such as e-mail, log files, call center notes, claims forms, and survey comments, as well as external sources such as tweets, blogs, and news. Text data can be a valuable source of input to predictive models because it often addresses the question of *why*. Why did a customer switch to another provider? Why did a customer buy one product and not another?

Increasingly, companies are combining text data with structured data in predictive analytics in an attempt to increase the effectiveness or lift of a model. This is a popular use case for combining structured and text data. For example, in churn analysis, companies combine the text from call center notes (which includes insight about why a customer called as well as sentiment associated with the call) with structured data such as demographic data. Using text analytics to extract entities, themes, or sentiments provides additional sources of attributes for a predictive model.

Real-time data is poised to grow.

Real-time data is poised for growth. Companies often view real-time data as a second phase of their advanced analytics strategies. Of course, this varies by user. Some wireless providers, for instance, have established real-time analytics projects to monitor and predict network health. These may be separate projects from the mainstream analytics work. Real-time data will also be used with operational predictive analytics, such as in operational intelligence systems that make decisions automatically.

Big data continues to get bigger. Of course, big data adoption is also on the rise. Big data is a big topic and is covered in its own section of this report (see page 22).

Challenges and Barriers to Adoption

Although the adoption of predictive analytics is rising, respondents still face a number of challenges, many of which are related to people and processes rather than technology. Challenges also vary based on how far an organization has progressed in its predictive analytics efforts. These challenges are shown in Figure 5. Both the active group and those investigating the technology weighed in on the challenges.

Lack of skills and understanding of technology are top challenges. For the active group, lack of skills was cited as the top challenge (24%). For those investigating the technology, this challenge ranked third on the list (at 19%), behind lack of understanding of predictive analytics (30%) and lack of a strong business case (21%). Predictive analytics can be quite complex. The skills required to build a predictive model include technical skills as well as analysis and critical thinking skills. Although vendors have come a long way in making predictive analytics easier to use, that doesn't negate the need for a particular skill set to build a very advanced model. Staffing is a major issue in predictive analytics and needs to be addressed as part of the planning process. Predictive analytics adopters would do well to heed this factor.

Lack of understanding of predictive analytics is also a key challenge. In a separate question, we asked: "Where would you like to see improvements in your predictive analytics deployment?" Seventy percent of all respondents (not shown) answered "education." Education is needed to understand what predictive analytics is all about so organizations can understand opportunities and challenges. It's not just a matter of education about the technology. As one respondent said, "There is a lack of understanding of the business potential" for predictive analytics in the organization, as well.

What are your top predictive analytics challenges? Select three or fewer.

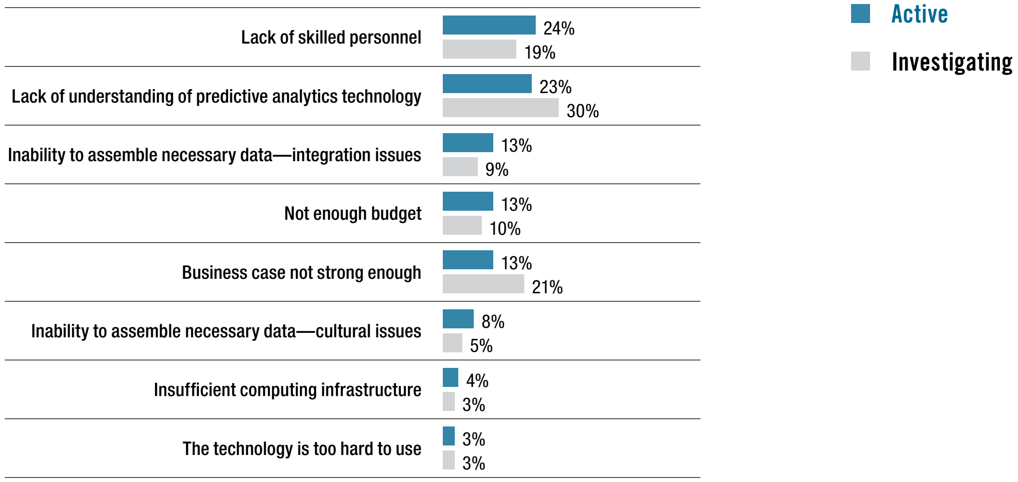


Figure 5. *Based on 126 respondents in the active group and 195 in the investigating group.*

Organizing to execute. The survey did not explicitly inquire about organizational structure, but the subject was raised in discussions with respondents about challenges. Some organizations have hired chief analytics officers, a role that is a bit different from a BI executive who, in some companies, is more concerned with IT pieces and doesn't necessarily understand how data is a corporate asset. A chief analytics officer actually helps to drive analytics in an organization. Other organizations have assembled centers of excellence (COEs) that support the whole organization. These COEs consist of people with different analytics skill sets. Within the center, different teams often support different parts of the business, although they share best practices. The short user story below illustrates the team approach.

> **USER STORY A TEAM APPROACH TO PREDICTIVE ANALYTICS.**
> Some companies have found that a team approach can be useful in solving complex predictive analytics problems. "It helps to assemble small teams that can get all the value out of the data," said a director of healthcare analytics. "Occasionally you can find one data scientist rock star who can do it all. We use small teams that change data formats, integrate data, apply advanced modeling, interpret results, and communicate those results to the company. We found a balance with the small-team approach (around three people) for many of our projects."

Business case woes. Lack of a strong business case is also a challenge for organizations, especially those just starting out with predictive analytics. Twenty-one percent cited this as a top challenge. Some respondents stressed the need for a proof of concept or proof of value (POV). Ideally, the POV is a collaborative effort between business and IT. The POV is a small yet manageable project that can show impressive results. This can help to make the case for predictive analytics.

Cultural issues can impede progress with predictive analytics.

Culture is also an issue. Looking at challenges from another angle, we asked the active group, "How satisfied are you with the following aspects of your predictive analytics deployment?" See Figure 6. Respondents used a five-point scale, where 1 was completely dissatisfied and 5 was completely satisfied. The responses pointed to the fact that cultural challenges can be just as significant as technical and business challenges. Culture around analytics barely scored a neutral rating. For instance, some users commented about the "lack of enterprise desire to change" and "organizational readiness." One respondent put it this way:

> The hard part about predictive analytics is not the technical change. It is not the information technology change. It is a behavior change. Analytics don't just happen. For instance, accountability for solid data is part of the process. This might start with the data entry team right at the beginning of the analytics process.

Companies that have been using predictive analytics for years realize that it takes time for an analytics culture to permeate an organization. As one respondent put it, "Analytics work best when they are not front and center. If it has to force itself to be heard or used, then something is wrong." This is true, but it also takes time for analytics to become part of the culture. In addition, analytics often requires plenty of selling because you are changing how decisions and operating procedures are made and implemented. Everyone needs to be on board.

Rather than ripping and replacing systems, this company took managed steps toward a solution. Many of the respondents talked about taking steps as part of implementing predictive analytics. For instance, according to this respondent, the idea was to "Bring people along at a comfortable level while piloting new functionality. Ultimately, it's about building trust." Building that trust means collaboration between different parts of the business and building relationships. That can take time.

How satisfied are you with the following aspects of your predictive analytics deployment?
(Please rate on a scale of 1–5, where where 1 is completely dissatisfied and 5 is completely satisfied.)

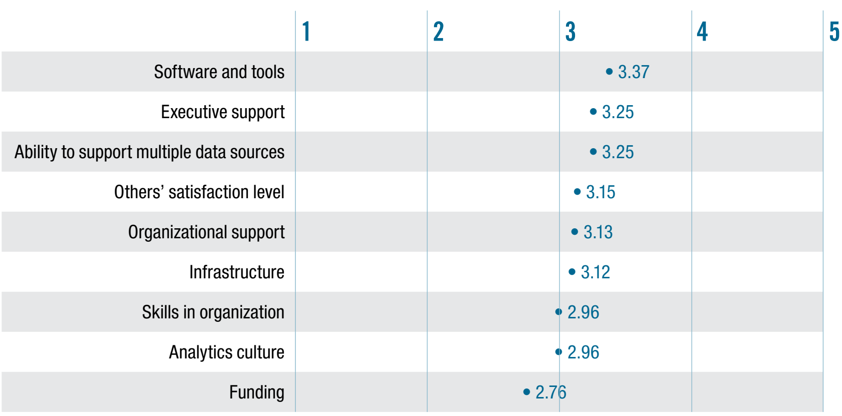


Figure 6. *Based on 126 active respondents.*

Current Value

Despite the challenges companies face, they still are experiencing value from predictive analytics. We asked the active group: "Overall, how satisfied are you with predictive analytics in your company?" Forty-four percent responded either "satisfied" or "completely satisfied," 40% responded "neutral," and only 16% were either "dissatisfied "or "completely dissatisfied" (not shown).

We also asked the active group what value they have measured using predictive analytics. (See Figure 7.) Forty-five percent of the respondents were able to measure a positive top- or bottom-line impact using predictive analytics. Another 30% believe that they have become more effective or efficient but have been unable to measure the impact. The remainder believes they have gained more insight from predictive analytics.

Later in this report, we examine some of the characteristics of companies that have gained measurable value from predictive analytics.

Which statement best describes the value you've seen from your predictive analytics efforts?

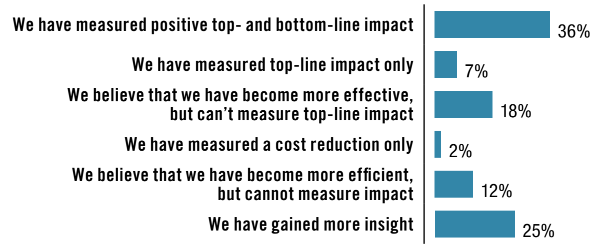


Figure 7. *Based on 126 active respondents.*

User Skills and Delivery Models

Say Hello to the Business User

Business analysts as well as statisticians are building and utilizing predictive analytics.

Along with the movement toward mainstream predictive analytics comes the movement to democratize it. In other words, a market goal has been to make the software easy enough to use to enable business analysts to build predictive models and make them consumable enough so that a host of end users can utilize them. Just five to seven years ago, statisticians or mathematicians (or others with quantitative backgrounds and advanced degrees) built most predictive models. They delivered the output in reports or tried other ways to incorporate the information into operations. Today, a shift is occurring in which data scientists and statisticians as well as business analysts are building these models.

In fact, when we asked respondents from the active group who is building predictive models, the top two answers were the statisticians/data scientists (76%) and business analysts (63%). (See Figure 8.) Of course, the kinds of analysis that the two groups perform might differ. For instance, statisticians or data scientists might make use of more complex data types such as time series, or they might be responsible for analyses where the cost for inaccuracy is high, such as a pricing model.

Who in your organization (or company) is using predictive analytics to actually build models? Select all that apply.

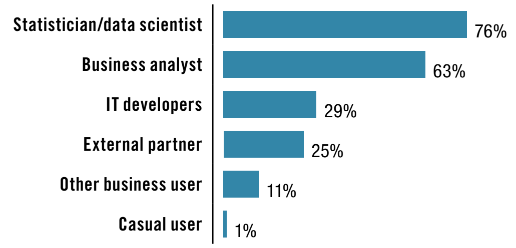


Figure 8. *Based on 126 active respondents.*

Regardless, the reality is that the builder role is changing. One reason is that vendors have made their software easier to use. They are providing wizards and other tools to guide or even suggest specific models to users. However, building a predictive model remains complex. It includes getting the data in shape for modeling as well as determining what variables to actually use. Expertise is required.

Get Ready for a Different Skill Set

We asked respondents in both the active and the investigating groups what skills were needed to perform predictive analytics. Rather surprisingly, both groups ranked a degree in statistics, math, or another quantitative discipline near the bottom of the list. This was true even for those who are currently using the technology. Figure 9 shows the percentage of respondents who believe the skills listed are necessary "to a large extent" in order to perform predictive analytics.

Respondents believe that knowledge of the business and critical thinking are key skills for predictive analytics.

To what extent do you believe the following skills are necessary to perform predictive analytics? (Not at all, to a little extent, to some extent, to a moderate extent, to a large extent.)

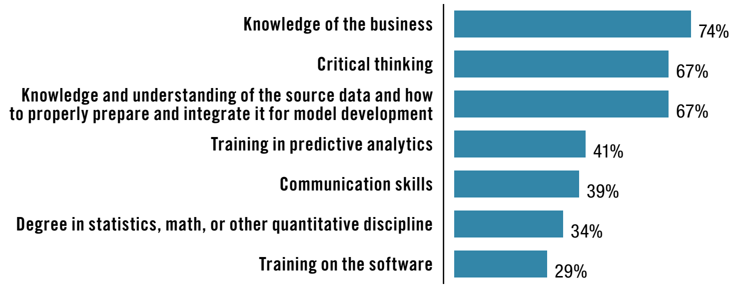


Figure 9. *Skills ranked necessary "to a large extent" for performing predictive analytics. Based on 330 respondents.*

The top skills cited were knowledge of the business (74%), critical thinking (67%), and knowledge of the source data and how to prepare it for analysis (67%). There is clearly a move to make predictive analytics easier to use and consume. In fact, respondents stated that in the near future, business analysts will be the top users of predictive analytics tools. Even the active group believes that the business analyst (86%) will be the primary user of predictive analytics tools in the near future (see Figure 10).

In the near future, who do you expect will be using predictive analytics tools in your company? Select all that apply.

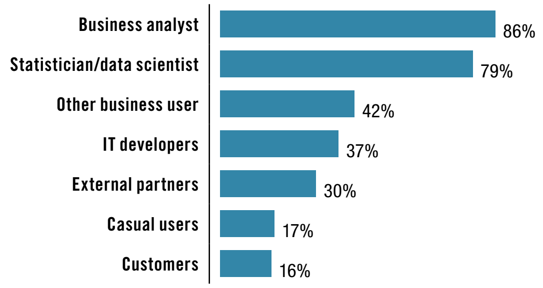


Figure 10. *Based on 126 active respondents.*

A cost-benefit approach can be utilized to manage the skills gap.

The question is whether these users realistically have the skills necessary to build models. The answer may be both yes and no. Business analysts often write complex SQL and get involved with sophisticated analysis, but their success will depend on the complexity of a given model. As discussed earlier, there is a higher cost for inaccuracy for some models than for others. Organizations should consider this factor as they decide who should build models. It will also be an important factor in staffing, as sophisticated model-building skills are in short supply. However, there are several methods for deploying predictive analytics where the business analyst is a key consumer of a model. These are discussed in the next section.

USER STORY A COST-BENEFIT APPROACH TO PREDICTIVE ANALYTICS.
The answer to who should be building the model often depends on the results of a cost-benefit analysis—especially when staffing is an issue. The VP of analytics at an international bank described it this way: "You need to factor in the cost of being wrong with an answer." In other words, if staffing and skills are a big issue, you need to plan your resources accordingly. He gave this example: "Say you're trying to build a model that predicts the probability of an action happening—like someone responding to a specific offer. You could pay a statistician to build that model, or you might ask a business analyst using software with wizards to build it. If the statistician builds it, you might get an additional 5% lift over the semi-automated approach. You need to ask yourself if it is worth the cost. You also need to ask whether what you're trying to model can take the hit. For instance, building a response model to an offer that is not as accurate as it could be is less risky than building a complex price sensitivity model that isn't accurate."

Operationalizing Predictive Analytics

Predictive models can be deployed in a number of ways in a company.

There are many ways that predictive analytics can be deployed in an organization. Figure 11 illustrates some of the top options according to the active group. A statistician or business analyst can build the model and then share the results for decision making (34%). A statistician can build the model and then a business analyst or other business user can interact with it (28%). For instance, a data scientist might build a model which other members of the organization might use to perform a what-if analysis. In addition, a model can be built by statisticians or other internal staff and then become operationalized as part of a business process (31%). These deployment options help to make predictive analytics more consumable.

Which statement best describes how predictive analytics is deployed in your organization?

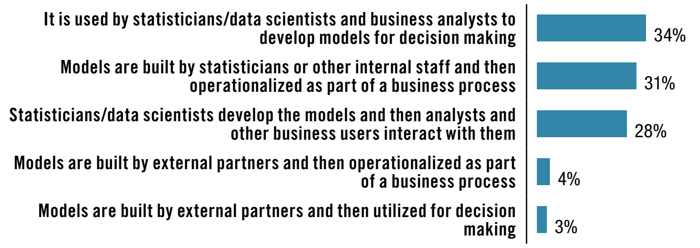


Figure 11. *Based on 126 active respondents.*

Here's how a few of these scenarios might work:

Data scientists set the foundation. In some companies, the data scientist is responsible for dealing with data issues related to feeding models that can be used by business analysts. For example, the data scientist might choose the input and target variables (i.e., the outcome of interest, such as fraud, churn, or positive response to an offer) for training a predictive model. The scientist is essentially determining "what can be modeled" by the business analyst. He or she might also construct derived variables or include other variables that might be useful in predictive models.

The business analyst then uses a software package geared to this approach to develop the models. Theoretically, the business analyst can't get into too much trouble because the data has already been determined. He or she can experiment and explore models that have, in essence, been approved by the data scientist. In some ways, this is analogous to OLAP BI for more advanced modeling. Using this approach, the business analyst can create numerous models without daily reliance on a statistician/data scientist to build the model. It can make the organization more agile.

Operationalizing the model. Some companies make predictive analytics available to greater numbers of users by operationalizing the models. For instance, suppose your company is interested in cross-sell and upsell opportunities (i.e., using predictive analysis to identify products or services to which customers are likely to respond). There might be three participants in this operationalizing scenario: the data scientist, the business administrator, and the call center agent. In this approach, the data scientist is responsible for developing the model (perhaps with the help of someone from marketing) and dealing with data issues related to feeding models.

Operationalizing predictive models as part of a business process makes the models more consumable.

The data scientist determines what models make sense for use with the call center and then passes the models off to a business administrator who can operationalize the model. The call center agent uses the model output without even necessarily knowing that there is a complex model working behind the scenes. All the call center agent might see is the next best offer to suggest to a customer with whom they are speaking.

The result is that a predictive model is taken to a wider set of end users in a one-to-many multiplier effect.

Tools, Techniques, and Processes

Top Techniques

Numerous techniques can be used for predictive analytics. We asked respondents to identify the kinds of techniques they are using or planning to use in their organizations.

Decision trees and linear regression lead the way. Decision trees and linear regression were the top two responses both for those using predictive analytics today and for those planning to use it. Both methods are fairly straightforward and relatively easy to understand. Linear regression tries to model the relationship between variables by fitting a line to the observed data. This simple model is widely used in statistics. It looks at the past relationship between variables to model the future. For instance, the price of a product might be strongly related to demand.

Decision trees and linear regression are two top techniques for predictive analytics.

Decision trees are often used for prediction because they are also fairly easy to understand, even by a non-statistician. A decision tree is a supervised learning approach that uses a branching or tree-like approach to model specific target variables or outcomes of interest. For instance, an outcome might be leave or stay; respond to or ignore a promotion; buy or not buy. A user would typically provide a set of training data (including data with known outcomes) to the tool. The decision tree then builds a model that can be interpreted as a set of rules with associated probabilities. For instance, in a churn model for a telecommunications company, a rule might be: "If a customer spends more than $200 a month for service, has been a customer for more than three years, and has not called the call center more than once a year, then there is an 80% probability that they will not churn." A test data set or a "holdout" sample is then used to see how well the rules perform with new data.

Current users: What are the most popular techniques for predictive analytics in your organization? Those in the investigation phase: Which techniques are you looking at? Select 3 or fewer.

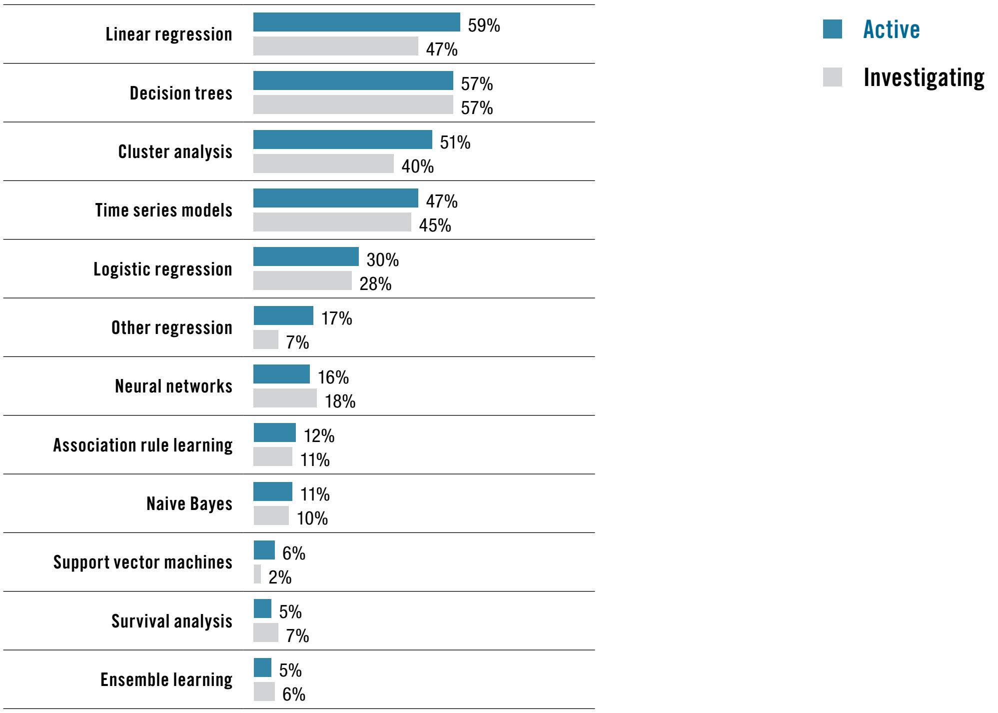


Figure 12. *Based on 126 respondents in the active group and 195 in the investigating group.*

Clustering and time series analytics are also popular.

Clustering and time series are also popular. As indicated in Figure 12, clustering and time series analysis are also popular techniques for predictive analytics. In fact, time series analysis seems to be more popular than clustering among those investigating the technology than by those actually using it. This might be because clustering is very useful in market segmentation, and marketing and sales are popular areas for predictive analytics among current users. Clustering is an unsupervised technique where grouping is based on similarities and the target variable is not known—such as a segment.

Time series analysis is used when there is a time-dependent nature to the data. It is very popular for forecasting. However, it is also being used in operations management and monitoring. More than 40% of respondents investigating predictive analytics cited this as a popular technique, which points to the value it can provide in a range of applications (such as operational intelligence).

Ensemble modeling is not widely used—yet. In ensemble modeling, predictions from a group of models are used to generate more accurate results. Only a small percentage of respondents cited ensemble learning as the most popular technique used, but it can be powerful and should garner more attention in the future.

Key Features and Processes Supporting Predictive Analytics

We asked the active group as well as those investigating the technology to rank certain functionality and processes in terms of level of importance on a five-point scale, where 1 is not at all important and 5 is extremely important. The rating results from the two groups are illustrated in Figure 13.

Data integration is key to predictive analytics.

Clearly, data integration is a key component of any predictive analytics effort. Both those using the technology as well as those investigating it ranked it at the top of the list. For those already using the technology, operationalizing (4.18) and ease of use (4.17) also ranked in the top three, no doubt for reasons already discussed. For those investigating predictive analytics, ease of use also ranked above a 4.

How important are the following currently to your predictive analytics efforts?

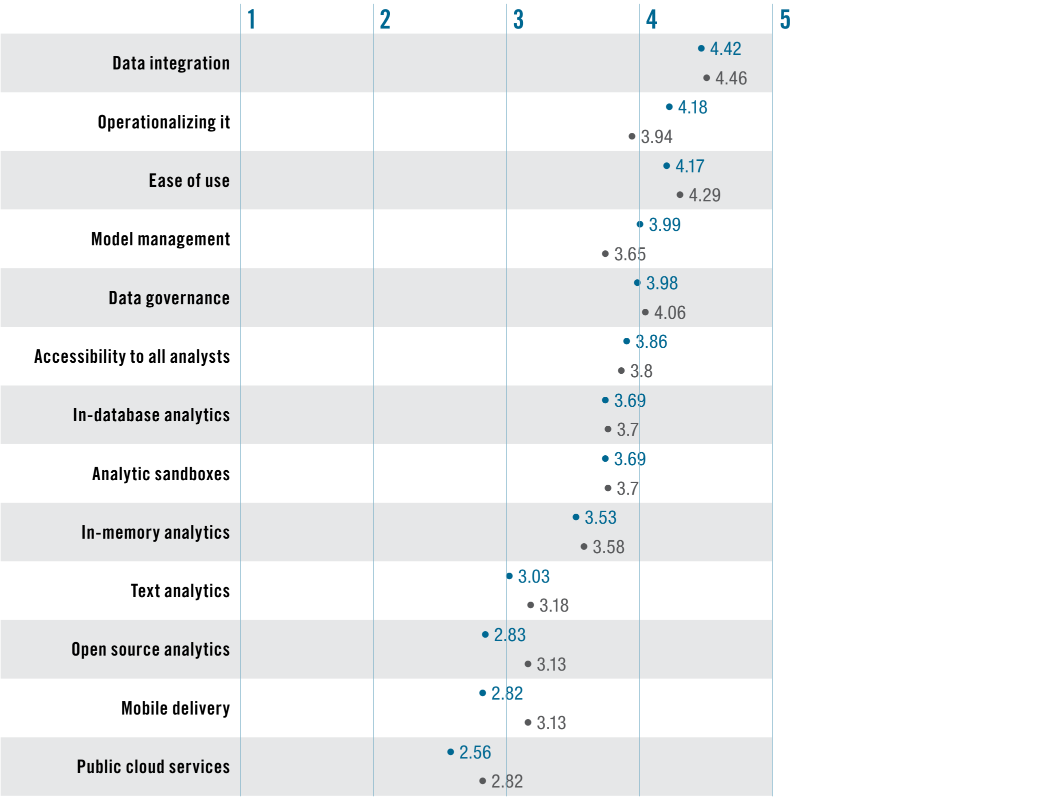


Figure 13. *Based on an active group of 126 respondents and investigating group of 195 respondents.*

Other interesting patterns in the results suggest best practices for predictive analytics.

Data integration. Data integration is a key component of an analytics infrastructure. If you can't utilize your data effectively, your models won't be as valuable. Integration becomes particularly important as companies start to use more disparate data types from different data sources. Data integration is not just about ETL. It is a family of techniques that includes data quality, master data management, data federation, and data blending. Data integration can actually have its own architecture.[2]

Those considering predictive analytics would do well to think about model management, too.

Think about model management. An interesting result from this question is that the active group realizes the importance of model management in predictive analytics (rating 3.99) significantly more than does the investigating group. Once an organization starts creating models, it needs a way to manage

[2] For more on data integration, please see the TDWI Best Practices Report *Next Generation Data Integration* by Philip Russom, available at tdwi.org/bpreports.

them. In addition, models cannot simply be built once and run forever. Internal and external variables can change, which can impact a model.

Model management can take different forms, such as a directory structure. Vendors also provide robust systems for model versioning that can send alerts to analysts when models might be going stale.

Data governance is also key.

Think about data governance. Both the active group and the investigating group realize the importance of data governance in predictive models. Data governance covers considerable territory: from deciding who owns and can access data to developing metadata to choosing appropriate data sources. For example, a government agency might use socioeconomic data to predict future teacher staffing requirements. However, if the agency hasn't agreed on which socioeconomic data to use, results can vary between departments. This governance can also extend to analytics. For instance, some companies are creating analytics councils that consist of members throughout the company who share models and best practices. This can be especially useful when analytics teams are geographically dispersed.

Think about operationalizing analytics. As discussed, operationalizing analytics (making analytics part of a business process) is one way to make it more consumable. This can take many forms. For instance, some organizations utilize the scores associated with certain models to route cases to appropriate operational groups. This approach is popular in claims analysis for fraud. Other organizations embed the models in systems that feed other operational systems such as call center systems (described earlier). Operationalizing analytics can take time, especially if internal systems can't support it.

In-memory computing and in-database analytics are gaining traction. Both in-memory computing and in-database analytics were ranked as important by respondents actively using predictive analytics and by those investigating it. In-memory computing refers to data processing where data is stored in memory to reduce disk I/O. Models can run faster with in-memory computing, which can be good for iterative models. In-memory computing is also useful for interactive work such as visualization and data discovery. In-database analytics embeds analytics in the database. When the amount of data is large, it can be cheaper when computation is closer to the data (rather than the other way around).

Open source analytics. Although open source analytics did not rate highly compared to some other features or functionality, it is becoming more popular—especially with those investigating predictive analytics. Open source solutions enable a wide community to engage in innovation. Tools such as open source R can be difficult to use in native form, but some vendors are creating front ends that make it more accessible. In addition, many universities are training students in these tools, which means that the hiring pool may be versed in them. One respondent put it this way: "Analytics like R provide more innovative packages and more sophisticated models. It may not be an enterprise system rigor-wise, but it helps us to come up with good initial model structure. It can also provide business rules that become part of a process."

Infrastructure for Predictive Analytics

Most of the respondents using predictive analytics today have a tried-and-true infrastructure in place for it, such as a data warehouse (89%), desktop applications (78%), or flat files on servers (75%). However, newer components are finding their way into the mix.

Analytic platforms are becoming more important to predictive analytics users.

Analytic platforms and appliances are gaining in popularity. Although the active group is currently running analytics using the tried-and-true infrastructure, there is definitely a move to newer forms of technology to support predictive analytics such as the analytics platform (55%) and the appliance

(41%). Analytics platforms can range from a database management system that manages data for analytics (typically query based), to a tool featuring a collection of algorithms, to any platform that happens to have analytics embedded. In addition, Figure 14 shows that if respondents stick to their plans, more than 80% of the active-group respondents will be using analytics platforms in the next three years. Slightly more than 60% will use an appliance of some kind.

What infrastructure do you have in place for predictive analytics? Now? Three years from now?

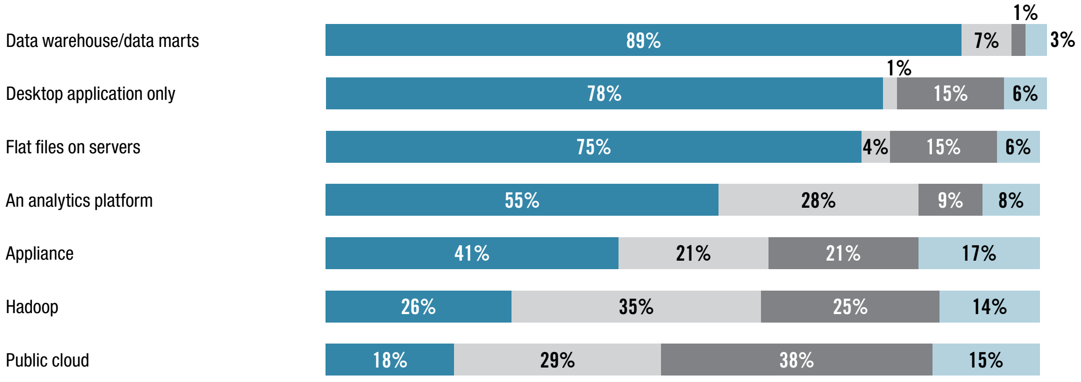


Figure 14. *Based on 126 active respondents.*

Likewise, those investigating the technology (not shown) are also looking at analytics platforms (73%) and appliances (46%) over the next three years to help support their predictive analytics needs.

Hadoop is gaining momentum among those using predictive analytics. Perhaps not surprisingly, 61% of the active group has plans for utilizing Hadoop over the next three years as part of their analytics efforts. Analytic ecosystems are the next wave, especially for those adopting big data solutions.

There is no one-size-fits-all approach to assembling a data infrastructure. An ecosystem approach might include a data warehouse along with Hadoop as the data management infrastructure supporting analytics. Most companies will want to make use of their existing infrastructure, so an ecosystem approach makes sense.

Predictive analytics in the cloud is limping along.

Clouds still not taking off. The public cloud for predictive analytics ranks low on the list for both current users and those planning to use it. Why so little interest? The are reasons vary. The next user story highlights some of these reasons as well as instances when the public cloud may be appropriate.

USER STORY ANALYTICS IN THE PUBLIC CLOUD.

When asked about the public cloud, one respondent said:

"We used to use a public cloud, really more of a hosting situation for analytics. As we grew, though, there were certain obstacles to analyzing data in the cloud environment for us.

"First was privacy, since most of the data we'd put in the cloud would be about our customers. We didn't want to lose control.

"Second was the volume issue. We have terabytes of data. Trying to push that over the wire was a big headache. Of course, some cloud vendors now have solutions for that, but we don't feel that they fully resolve the problem.

"Finally, we had a cost accounting issue with the pay-as-you-go financial model that made the economics infeasible.

"I think for data created in the cloud, it may be easier to analyze it in the cloud. If it doesn't start in the cloud and there is lots of it, then that is difficult. We do use certain apps that create data in the cloud, but it needs to be integrated with other data sources for analysis, so we bring it into our on-premises data warehouse."

Big Data and Predictive Analytics

No report about analytics would be complete without considering big data analytics. As mentioned, more than 70% of respondents in both the active and investigating groups claim to have some sort of big data push inside their organizations. Although big data analytics is relatively immature, it is still interesting to see what these companies are doing with big data and predictive analytics.

Analytics efforts are driving big data predictive analytics. Sixty-two percent of the respondents using big data predictive analytics stated that their big data efforts are part of their overall analytics efforts (not shown). Twenty-five percent stated it was a separate effort. That is not unusual. Although the ideal end state for most organizations will be a big data analytics ecosystem, the reality is that some companies are starting off in different places. For instance, a telecommunications provider might deploy a separate big data project to monitor the health of the network. This is often separate from the rest of the analytics going on in the company. In fact, it might lead the way for the rest of the company.

Big data isn't necessarily about predictive analytics. Although many people seem to lump together and confuse big data with predictive analytics, the study results suggest that most companies that claim to be using big data analytics are using it in BI reports and dashboards (85%) in conjunction with their existing data warehouse (83%). (See Figure 15.) No doubt, the majority of these respondents have large amounts of structured data (perhaps in the terabyte range) that they believe should be considered big data.

In-database analytics and in-memory databases are popular. Interestingly, close to 50% of those respondents who claim to have a big data effort under way use in-database analytics. Slightly over 40% use in-memory databases. These numbers are poised to almost double in the next three years.

Companies are considering many tools for their big data arsenals.

Enterprises have big plans for technology. Figure 15 also points to the potential growth of big data infrastructure and tools to support big data and predictive analytics. For instance, advanced analytics is projected to more than double, as is the use of real-time analytics and Hadoop. Of course, as companies start to implement big data projects, they will hopefully choose which tools and technologies are most suitable to solve the problem they have identified as worth solving. Whether these numbers actually come to fruition is as yet unclear.

What are the critical parts of your big data infrastructure? Now? Three years from now?

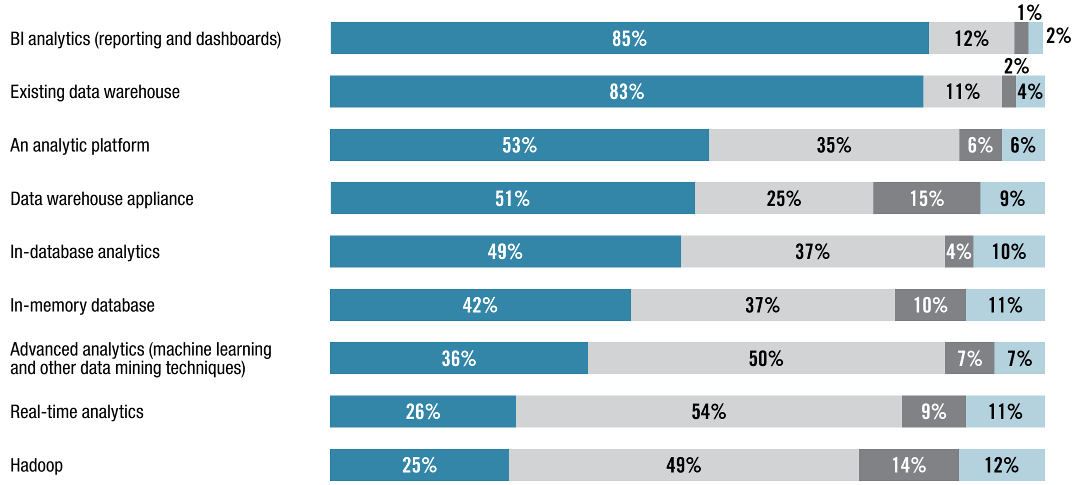


Figure 15. *Based on 242 respondents using big data.*

Companies face a host of challenges with big data analytics. Those respondents who are current users of predictive analytics and users of big data analytics report that their number one challenge is data integration complexity (44%), followed by lack of skills (25%) and data ownership and other political issues (25%) (not shown). Architecting the system was also relatively high on the list, with 24% citing this as a major challenge.

What Drives Measurable Value?

We have been citing best practices for predictive analytics throughout this report. However, in order to explore best practices further, we wanted to explore the characteristics of those companies that are monitoring and obtaining measurable value from predictive analytics. We also wanted to see if those companies differed from companies that were not reporting measurable value.

We separated those using the technology into two groups—those who measured either a top- or bottom-line impact (or both) from predictive analytics (73 respondents) and those who could not measure an impact (124 respondents). (Those who could not measure an impact stated that they thought they might gain efficiencies or insights, but could not measure anything.) Both groups contained respondents from the active group and those who are investigating the technology and were using it for less than one year.

We compared the two groups across a number of dimensions, including demographics, infrastructure, drivers, variety of data used, structure in the decision-making process, and the predominant user roles. This is the same data we've been utilizing throughout this report, but some interesting findings emerged by examining the data this way.

	Measured Value (n=73)	Didn't Measure Value (n=124)
Satisfaction (completely satisfied or satisfied)	47%	27%
Standardized analytics	26%	5%
Measured ROI	41%	10%
Disparate data types	41%	22%
Statisticians building models	87%	67%

Table 1. *Responses between two groups for certain characteristics.*

Satisfaction and value are linked. Not surprisingly, those who measured value were generally more satisfied with their predictive analytics efforts than those who didn't. That is not to say that those who are more deliberate about predictive analytics cannot be dissatisfied. Many were, especially regarding budget. However, this group was more satisfied (47% versus 27%) than those who could not measure value. See Table 1.

This suggests that it is not enough to simply perform predictive analytics and hope that you get some value out of it. It is important to put a program in place to actually measure this value, which includes defining the project candidates where predictive analytics is most likely to have an impact. This makes intuitive sense. If you can measure value and find that value is positive, then there is a good chance you're doing something right. Others will see this, too, and your efforts will be regarded as successful. As discussed earlier, success breeds success.

Given the survey data, there is no way to definitively know whether the group that measured an impact put a program together specifically for this purpose. However, the results suggest that these companies have a process in place and are more deliberate about their predictive analytics efforts (see below).

Companies that were able to measure value from predictive analytics were more likely to have a process in place for it.

A standardized analytics process drives value. Results also suggest that those respondents who were able to measure top- and bottom-line impact from predictive analytics were more likely to have a standardized analytics process or repeatable framework in place across the company than those who did not (26% versus 5%). This group was also more likely to say that analytics underpins their decision-making process (87% versus 74%; not shown). Finally, they also measured and achieved a return on investment (41% versus 10%). This may seem obvious or even circular and self-serving, but it underscores the need for a sound process and governance.

Those who have a standardized analytics process in place are more likely to be productive because they are not held up debating who owns the data, what the data means, how to approach the data, and so on. Note that just because something is standardized doesn't mean that innovation and experimentation don't occur. The majority of both groups had been using predictive analytics for one to five years, although those that measured value were more likely to have 5 years' experience. This also suggests it takes time to put the processes in place to measure value.

Utilizing disparate data types can add value to predictive analytics.

Adding different data types to the mix adds value. Interestingly, the data indicates that those who utilize disparate data types (other than structured and demographic data) were those who measured top- and bottom-line impact and were more satisfied with their efforts. Of course, this group might also be further along in their predictive analytics efforts and more sophisticated about them. This group was generally almost twice as likely to use geospatial data, clickstream data, or internal text data as part of their analysis. Such data can enrich a data set and potentially drive more meaningful results that can have measurable impact. These users may be seeing that impact.

Skills and culture play a role in value. Again, not surprisingly, those who have not measured value in their predictive analytics efforts are also more likely to have issues with skills and the analytic culture in their organization. When we asked respondents to rate how satisfied they were with particular aspects of analytics deployment, both groups rated funding the lowest (not shown). However, those who had not measured value had lower rating scores (2.8 range) for skills and culture compared to those who had measured value (3.3 range). We also checked between groups regarding infrastructure components supporting predictive analytics, and there wasn't much difference. Interestingly, though, those who saw value had a higher percentage of statisticians building models (82% versus 67%) than those who could not measure value.

The upshot? Those who measure positive impact more often have more experience with models, may have more sophisticated builders of the models, can make use of disparate data sources, and have an analytics-driven culture and structure in place to deal with analytics. This is a laudable goal and one that takes time.

Vendor Predictive Analytics Solutions

The firms that sponsored this report are among the leaders and innovators in analytics. The next section provides a brief overview of their solutions for predictive analytics. Their approaches vary, so they represent a good cross section of some of the solutions on the market today.

Actuate

Founded in 1993 and headquartered in Silicon Valley, Actuate Corporation provides development and deployment software to developers and OEMs (original equipment manufacturers). Actuate creates and supports Eclipse BIRT (2004) (the open source IDE) and BIRT iHub (a deployment platform) targeted to developers to help them design and deploy intuitive BI visualizations. The platform allows developers to develop business and customer analytics applications that integrate multiple sources and any volume of data, including unstructured data.

Actuate's BIRT Analytics solution provides easy-to-use visual data mining as well as predictive analytics capabilities to business analysts. Its advanced predictive analytics toolkit includes support for time series, decision trees, and purchase associations analysis. It also provides marketing-specific components for campaign management. It provides an in-memory engine and supports integration from Hadoop (including Cloudera, Hortonworks, and others) and virtually any source in any volume of big data. In its most recent release, the company's goal is to provide an easy-to-use interface for business analysts that does not require advanced knowledge of statistics.

Alteryx

Alteryx, founded in 2010, provides an analytics platform for line-of-business analysts looking to integrate, analyze, and share data analysis (via applications). Its flagship product, Alteryx Strategic Analytics, enables business analysts to combine various data sources such as location data, log files, social media, and other big data and analyze it on the platform. In addition, business analysts can produce analytic applications that can be shared via the private cloud or the Alteryx Analytics Gallery public cloud.

Alteryx supports predictive analytics based on open source R. The company supports a range of algorithms for predictive analytics with the goal to provide these techniques in a way that does not require coding. It now has the ability to drag and drop predictive analytics procedures directly into a workflow environment. Alteryx is headquartered in Irvine, California, with offices in Boulder and Silicon Valley.

Pentaho

Pentaho, founded in 2004, provides a business analytics platform that combines data integration and business analytics. It provides an enterprise data integration component based on the Kettle open source ETL project. The platform also includes a big data layer that provides integration across Hadoop, NoSQL, and analytic database systems and enables data blending across these and traditional data stores. It also ensures portability across Hadoop distributions from Cloudera, Hortonworks, MapR, Intel, Cassandra, MongoDB, and Splunk.

In addition to traditional reporting and dashboards, the company's business analytics tools include visualization tools, data mining, and predictive algorithms along with an analytic modeling workbench. Its predictive analytics include visualizations and integration with R and Weka today, and Pentaho labs are actively developing new predictive approaches for the future.

SAP

Software giant SAP is actively ramping up its predictive analytics and big data predictive analytics capabilities. It provides traditional predictive analytics capabilities to its users through its BusinessObjects BI platform. More recently, it purchased predictive analytics vendor KXEN, which provides marketers and other business users with the capabilities to build and deploy models that may change.

Along with a comprehensive set of solutions for BI, SAP also offers HANA, its in-memory database appliance and real-time analytics solution. SAP Predictive Analytics users can query and explore Hadoop data directly.

Tableau Software

At Tableau, founded in 2003, the goal is to help people understand their data. The company aims to help people of all skill levels to derive insight and tell stories from data. Tableau Desktop and Server products offer drag-and-drop functionality that does visual query and analysis. The new visualization engine in Tableau 8 provides support for a wide array of visualizations, including tree maps, bubble clouds, and forecasting extrapolations.

In addition to visual discovery for discerning patterns for prediction, Tableau also offers a forecasting capability that can project data values based on historical data using both additive and multiplicative models.

Recommendations

Recognize that building trust takes time, but don't wait forever. Building out a predictive analytics effort takes time. Cultivate relationships and build trust because sometimes professionals are suspicious of new forms of analysis. This is a question of culture, and it's important to involve the business and other parts of the organization. Collaboration between business users and IT as well as other groups is key. However, do not wait too long. Spinning your wheels experimenting with predictive analytics will only take you so far. Start proof-of-concept projects (with a business sponsor) that show value to get the ball rolling.

Work in steps. If you try to do too many big things at once, you may not succeed. For instance, business users might get frustrated because they are not part of the change. Everyone needs to be involved. Take it one step at a time.

Think about the skills you need. Clearly, the democratization of analytics is moving forward. However, you must think about the skills you will need to build your models and manage your data. With statisticians and other quantitative experts in short supply, list the skills you will need for the kinds of models you want to build. Part of this process is determining the cost/benefit of the models you want to build and deploy. Consider training the staff you have today rather than hiring a "superhero." Allocate your resources wisely.

Cultural issues are just as important as technical issues for predictive analytics.

Organize to execute. It is the rare company that can assemble a group of rock-star statisticians to build and deploy predictive models. Even where that is possible, predictive analytics is not simply about building a model. Remember, different people across your organization will have to get involved, especially if you plan to operationalize the model.

Enterprises are organizing around predictive analytics in different ways. Some are building out dedicated analysis teams. Others are building cross-functional centers of excellence and may have teams within the centers that serve different business areas. Information and best practices are shared across the entire team.

Manage your models. It's easy to start building predictive models when you don't have many of them. However, they can pile up quickly and get stale. Make sure you have a good model management approach for predictive analytics before you start your projects.

Think about utilizing different kinds of data. Although structured data and demographic data are the mainstay of predictive modelers, disparate data types can enrich a data set and provide lift to predictive models. Think about incorporating data beyond the traditional data that you currently have in your data warehouse or on your servers.

The data warehouse still has lots of value. The data warehouse is not going away anytime soon, especially not for predictive analytics. As companies grow in their sophistication and in the kinds and amounts of data they collect, it is best to think about data using an ecosystem approach where your data warehouse is one component of the ecosystem—no matter which groups build the predictive models or which group uses predictive analytics tools.

Leverage what you already have in place, if possible.

Consider newer technology in the mix. However, in addition to the data warehouse, you may need other infrastructure components, especially as volumes increase and disparate and real-time data sources are added. These would be part of the analytics ecosystem and might include analytics platforms, appliances, in-database analytics, in-memory analytics, NoSQL, and Hadoop.

Data integration technologies are key. Data integration is about more than ETL. With the move to larger and more diverse data sets and sources, companies need to start thinking about data integration architectures to support the unification of data.

A solid BI program can ease the way to predictive analytics.

A solid BI infrastructure helps make predictive analytics easier. You must walk before you can fly; the same holds true for predictive analytics. As our research indicates, most of the companies either using predictive analytics or investigating it have a solid BI infrastructure in place. They grow from there.

Governance is always needed. As analytics efforts grow out across the organization, governance will become more important. Some companies put this off, stating that they don't have time for it now. However, issues such as data provenance, ownership, metadata, and the like can come back to haunt you.

Actuate
www.actuate.com

Actuate provides software to more than three million BIRT developers and OEMs who build scalable, secure solutions that save time and improve brand experience by delivering personalized analytics and insights to over 200 million of their customers, partners, and employees. Actuate founded and supports BIRT—the open source IDE—and develops BIRT iHub—the world-class deployment platform—to significantly improve productivity of developers working on customer-facing applications. Actuate's BIRT Analytics™ delivers self-service predictive analytics to enhance customer engagement using big data. Actuate for CCM empowers ECM architects to easily transform, process, personalize, and archive high-volume content. Actuate is headquartered in Silicon Valley with more than 5,000 enterprise customers in financial services, technology, healthcare, and government. Visit actuate.com and developer.actuate.com.

Alteryx
www.alteryx.com

Alteryx provides indispensable analytic solutions for enterprise companies making critical decisions about how to expand and grow. Our product, Alteryx Strategic Analytics, is a desktop-to-cloud agile BI and analytics solution designed for data artisans and business leaders that brings together the market knowledge, location insight, and business intelligence today's organizations require. For more than a decade, Alteryx has enabled strategic planning executives to identify and seize market opportunities, outsmart their competitors, and drive more revenue. Customers like Experian Marketing Services and McDonald's rely on Alteryx daily for their most important decisions. Headquartered in Irvine, California, and with offices in Boulder and Silicon Valley, Alteryx empowers 250+ customers and 200,000+ users worldwide. Get inspired today at www.alteryx.com or call 888.836.4274.

Pentaho
pentaho.com

Pentaho is building the future of business analytics. Pentaho's open source heritage drives our continued innovation in a modern, integrated, embeddable platform built for accessing all data sources. With support for all of the leading Hadoop distributions, NoSQL databases, and high-performance analytic databases, Pentaho provides the broadest support for big data analytics, as well as integration and orchestration of big data and traditional sources. For more information, visit pentaho.com or call 866.660.7555.

SAP
www.sap.com

SAP is at the center of today's technology revolution, developing innovations that not only help businesses run like never before, but also improve the lives of people everywhere.

As the market leader in enterprise application software, we help companies of all sizes and industries run better. From back office to boardroom, warehouse to storefront, 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 more than 248,500 customers to operate profitably, adapt continuously, and grow sustainably.

Tableau Software
www.tableausoftware.com

Tableau Software helps people see and understand data. Tableau helps anyone quickly analyze, visualize, and share information. More than 15,000 organizations get rapid results with Tableau in the office and on the go. And tens of thousands of people use Tableau Public to share data in their blogs and websites. See how Tableau can help you by downloading the free trial at www.tableausoftware.com/trial.

TDWI RESEARCH

TDWI Research provides research and advice for business intelligence and data warehousing professionals worldwide. TDWI Research focuses exclusively on BI/DW issues and teams up with industry thought leaders and practitioners to deliver both broad and deep understanding of the business and technical challenges surrounding the deployment and use of business intelligence and data warehousing solutions. TDWI Research offers in-depth research reports, commentary, inquiry services, and topical conferences as well as strategic planning services to user and vendor organizations.

tdwi

555 S Renton Village Place, Ste 700
Renton, WA 98057-3295

T 425.277.9126
F 425.687.2842
E info@tdwi.org

tdwi.org