The Forrester Wave™: Big Data Predictive Analytics Solutions, Q1 2013

by Mike Gualtieri, January 3, 2013 | Updated: January 13, 2013

KEY TAKEAWAYS

Vendors Must Rise To The Challenge Of Big Data Predictive Analytics

With the rise of big data, the predictive analytics market has woken up; firms now understand the opportunity to use big data to increase their knowledge of their business, their competitors, and their customers. Firms can use predictive analytics models to reduce risks, make better decisions, and deliver more personal customer experiences.

The Big Data Market Is Growing As Organizations Scramble To Harness Predictive Analytics

The big data predictive analytics market is growing because more business and technology professionals see these solutions as a way to address opportunities. Big data involves finding patterns in heterogeneous sources of data; business analysts and data scientists can use these to create predictive models to improve business outcomes.

Big Data Handling, Modeling Tools, And Algorithms Are Key Differentiators

Better big data handling, easy-to-learn/use modeling tools, and a wide choice of analysis algorithms for structured and unstructured data dictate which of these 10 vendors will lead this market. We expect the market for big data predictive analytics solutions to be vibrant, highly competitive, and flush with new entrants over the next three years.

The Forrester Wave™: Big Data Predictive Analytics Solutions, Q1 2013

by Mike Gualtieri
with Stephen Powers and Vivian Brown

WHY READ THIS REPORT

Predictive analytics enables firms to reduce risks, make intelligent decisions, and create differentiated, more personal customer experiences. But predictive analytics is hard to do without the right tools and technologies, given the increasing challenge of storing, processing, and accessing the volume, velocity, and variety of big data. In Forrester's 51-criteria evaluation of big data predictive analytics solution vendors, we evaluated 10 solutions from Angoss Software, IBM, KXEN, Oracle, Revolution Analytics, Salford Systems, SAP, SAS, StatSoft, and Tibco Software. This report details our findings about how well each solution fulfills the criteria and where they stand in relation to each other, and it helps application development and delivery professionals select the right big data predictive analytics solution.

Table Of Contents

Notes & Resources

Forrester conducted lab-based evaluations in Q3 2012 of nine big data predictive analytics solutions from the following vendors: Angoss Software, IBM, KXEN, Revolution Analytics, Salford Systems, SAP, SAS, StatSoft, and Tibco Software. In addition, Forrester spoke with at least two customer references provided by each vendor. Forrester evaluated Oracle separately because Oracle chose not to participate in this research.

Related Research Documents

The Forrester Wave™: Customer Analytics Solutions, Q4 2012
October 26, 2012

PREDICTIVE POWERS DRIVE BREAKTHROUGH BUSINESS OUTCOMES

Sir Francis Bacon, an 18th-century founder of the modern scientific method, famously argued that "Knowledge is power."[1] With all due respect to this great scientist, in the 21st century, "knowledge is profit" for those firms that deploy big data predictive analytics solutions to reduce risks, make smart decisions, and create differentiated, more personal customer experiences. The answers are in the data — but only if companies look for them.

Firms have spent many years building enterprise data warehouses (EDWs) and using business intelligence (BI) tools to report on the business. But predictive analytics is different. Advanced statistical, data mining, and machine learning algorithms dig deeper to find patterns that traditional BI tools may not reveal. Many of these techniques are not new, but big data has breathed new life into the possibilities because more data can mean more and better predictive models. Big data is the fuel and predictive analytics is the engine that firms need to discover, deploy, and profit from the knowledge they gain.[2]

Forrester defines big data predictive analytics solutions as:

> *Software and/or hardware solutions that allow firms to discover, evaluate, optimize, and deploy predictive models by analyzing big data sources to improve business performance or mitigate risk.*

A Continuous Process Fuels Big Data Predictive Analytics

Predictive analytics uses algorithms to find patterns in data that might predict similar outcomes in the future. A common example of predictive analytics is to find a model that will predict which customers are likely to churn. For example, telecommunications firms can use customer data such as calls made, minutes used, number of texts sent, average bill amount, and hundreds of other variables to find models that will predict which customers are likely to change mobile carriers. If a carrier can predict the reasons why customers are likely to churn, it can try to take preemptive action to avoid this undesirable outcome. But this isn't a one-time operation; firms must rerun their analysis on new data to make sure the models are still effective and to respond to changes in customer desires and competitors. Many firms analyze data weekly or even continuously.

In order to maximize success with predictive analytics programs, organizations must (see Figure 1):

■ **Set the business goals.** Clearly stated business goals lie at the center of any successful predictive analytics project. For example, the goal might be to recommend items to upsell to existing customers — or to prevent life-threatening and costly hospital re-admittance. Businesses can also use predictive analytics to achieve more generic business goals, such as increasing revenue, because it enables them to discover correlations that may suggest strategy improvements.

- **Understand data from a variety of sources.** In large organizations, potentially valuable data often exists in multiple silos. In addition, many firms are now using external data from social media, government data, and other public sources of data to augment their internal data. Advanced data visualization tools can help data analysts explore the data from various sources to determine what might be relevant for a predictive analytics project. Increasingly, many data analysts collect every shred of data available to let the predictive analysis algorithms find what is most relevant.

- **Prepare the data.** Data preparation for predictive analysis is a key challenge.[3] Raw data is often unsuitable for predictive analytics. Data analysts must often perform extensive preprocessing of the data before running analysis algorithms. For example, data analysts might need to enrich the data with calculated aggregate fields, strip out extraneous characters or information that would choke the algorithms, or combine data from multiple sources.

- **Create the predictive model.** Data analysts use predictive analytics modeling tools to run analysis algorithms against the data. There are hundreds of different statistical and machine learning algorithms and combinations that data analysts can run against the data to find predictive models. Data analysts typically run the analysis on a subset of the data called "training data" and set aside "test data" that they will use to evaluate the model. For example, data analysts may run the algorithms on a training data set that is 70% of the entire data set; they will then use the remaining 30% as the test data set to evaluate the predictive model.

- **Evaluate the model.** Predictive analytics is not about absolutes; it is about probabilities. To evaluate the predictive power of the model, data analysts run it against the test data set. If the predictive model is more effective than a random selection of the outcome, then they've found an effective predictive model. Data analysts can continue to run other types of algorithms until they find the one that is most predictive; alternatively, they may not find any because there is not enough data or the data is too random to uncover a predictive model for the desired business outcome.

- **Deploy the model.** Analysts must deploy effective predictive models in production applications to accrue the business benefits. A deployed model consists of logic to run the predictive rules and/or formulas and a method to get the data that the model needs and return the result.

- **Monitor the effectiveness of the model.** As financial firms caution, "Past results do not guarantee future performance." It is essential to monitor the effectiveness of the predictive model. For example, if mobile carrier A starts to offer a free data plan, then the reasons why customers churn from carrier B can change. Firms must continue the predictive analytics process to stay on top of business goals, understand new data, prepare better data, refine models with new algorithms, evaluate the models, and deploy and monitor the models in a never-ending cycle.

Figure 1 The Predictive Analytics Process Must Be Continuous To Ensure Effectiveness

85601

Source: Forrester Research, Inc.

Market Overview: General Purpose Predictive Analytics Solutions

The vendors evaluated in this Forrester Wave provide general purpose big data predictive analytics solutions to facilitate the predictive analytics process and ease the burden of this never-ending, continuous cycle of model discovery, deployment, and optimization that can be applied to most industries and business domains. In addition to the general purpose solutions evaluated in this Forrester Wave, firms that wish to benefit from big data predictive analytics solutions can also choose among:

- **Vertical or horizontal solutions.** Many vendors provide solutions that focus on specific industry or horizontal domains, such as customer analytics. For example, Forrester has evaluated solutions offered by Fair Isaac (FICO) and Pitney Bowes that specifically focus on customer-focused programs and initiatives that drive acquisition, retention, cross-sell/upsell, and targeted marketing campaigns.[4] Other examples of vertical solutions include cloud-based offerings such as BloomReach, which uses predictive analytics to help eCommerce companies sell more online by showing customers more relevant content, and startup company Objective Logistics, which uses big data predictive analytics to help restaurants increase sales by improving workforce planning.

- **Embedded solutions.** Other platforms increasingly embed predictive analytics capabilities. BI platforms such as Alteryx and Pentaho include embedded predictive analytics features in addition to BI functionality. Business process management (BPM) platforms such as Pegasystems and Rage Frameworks also offer predictive analytics capabilities.

- **Database analytics.** Relational database management systems (RDBMS), EDWs, NoSQL, Hadoop, and other data-focused hardware and software have some predictive analytics capabilities, but they tend to be oriented toward technical users and require programming or SQL. For example, Teradata's Aster provides a big data predictive analytics capability that allows developers to use SQL and MapReduce together to perform sophisticated analysis on very large data sets. Similarly, programmers can use open source machine learning libraries such as Apache Mahout for Hadoop or a Java library such as Weka for predictive analytics.

- **Offerings from consulting firms.** Enterprises that lack expertise in predictive analytics or that wish to outsource can choose from among many mainstream or boutique consulting firms that focus on predictive analytics. For example, Opera Solutions is a consulting firm that creates predictive models that focus on specific business outcomes. These firms will often use general purpose solutions such as those evaluated in this Forrester Wave, but they also provide deep knowledge and expertise in analyzing data and creating predictive models.

BIG DATA PREDICTIVE ANALYTICS EVALUATION OVERVIEW

To assess the state of the big data predictive analytics market and see how the vendors stack up against each other, Forrester evaluated the strengths and weaknesses of top big data predictive analytics solutions vendors. Forrester expects the market for big data predictive analytics solutions to be vibrant, highly competitive, and flush with new entrants over the next three years.

Evaluation Criteria: Current Offering, Strategy, And Market Presence

After examining past research, user need assessments, and vendor and expert interviews, we developed a comprehensive set of evaluation criteria. We evaluated vendors against 51 criteria, which we grouped into three high-level buckets:

- **Current offering.** We evaluated each solution's architecture; data handling capabilities; discovery and modeling tools; algorithms; model deployment options; life-cycle tools; integration; extensibility; support for standards; and other features.

- **Strategy.** We reviewed each vendor's strategy to assess how each vendor plans to meet current and emerging customer demands. Core evaluation criteria included licensing and pricing; the resources dedicated to their solution; R&D spending; the ability to execute their strategy; and their solution road map.

■ **Market presence.** To determine each vendor's market presence, we evaluated their financials; global presence; installed base; as well as partnerships with other software vendors, software-as-a-service (SaaS)/cloud/hosting providers, and professional services firms.

Lab Evaluations: A Close Look At Each Solution

Forrester also conducted a 3-hour interactive lab session with all participating vendors to get a close look at each solution. The lab was the opportunity for the participating vendors to demonstrate the core capabilities and differentiators of their big data predictive analytics solutions. Each lab session included:

■ **A solution and architectural overview.** Each vendor was asked to include an overview of the product's technical architecture, features, and functionality. The purpose of this session was to familiarize Forrester with the solution and demonstrate the solution's key differentiating features. Vendors used a combination of presentations and/or product demonstrations to provide this overview.

■ **A solution approach using a vendor-provided data set.** Vendors were asked to bring their own data set and show how their solution can be used to understand the data set, prepare the data set, create models, evaluate models, and deploy the solution. Forrester advised vendors to choose a business use case that highlighted the differentiating features of their solution.

■ **A solution approach using the Million Song Dataset.** In addition to demonstrating their solution using their own data set, Forrester also asked vendors to demonstrate how their solution can be used to create a recommendation engine to predict song choices by analyzing the Million Song Dataset.[5] Forrester's objective in specifying the Million Song Dataset was not to evaluate or compare the accuracy of solutions. Rather, Forrester wished to understand how each vendor would approach finding a solution using their respective tools.

Evaluated Vendors: Cross-Domain Focus, Market Presence, And Client Interest

Forrester included 10 vendors in the assessment: Angoss Software, IBM, KXEN, Oracle, Revolution Analytics, Salford Systems, SAP, SAS, StatSoft, and Tibco Software. While Oracle chose not to provide full information for its big data predictive analytics solution, we included it in the Forrester Wave based on our analysis of publicly available information. Each of these vendors has (see Figure 2):

■ **A comprehensive core predictive analytics functionality.** We included vendors that offer one or more solutions that were generally available by May 20, 2012 and that provide at least the following core predictive analytics functional components, tools, and features: 1) connect, extract, transform, cleanse, load, and otherwise prepare analytical data sets; 2) develop and evaluate predictive models using both statistical and machine learning algorithms; 3)

deploy predictive models; 4) manage the predictive modeling life cycle; and 5) tools that data scientists, business analysts, and application developers can use to manage the predictive analytics life cycle.

- **An original, cross-domain predictive analytics solution.** The products included in this evaluation are general purpose predictive analytics solutions that aren't technologically or functionally focused upon particular functional or horizontal applications — such as enterprise resource planning (ERP); customer analytics; customer relationship management (CRM); business intelligence (BI); data warehousing (DW); extract, transform, load (ETL); or the middleware stack. The vendor must develop, market, sell, and implement the solution as a self-sufficient, general purpose big data predictive analytics offering that can stand alone, meaning that it does not need to be embedded in other applications.

- **Significant market presence with a referenceable customer base.** We included vendors that were likely to report or estimated to have at least US$5 million in big data predictive analytics-specific revenue in the latest fiscal or calendar year; at least 80% of those revenues must come from predictive analytics solutions (software licenses, maintenance contracts, and/or subscription offerings), exclusive of consulting or other professional services. The vendors must have 50 in-production big data predictive analytics solutions customers that span more than one major geographical region and that represent five or more industry verticals.

- **Sparked client inquiries and/or has technologies that put the vendor on Forrester's radar.** Forrester clients often discuss the vendors and products through inquiries; alternatively, the vendor may, in Forrester's judgment, warrant inclusion in this evaluation because of technology trends or market presence.

Figure 2 Evaluated Vendors: Product Information And Selection Criteria

	KnowledgeStudio	8.0	May 15, 2012
	KnowledgeCloud		February 14, 2012
IBM	IBM InfoSphere		
	IBM InfoSphere Streams	2.0.0.4	March 29, 2012
	IBM Netezza		
	IBM Netezza 1000 "Twinfin" (as of 10/9/2012, known as IBM PureData System for Analytics N1001) with NetezzaPlatform Software, NPS v6.0.3	6.0.3	June 2011
	IBM Netezza Analytics with Netezza Platform Software v6.0.5	2.0	April 2012
	IBM SPSS Predictive Analytics		
	IBM SPSS Data Collection	6.0.1	September 2011
	IBM SPSS Statistics	20	August 2011
	IBM SPSS Modeler	14.2	June 2011
	IBM SPSS Collaboration and Deployment Services	4.2.1	June 2011
	IBM SPSS Decision Management	6.2	June 2011
	IBM Cognos Consumer Insight	1.1.0.1	December 1, 2011
KXEN	InfiniteInsight	6	March 8, 2012
Revolution Analytics	Revolution R Enterprise	5.0	November 15, 2011

Source: Forrester Research, Inc.

Figure 2 Evaluated Vendors: Product Information And Selection Criteria (Cont.)

Vendor	Product evaluated	Product version evaluated	Version release date
Salford Systems	The SPM Salford Predictive Modeler software suite	6.8	September 6, 2011
	CART Classification and Regression Trees	6.8	September 6, 2011
	MARS Automated Non-linear Regression	6.8	September 6, 2011
	Tree Net/MART Stochastic Gradient Boosting	6.8	September 6, 2011
	Logit	6.8	September 6, 2011
SAP	SAP HANA	1.0	June 2011
	SAP Sybase IQ	15.4	November 2011
	SAP Predictive Analysis	1.0 SP0	April 2012
	SAP Data Services	4.1	May 2012
	SAP BusinessObjects BI	4.0 FP3	May 2012
SAS	SAS Analytics Suite		
	JMP	10	March 2012
	JMP Pro	9	October 2010
	SAS Social Media Analytics	5.4	June 2011
	SAS Enterprise Miner	7.1	July 2011
	SAS Enterprise Miner for Desktop	7.1	July 2011
	SAS Forecast Server	4.1	July 2011
	SAS Model Manager	3.1	July 2011
	SAS Text Miner	5.1	July 2011
	SAS Operations Research (SAS/OR)	9.3	July 2011
	SAS Rapid Predictive Modeler	1.3.1	July 2011
	SAS Analytics Pro	9.3	July 2011
	SAS/IML Studio	9.3	July 2011
	SAS/ETS	9.3	July 2011

Source: Forrester Research, Inc.

Figure 2 Evaluated Vendors: Product Information And Selection Criteria (Cont.)

Vendor	Product evaluated	Product version evaluated	Version release date
SAS	SAS/STAT	9.3	July 2011
	SAS Credit Scoring for Enterprise Miner	7.1	July 2011
	SAS Enterprise BI Server	4.3.1	July 2011
	SAS Analytics Accelerator for Teradata	2.1	July 2011
	SAS/ACCESS	9.3	July 2011
	SAS DataFlux Data Management Platform	2.2	November 2011
	SAS Scoring Accelerator	4.1	December 2011
	SAS Data Integration Studio	4.3	January 2012
	SAS Enterprise Guide	5.1	February 2012
	SAS Add-in for Microsoft Excel	5.1	February 2012
	SAS Visual Analytics	5.1	March 2012
	SAS Social Network Analysis	2.3	April 2012
	SAS High-Performance Analytics Server	12.1	April 2012
StatSoft	Statistica	11.0A	April 13, 2012
Tibco Software	Tibco Spotfire S+	8.2.0	November 1, 2010
	Tibco Spotfire Miner	8.2.0	November 1, 2010
	Tibco Spotfire Statistics Services	4.5.0	May 1, 2012
	Tibco Spotfire Analytics	4.5.0	May 1, 2012

Source: Forrester Research, Inc.

## BREADTH AND DEPTH DISTINGUISH THE LEADERS

The Leaders in this Forrester Wave offer a rich set of algorithms to analyze data, architectures that can handle big data, and tools for data analysts that span the full predictive analytics life cycle. SAS and IBM resoundingly provide leading solutions, but each of the 10 vendors we evaluated has unique points of differentiation that many customers will find attractive. Forrester's evaluation of general purpose big data predictive analytics solutions reveals three Leaders, five Strong Performers, and two Contenders (see Figure 3):

- **SAS and IBM are unshakeable Leaders, while newcomer SAP performs well.** SAS, with its 36-year history of providing analytics software, is a Leader in this evaluation because it scored well in all categories. Its SAS Enterprise Miner tool is easy to learn and can run analysis in-database or on distributed clusters to handle big data. IBM's Smarter Planet campaign and acquisitions of SPSS, Netezza, and Vivisimo represent its commitment to big data predictive analytics. IBM's complementary solutions, such as InfoSphere Streams and Decision Management, strengthen the appeal for firms that wish to integrate predictive analytics throughout their organization. SAP is a newcomer to big data predictive analytics but is a Leader due to a strong architecture and strategy. SAP also differentiates by putting its SAP HANA in-memory appliance at the center of its offering, including an in-database predictive analytics library (PAL), and offering a modeling tool that looks a lot like SAS Enterprise Miner and IBM SPSS Modeler.

- **Tibco, Oracle, StatSoft, and KXEN are Strong Performers with unique approaches.** In general, the Strong Performers had lower architecture scores than the Leaders. Tibco's Spotfire advanced data visualization tool offers core support for S+ and R, which makes it attractive to data scientists who know those languages.[6] Oracle's solution centers on offering in-database R and the strength of its in-database analytics technology. StatSoft has a comprehensive number of analysis algorithms and is very strong in manufacturing use cases. KXEN collapses the normal predictive analytics life cycle by automating the predictive model discovery process; it also offers strong social network analysis.

- **Angoss, Revolution Analytics, and Salford Systems are Contenders with a narrower focus.** Smaller vendors Angoss Software, Revolution Analytics, and Salford Systems have a narrower focus than others in this evaluation, but firms have valid reasons to consider them. Angoss offers the best tooling for decision trees that we have seen and offers cloud solutions that firms can use to improve results quickly. Revolution Analytics aims to be the de facto commercial provider of solutions based on the very popular open source statistics programming language R; other vendors in this evaluation offer or plan to offer R-based solutions. Salford Systems claims superior implementations of analysis algorithms including CART, MARS, TreeNet, and random forests; it has made a strong name for itself in particular among data scientists who have an interest in the algorithms that Salford supports.

Figure 3 Forrester Wave™: Big Data Predictive Analytics Solutions, Q1 '13
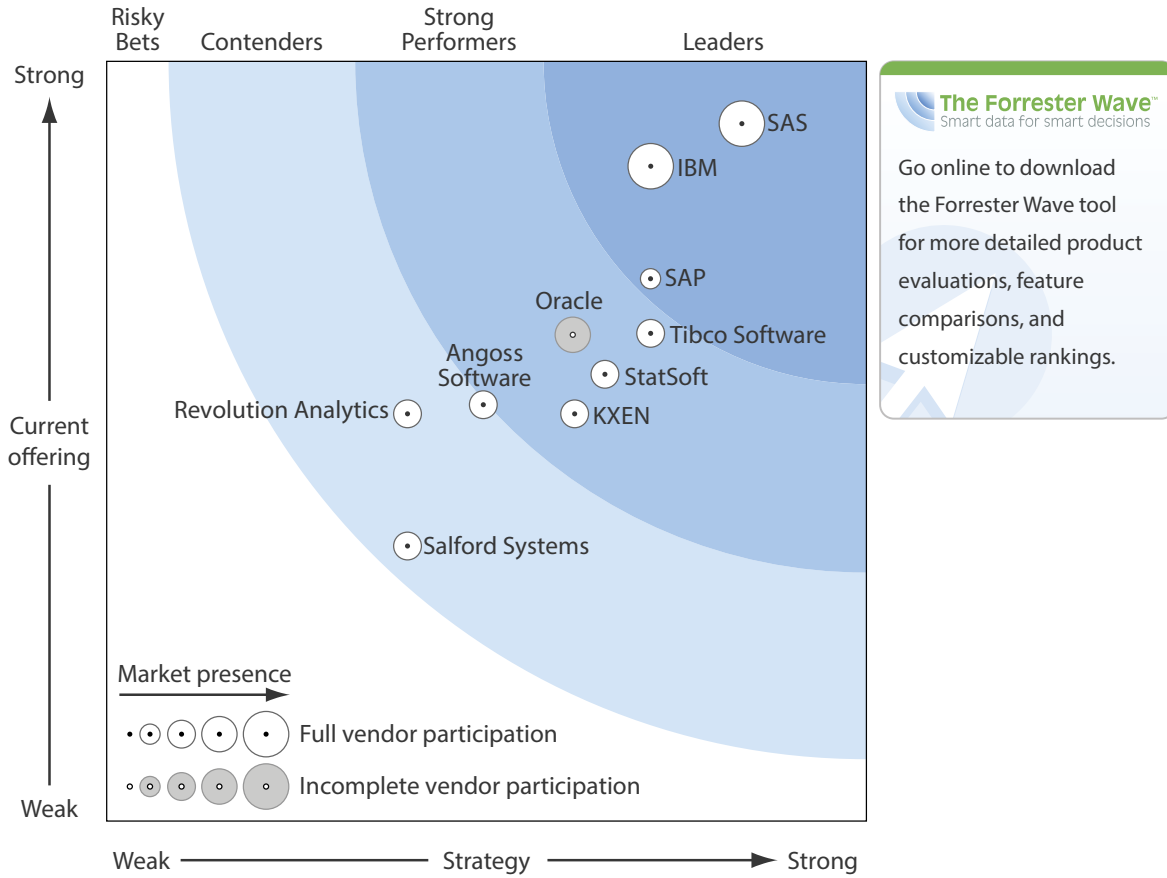
Source: Forrester Research, Inc.

Figure 3 Forrester Wave™: Big Data Predictive Analytics Solutions, Q1 '13 (Cont.)

	Forrester's Weighting	Angoss Software	IBM	KXEN	Revolution Analytics	Salford Systems	SAP	SAS	StatSoft	Tibco Software
CURRENT OFFERING	50%	2.74	4.31	2.68	2.68	1.81	3.57	4.59	2.94	3.21
Architecture	20%	2.82	5.00	2.50	3.07	1.10	4.96	5.00	1.30	2.10
Data	15%	3.00	4.00	4.00	2.50	2.50	4.50	4.50	4.00	3.00
Discovery	25%	1.70	3.85	2.90	2.85	1.00	3.05	4.40	2.85	3.10
Evaluation and optimization	5%	2.00	5.00	2.20	2.00	1.60	1.20	4.20	3.80	4.60
Deployment	5%	4.60	5.00	4.60	3.40	3.40	2.60	5.00	4.60	3.80
Tools	25%	3.60	4.00	1.60	2.50	2.80	3.40	4.40	3.40	3.80
Standards, integration, solutions, and extensibility	5%	1.50	5.00	2.25	1.75	0.25	2.00	5.00	2.00	4.00
STRATEGY	50%	2.48	3.58	3.08	1.98	1.98	3.58	4.18	3.28	3.58
Licensing and pricing	25%	2.50	2.50	2.50	2.50	2.50	2.50	2.50	2.50	2.50
Commitment	50%	2.20	3.40	3.40	2.20	2.20	3.40	4.60	3.80	3.40
Product road map	25%	3.00	5.00	3.00	1.00	1.00	5.00	5.00	3.00	5.00
MARKET PRESENCE	0%	2.09	4.21	2.94	2.20	2.15	1.93	4.46	2.66	2.09
Company financials	30%	1.00	3.50	2.00	3.00	1.00	1.00	4.00	2.00	1.00
Global presence and installed base	60%	2.68	4.60	3.24	1.80	2.68	2.28	4.60	3.24	2.28
Partnerships	10%	1.80	4.00	4.00	2.20	2.40	2.60	5.00	1.20	4.20

All scores are based on a scale of 0 (weak) to 5 (strong).

Source: Forrester Research, Inc.

VENDOR PROFILES

Leaders

■ **SAS proves an analytics powerhouse.** With an unwavering focus on data analytics since 1976, SAS offers a broad set of tools for predictive analytics, an architecture that supports multiple platforms, in-database analytics, in-memory analytics, and significant market presence. SAS has more than 19,500 unique customers in 135 countries. Forrester estimates that well over 3,000 are using predictive analytics. For SAS to stay a Leader, it must continue to support the largest data sets in the world; provide more sophisticated solutions for real-time analytics, such as stream processing; offer predictive modeling tools that business analysts find more usable; and address the many upstarts entering this market now that big data and predictive analytics are hot.

- **IBM's Smarter Planet defines big data predictive analytics.** IBM says of its Smarter Planet, "We want to change the paradigm from react to anticipate."[7] Forrester estimates that IBM has well over 1,500 predictive analytics customers in production in the Americas, Europe, Middle East, and Africa (EMEA), and Asia Pacific. IBM has a full range of assets to help customers find predictive models in their big data, including SPSS analysis software and Netezza data warehouse appliance. IBM needs to do a better job of reconciling its go-to-market strategy for its broad portfolio to make it less confusing. In addition, IBM needs to continue its efforts in creating predictive analytics solutions — IBM calls these blueprints — that customers can use out of the box.

- **SAP's strategy hinges on adoption of its HANA in-memory database appliance.** SAP's strategy of centering its big data predictive analytics solution on its SAP HANA in-memory database appliance is bold. To be successful with this strategy, SAP must continue to build out cloud-based HANA options to make it easy to adopt for customers who don't want to install new hardware on-premises. SAP must also enhance its modeling and analytics life-cycle tools to compete with SAS and IBM. Forrester estimates that SAP has fewer than 100 predictive analytics customers in production.

Strong Performers

- **Tibco Software's visualization tools combine with R to differentiate it.** Tibco was either prescient or lucky with its acquisition of S+ technology from Insightful in 2008. Because S+ is the precursor to R, Tibco has a headstart in making R its core analytics engine. Combined with Tibco's leading advanced data visualization tool, Tibco will be well-positioned to be a strong choice for enterprises that wish to have an R-based predictive analytics capability. However, Tibco must also work quickly to overcome the memory-bound data processing limitations of R. Forrester estimates that Tibco has fewer than 500 predictive analytics customers in production, with more than 25% installed outside North America.

- **Oracle can do better than R.** Oracle's solution centers on in-database analytics capabilities and prepackaged solutions within its widely adopted enterprise RDBMS, applications, and Exadata appliance rather than predictive analytics tools that are database-agnostic. Oracle's recent strategy is to incorporate R into its database and data warehouse offerings. R is a powerful statistical programming language that is difficult to learn, so Oracle must work to provide tooling that hides that complexity. Oracle needs to fill in modeling and life-cycle tools gaps to advance the product.

- **StatSoft goes deep on algorithms but must invest to get to the next level.** StatSoft has a comprehensive library of analysis algorithms and modeling tools but lags behind in providing big data processing capabilities such as in-database analytics and distributed processing architecture. StatSoft is well-known in data scientist circles — and has lots of potential if it

invests in marketing its solution to business executives in addition to data scientists. StatSoft must also invest in a redesign of its tools to make them simpler for business analysts to use. Forrester estimates that StatSoft has less than 1,500 predictive analytics customers in production in the Americas, EMEA, and Asia Pacific.

- **KXEN's low-touch approach to predictive analytics will boom in popularity.** KXEN focuses on providing new clients with better business outcomes as fast as possible. To achieve this, KXEN offers an automated predictive model discovery process using proprietary structured algorithms. It is also one of two vendors in this evaluation that offer a social network analysis algorithm. Forrester expects KXEN's low-touch approach to become more popular given that many firms wish to quickly deploy models but do not have data scientists to use the more traditional tools. However, many firms will also need an additional solution that allows data analysts to have more control over the predictive analytics process. KXEN must either sharpen its differentiation or invest in the capability to handle nonproprietary algorithms; if not, it will continue to be a niche player. KXEN reports 577 predictive analytics customers in the Americas, EMEA, and Asia Pacific.

Contenders

- **Angoss Software has the most impressive user tools for decision trees that we have seen.** Although Angoss supports several types of algorithms, it really shines in providing a highly functional yet simple-to-use interface for creating, exploring, and modifying decision trees. Angoss also offers cloud-based customer analytics solutions that are especially attractive to financial services firms. But great decision trees and customer analytics solutions are not enough for Angoss to really shine in this space. Angoss must decide whether to focus on customer analytics or invest heavily to keep up with the Leaders in this evaluation. Angoss has 266 predictive analytics customers in production deployed in the Americas, EMEA, and Asia Pacific.

- **Revolution Analytics wants to become R's enterprise choice.** Revolution Analytics focuses on providing enterprise features for R, the popular programming language and environment for statistical computing and graphics. R is the open source darling among data scientists, but performance on large data sets is an architectural barrier that Revolution Analytics hopes to overcome.[8] Revolution is investing to re-implement many R functions to take advantage of advanced CPUs and distributed computing such as Hadoop. This could help Revolution Analytics become the de facto standard for enterprise R tools. However, performance alone may not be enough for Revolution Analytics to differentiate its R solution. It must also invest in much better tools to catch up to vendors that have enterprise features. Revolution Analytics is a strong choice for firms that want an R-based solution that is supported by a commercial firm and provides some enterprise features. It will also face competition from other vendors in this evaluation that provide technology to integrate with R. Revolution Analytics reported 175 customers with production deployments in North America, EMEA, and Asia Pacific.

■ **Salford Systems will remain a niche player.** Salford Systems is well known among the data scientist community and especially in academic circles. Its strategy has been to establish exclusive relationships with academic data mining rock stars, such as Jerome Friedman of Stanford University, who implement algorithms used in its SPM Salford Predictive Modeler software suite. Salford focuses on three core decision tree algorithms — CART, TreeNet, and random forests — as well as one regression algorithm — MARS. Other software solutions also offer these algorithms, but Salford's references claim that Salford has the best implementations. Salford is very active in data mining competitions and is a well-known vendor among data scientists who focus on tree algorithms. Unless Salford broadens its focus in terms of architecture, support for big data, and life-cycle tools, it will remain a niche player. Salford Systems reported 353 customers with production deployments in the Americas, EMEA, and Asia Pacific.

SUPPLEMENTAL MATERIAL

Online Resource

The online version of Figure 3 is an Excel-based vendor comparison tool that provides detailed product evaluations and customizable rankings.

Data Sources Used In This Forrester Wave

Forrester used a combination of data sources to assess the strengths and weaknesses of each solution:

■ **Hands-on lab evaluations.** Vendors spent 3 hours with a team of analysts who performed a hands-on evaluation of the product using a scenario-based testing methodology. We evaluated each product using the same scenarios, thus creating a level playing field by evaluating every product on the same criteria.

■ **Vendor surveys.** Forrester surveyed vendors on their capabilities as they relate to the evaluation criteria. Once we analyzed the completed vendor surveys, we conducted vendor calls where necessary to gather details of vendor qualifications.

■ **Customer reference calls.** To validate product and vendor qualifications, Forrester also conducted reference calls with at least two of each vendor's current customers.

The Forrester Wave Methodology

We conduct primary research to develop a list of vendors that meet our criteria to be evaluated in this market. From that initial pool of vendors, we then narrow our final list. We choose these

vendors based on: 1) product fit; 2) customer success; and 3) Forrester client demand. We eliminate vendors that have limited customer references and products that don't fit the scope of our evaluation.

After examining past research, user need assessments, and vendor and expert interviews, we develop the initial evaluation criteria. To evaluate the vendors and their products against our set of criteria, we gather details of product qualifications through a combination of lab evaluations, questionnaires, demos, and/or discussions with client references. We send evaluations to the vendors for their review, and we adjust the evaluations to provide the most accurate view of vendor offerings and strategies.

We set default weightings to reflect our analysis of the needs of large user companies — and/or other scenarios as outlined in the Forrester Wave document — and then score the vendors based on a clearly defined scale. These default weightings are intended only as a starting point, and we encourage readers to adapt the weightings to fit their individual needs through the Excel-based tool. The final scores generate the graphical depiction of the market based on current offering, strategy, and market presence. Forrester intends to update vendor evaluations regularly as product capabilities and vendor strategies evolve.

ENDNOTES

[1] Sir Francis Bacon was an English philosopher and scientist. He was an advocate of the scientific method during the scientific revolution.

[2] Forrester defines big data as *the frontier of a firm's ability to store, process, and access (SPA) all the data it needs to operate effectively, make decisions, reduce risks, and serve customers. Source: Mike Gualtieri, "The Pragmatic Definition Of Big Data," Mike Gualtieri's Blog, December 5, 2012 (http://blogs.forrester.com/mike_ gualtieri/12-12-05-the_pragmatic_definition_of_big_data).*

[3] For more information on cleaning dirty data, check out "Overcoming Data Mining Challenges: Dirty Data," Rexer Analytics (http://www.rexeranalytics.com/Overcoming_Challenges_Dirty_Data.html).

[4] In Forrester's 70-criteria evaluation of customer analytics vendors, we identified the six most significant software providers — Angoss Software, FICO, IBM, KXEN, Pitney Bowes, and SAS — in the category and researched, analyzed, and scored them. This evaluation details our findings about how well each vendor fulfills our criteria and where the vendors stand in relation to each other and will help customer intelligence (CI) professionals select the right partner for their customer analytics needs. See the October 26, 2012, "The Forrester Wave™: Customer Analytics Solutions, Q4 2012" report.

[5] Visit http://labrosa.ee.columbia.edu/millionsong/ for more information on the Million Song Dataset.

[6] Enterprises find advanced data visualization (ADV) platforms to be essential tools that enable them to monitor business, find patterns, and take action to avoid threats and snatch opportunities. In Forrester's 29-criteria evaluation of ADV vendors, we found that Tableau Software, IBM, Information Builders, SAS, SAP, Tibco Software, and Oracle led the pack due to the breadth of their ADV business intelligence

(BI) functionality offerings. Microsoft, MicroStrategy, Actuate, QlikTech, Panorama Software, SpagoBI, Jaspersoft, and Pentaho were close on the heels of the Leaders, also offering solid functionality to enable business users to effectively visualize and analyze their enterprise data. See the July 17, 2012, "The Forrester Wave™: Advanced Data Visualization (ADV) Platforms, Q3 2012" report.

[7] For more information on IBM's "Smarter Planet" initiative, check out "What is a Smarter Planet," IBM (http://www.ibm.com/smarterplanet/us/en/overview/ideas/index.html?lnk=ussph1.16).

[8] The KDnuggets Software Poll provides data on the adoption of analytics, data mining, and big data tools. Source: KDnuggets (http://www.kdnuggets.com/polls/2012/analytics-data-mining-big-data-software.html).

About Forrester

A global research and advisory firm, Forrester inspires leaders, informs better decisions, and helps the world's top companies turn the complexity of change into business advantage. Our research-based insight and objective advice enable IT professionals to lead more successfully within IT and extend their impact beyond the traditional IT organization. Tailored to your individual role, our resources allow you to focus on important business issues — margin, speed, growth — first, technology second.

FOR MORE INFORMATION

To find out how Forrester Research can help you be successful every day, please contact the office nearest you, or visit us at www.forrester.com. For a complete list of worldwide locations, visit www.forrester.com/about.

CLIENT SUPPORT

For information on hard-copy or electronic reprints, please contact Client Support at +1 866.367.7378, +1 617.613.5730, or clientsupport@forrester.com. We offer quantity discounts and special pricing for academic and nonprofit institutions.

Forrester Focuses On
Application Development & Delivery Professionals

Responsible for leading development and delivery of applications that support your company's business strategies, you also choose technology and architecture while managing people, skills, practices, and organization to maximize value. Forrester's subject-matter expertise and deep understanding of your role will help you create forward-thinking strategies; weigh opportunity against risk; justify decisions; and optimize your individual, team, and corporate performance.

« **ANDREA DAVIES,** client persona representing Market Insights Professionals