



Top 12 Building Blocks of Tomorrow's Data Science Platforms

FEATURING RESEARCH FROM FORRESTER

The Forrester Wave™: Predictive
Analytics And Machine Learning
Solutions, Q1 2017

Evolution of the Data Science Platform

AN INTRODUCTION TO THE FORRESTER WAVE™: PREDICTIVE ANALYTICS AND MACHINE LEARNING SOLUTIONS, Q1 2017

A Data Science Platform is arguably the most important asset companies acquire as part of their data analytics strategy. In the 2017 Predictive Analytics and Machine Learning Solutions Wave, “Forrester forecasts a 15% compound annual growth rate (CAGR) for the PAML market through 2021.” As market adoption grows, so does the need for Data Science Platforms to adapt to noticeable industry-consistent trends within enterprise IT. Here is a list of key IT trends that are shaping the way Data Science Platforms are designed today.

IN THIS DOCUMENT

- 1 Top 12 Building Blocks of Tomorrow’s Data Science Platforms
- 6 Research From Forrester: The Forrester Wave™: Predictive Analytics And Machine Learning Solutions, Q1 2017

- Adoption of Big Data Technologies
- Use of New Data Sources for Decision Making
- Adoption of Open Standards and Libraries
- Integration with Enterprise Applications & Tools
- IT Centralization
- Security and Governance

“Data scientists who are coders are increasingly using more than one programming language because of open source add-on libraries such as CRAN for R and scikit-learn for Python.”

- Forrester Research

22 About Angoss

Data Science Platforms that align with these trends will equip users with a single, fully-functional platform that is capable of unifying infrastructure, technology, and data science teams. This will simplify data access and enable CIOs and Data Scientists to easily attain data from a multitude of new data sources like social media, Internet of Things (IoT), Hadoop and Amazon S3 buckets, to build actionable strategies. As organizations build more heterogeneous data science teams, Data Science Platforms will need to cater to different skill levels and requirements. These could range from different visual analytics or custom coding environments (R, Python, Language of SAS, Tableau, Qlik), or out-of-the-box integration with open source machine learning packages like Spark ML. Essentially, having a single, integrated platform that is compliant with current data and technology trends will lead to better analytic results.

TOP 12 BUILDING BLOCKS OF TOMORROW'S DATA SCIENCE PLATFORMS

With the growing need for more dynamic Data Science Platforms, adapted to Big Data needs, organizations are seeking analytics tools that satisfy the requirements of the modern day data scientist as well as automate and ease current decision making processes without sacrificing functionality. When evaluating current legacy applications or acquiring an enterprise-wide data science solution, organizations should look for the following essential building blocks that comprise a comprehensive end-to-end Data Science Platform for tomorrow's data needs.

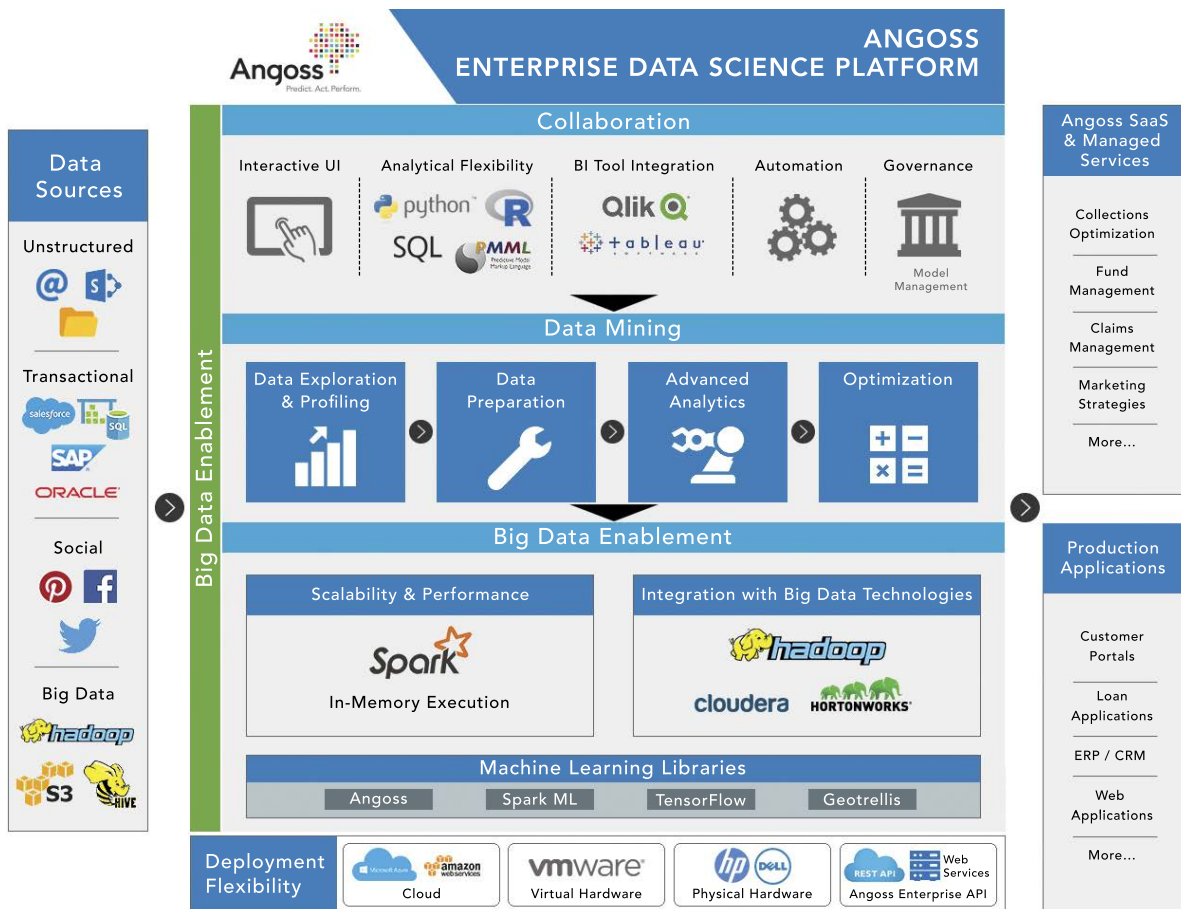
- 1 User-Friendly and Interactive Interface** - Data Science Platforms with a consistent look-and-feel, wizard-driven functionality, and an automated workflow minimize the learning curve, ease navigation, optimize model development, and enable users to easily construct workflows in a fraction of the time.
- 2 Data Access** - Customer information is streamed from a variety of sources. A Data Science Platform must be able to support structured data sources, like Excel, as well as unstructured data sources such as social media feeds, audio, transactional, IoT, and from Big Data environments such as Amazon S3, Hadoop and Teradata. Being able to blend both data types in a single view improves predictive and exploratory power of models.
- 3 Data Preparation** - A Data Science Platform that provides users with the option to code or not to code can significantly increase productivity of data mining projects by reducing the 80% of time spent on data preparation.
- 4 Data Exploration, Profiling, and Visualization** - A wide range of exploratory features such as cross tabulations, segmentation, characteristics analysis, and the segment viewer, help data scientists quickly assess data quality, derive insights, and detect patterns and trends making every stage of the data mining process easily interpretable and presentable. Additionally, platforms that support BI tools like Tableau and Qlik dashboards provide transparency to advanced analytic results and ease communication of data insights with decision makers.
- 5 Comprehensive Advanced Analytical Flexibility** - Modeling techniques such as advanced scorecard development, regression analysis, optimization, and machine-learning methods are a critical component of a platform when building data science solutions for an extensive variety of business applications. Data Science Platforms that extend analytical capabilities beyond their own, such as programming in multiple languages like R, Python, SQL, and PMML, offer immense flexibility and collaboration potential for diverse data science teams.

- 6 **Integration & Extensibility with Open Source Machine Learning Libraries** - Enterprises with Big Data needs require Data Science Platforms integrated with Machine Learning libraries such as Spark ML, Tensorflow, or Geotrellis, for large scale data analytics processes.
- 7 **Governance** - To satisfy security, governance, and compliance regulations, Data Science Platforms must possess a project and model management framework with documentation capabilities, logging, auditing, as well as access to streamlined reporting (historical comparison results, validation reports, and model validation charts and tables) for model comparison and testing.
- 8 **Collaboration** - In order to provide transparency at every step of the modeling process, enable communications across all modeling steps, and facilitate project communications between multiple stakeholders, Data Science Platforms must be equipped with features such as self documentation, model documentation, model sharing and report sharing which create a continuous and simultaneous teamwork across different locations.
- 9 **Integration with Big Data Technologies** - Many companies have either adopted or plan to adopt Big Data technologies. More specifically, they use either Hortonworks or Cloudera as their distributed storage tier for enterprise data. The impact to the current Data Science Platform is that it now needs to access data within HDFS and Hive and provide visual analytics for Hadoop. Most importantly, the Data Science Platform must integrate with Spark for model training as well as scoring.
- 10 **Automation** - Automation is inherent in Data Science Platforms adapted for Big Data Frameworks. Features such as a visual canvas to build automated workflows - data collection, data prep, transformations, scoring, report generation - can ease the process of deployment on Hadoop clusters.
- 11 **Deployment Flexibility** - A platform's ability to efficiently deploy models via scoring, automatic code generation (SAS, SQL, SPSS, PMML and Java), in cloud or on-premises, as well as in a physical or virtual environment, make it functional for Big Data insights dissemination and is crucial when dealing with a magnitude of business scenarios. Additionally, in order to turn data into actionable insights, Data Science Platforms require API's (REST API and Web Services) to integrate with production applications that call for Real Time Analytics.
- 12 **Scalability, Performance, and IT Centralization** - Data Science Platforms adapted for Big Data functionality will be equipped to support large datasets, in-memory execution on Spark, and deployment of multiple projects in parallel. Furthermore, platforms that support a multi-tenant/shared environment will improve IT efficiency and lower costs associated with hardware and storage overhead.

FUTURE OUTLOOK

Adoption of Big Data enabled data science platforms in on the rise across the enterprise in all verticals and industries. Data Science Platforms capable of merging and harnessing the power of advanced analytics and Big Data technologies will play a vital role in overcoming the enterprise challenges of insight discovery.

“Forrester forecasts a 15% compound annual growth rate (CAGR) for the PAML market through 2021.”
- Forrester Research



Angoss' Data Science Platform, a leader in the 2017 Forrester Wave, offers 12 data science platform building blocks to help organizations institutionalize knowledge and promote collaboration across their heterogeneous teams of different skills and tools requirements.

Angoss Data Science Platform performs advanced analytics directly within the Big Data environment, eliminates data duplication, minimizes security breaches, and automates results which lead to insights. With Angoss' Enterprise Solution, businesses will be able to reduce infrastructure costs by leveraging centralized infrastructure for data science across the organization. **Angoss enables companies to truly live up to the Big Data philosophy by being able to continue to accrue vast amounts of data without compromising data access, performance, and analytics.**

The Forrester Wave™: Predictive Analytics And Machine Learning Solutions, Q1 2017

Enterprises Must Possess The Power To Predict In The Age Of The Customer

by Mike Gualtieri

March 7, 2017

Why Read This Report

Make your data scientists more productive — and happier — with a new predictive analytics and machine learning (PAML) solution. In our 23-criteria evaluation of PAML solution vendors, we identified the 14 most significant ones — Alpine Data, Angoss, Dataiku, Domino Data Labs, FICO, H2O.ai, IBM, KNIME, Microsoft, RapidMiner, Salford Systems, SAP, SAS, and Statistica — and researched, analyzed, and scored them. This report shows how each provider measures up and helps AD&D professionals make the best choice for their specific needs.

Key Takeaways

PAML Solutions Make Open Source Better

There is a tremendous amount of innovation in the open source community in terms of algorithms and platforms to build models from big data. Many PAML solutions have embedded open source in their solutions and made it easier for data scientists to leverage them.

Predictive Model Management Is A Key Differentiator

Predictive models can be fickle in that they can lose their predictive power over time. A distinguishing mark of the Leaders is that they have superior features to monitor models in production.

Machine Learning Is Fundamental To Artificial Intelligence

If your enterprise wants to leverage AI, then it must start with a PAML solution. All of the solutions in this Forrester Wave provide machine learning capabilities that AD&D pros can use to create AI applications. We found seven Leaders, five Strong Performers, and two Contenders.

The Forrester Wave™: Predictive Analytics And Machine Learning Solutions, Q1 2017

Enterprises Must Possess The Power To Predict In The Age Of The Customer



by [Mike Gualtieri](#)

with [Srividya Sridharan](#), [Holger Kisker, Ph.D.](#), and [Cristian Austin](#)

March 7, 2017

Table Of Contents

The Power To Predict Is Mighty

Predictive Analytics And Machine Learning Solutions Make It Possible

Hot Markets Spur Innovation And New Entrants

PAML Solutions Are DIY, But There Are Other Ways To Attain The Power To Predict

Predictive Analytics And Machine Learning Evaluation Overview

Evaluated Vendors And Inclusion Criteria

Vendor Profiles

Leaders

Strong Performers

Contenders

Supplemental Material

Related Research Documents

[A Machine Learning Primer For BT Professionals](#)

[Artificial Intelligence: What's Possible For Enterprises In 2017](#)

[Massive Machine-Learning Automation Is The Future Of Data Science](#)

[Prescriptive Analytics: The Black Belt Of Digital Decisions](#)

FORRESTER

Forrester Research, Inc., 60 Acorn Park Drive, Cambridge, MA 02140 USA
+1 617-613-6000 | Fax: +1 617-613-5000 | [forrester.com](#)

FORRESTER

© 2017 Forrester Research, Inc. Opinions reflect judgment at the time and are subject to change. Forrester®, Technographics®, Forrester Wave, TechRadar, and Total Economic Impact are trademarks of Forrester Research, Inc. All rights reserved. Forrester Research, Inc. is not responsible for any content or copyright infringement on this page. For more information, contact [Forrester@forrester.com](#) or +1 866-367-7378.

The Power To Predict Is Mighty

Yogi Berra once said, “It’s tough to make predictions, especially about the future.”¹ It is tough indeed, but enterprises that can make probabilistic predictions about customers, business processes, and operations will have an edge over enterprises that can’t. These predictions don’t have to be macroscopic to be consequential.² Predictions about what a customer is likely to buy next. Predictions about marketing content that will resonate with a prospect. Predictions about the next best action to take in a business process. Predictions about when an expensive asset is likely to break down. Virtually any customer journey, business process, and even strategic decision can be made better if permeated with the power to predict.

PREDICTIVE ANALYTICS AND MACHINE LEARNING SOLUTIONS MAKE IT POSSIBLE

Yes, making accurate predictions is tough, but PAML solutions provide application development and delivery (AD&D) professionals and data scientists with the tools to make it happen. In our third evaluation of this market, we emphasize the use of machine learning to create predictive models.³ Forrester defines PAML solutions as:⁴

Software that provides data scientists with 1) tools to build predictive models using statistical and machine learning algorithms and 2) a platform to deploy and manage predictive production models.

HOT MARKETS SPUR INNOVATION AND NEW ENTRANTS

Forrester forecasts a 15% compound annual growth rate (CAGR) for the PAML market through 2021.⁵ That’s a conservative estimate given that the PAML category includes and overlaps with AI and deep learning. The category continues to be hot since most large enterprises want the power to predict and have only scratched the surface of what is possible. PAML solutions today should:

- › **Drive data scientist productivity.** Data scientists are in demand, and it can be hard to find good ones. Instead of hiring three more data scientists, perhaps an enterprise could dramatically improve the productivity of existing data scientist teams. Many of the PAML solution vendors focus on speeding up analysis by using big data platforms such as Apache Spark, automating portions of the data science life cycle, and improving usability of the data science workbench.⁶
- › **Include multiple model deployment methods.** Production models shouldn’t sit statuesque on their own. They must be embedded in applications and business processes to provide business value. Enterprises must be able to deploy models in multiple ways, including as code embedded directly into applications, exposed as a service callable by applications, or injected into other platforms such as databases.⁷ Some of the more mature PAML vendors include or are integrated with decision management platforms that allow AD&D pros and business users to use a visual metaphor or express decision logic as a set of business rules that can also include models.⁸
- › **Provide sophisticated model management.** The very nature of predictive models is that they

may lose accuracy overtime. More mature PAML solutions include features to monitor the ongoing efficacy of models in production by comparing model output with established key performance indicators and testing new models using a champion/challenger or A/B testing scheme.

- › **Allow polyglot programming.** Data scientists who are coders are increasingly using more than one programming language because of open source add-on libraries such as CRAN for R and scikit-learn for Python. Enterprise data scientists who still use the SAS programming language may also use R, Python, Scala, Julia, and others because of this.
- › **Expand to Apache Spark.** Apache Spark is an open source, primarily in-memory cluster computing platform that also includes Spark ML, a set of machine learning libraries that data scientists are increasingly interested in using.⁹ In addition to Spark ML, other machine libraries such as H2O.ai's Sparkling Water and IBM's SystemML run on Spark. Most PAML vendors have moved from a Hadoop strategy for analyzing big data to an Apache Spark strategy because of the machine learning libraries and speed of in-memory processing.
- › **Build the foundation for AI and invest in deep learning.** Machine learning models are a key building block of AI applications.¹⁰ Data scientists can use any of the PAML solutions we evaluated in this Forrester Wave to build models for use in AI applications. Deep learning is a branch of machine learning that data scientists use to build models based on artificial neural networks.¹¹ This method is particularly good at creating models for image recognition (including facial recognition), but it is applicable to more traditional use cases as well. Vendors are incorporating numerous open source libraries, such as Caffe, MXNet, and TensorFlow, into PAML solutions, or they are creating their own deep-learning algorithms built into the platform.
- › **Accommodate citizen data scientists.** Many enterprises like the idea that non-data scientists in their organizations can create models without having in-depth data science knowledge. PAML vendors have responded by creating wizard-like tools in their PAML solutions to make it easy for these citizen data scientists to create simple models. While these tools may have some benefits for nonproduction models used for exploratory business intelligence (BI), an enterprise should not think that this will replace real data scientists, because there are too many complexities of model building, such as feature creation, model evaluation, overfitting, and creating ensembles.

PAML SOLUTIONS ARE DIY, BUT THERE ARE OTHER WAYS TO ATTAIN THE POWER TO PREDICT

The vendors evaluated in this Forrester Wave provide general-purpose PAML solutions that apply to any industry. Enterprises that long for the power to predict can also seek out software applications that offer predictive models as features within applications — no data scientist needed. Enterprises that don't have data scientists can also hire any number of consulting firms that offer predictive modeling services. The largest enterprises will likely have an in-house data science team to build differentiated predictive models, but they will also benefit from acquiring applications with predictive models built in. In addition to the vendors evaluated in this Forrester Wave, enterprises can also:

- › **Choose alternative PAML solutions not included in this evaluation.** The market for PAML is larger than the key vendors evaluated in this Forrester Wave. Startups such as datascience.com and bons.ai have emerging solutions based on open source components that are likely to be in our next PAML Forrester Wave.
- › **Use PAML capabilities embedded in other tools.** Enterprise software vendors such as Alteryx, Oracle, and TIBCO Software have deep PAML capabilities via R embedded in BI tools and/or database. Salesforce launched Salesforce Einstein billed as “AI for everyone,” which translates to a PAML solution.
- › **Buy or turn on predictive capabilities in vertical or horizontal solutions.** Many vendors provide solutions that focus on specific industry or horizontal domains, such as customer analytics.¹² For example, NGDATA and RichRelevance specifically focus on customer-experience-focused predictive models for acquisition, retention, cross-sell/upsell, and targeted marketing campaigns. Oracle offers “adaptive intelligent apps” that are predictive models for optimized payment terms, best-fit candidates for hiring, smart offers and actions, and freight planning.
- › **Hire an expert consulting firm.** 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. Large consulting companies such as Infosys and KPMG have data science practices. Boutique firms Beyond The Arc, Clarity Solutions, and Think Big (a Teradata company) provide focused expertise in predictive modeling. These firms often use general-purpose solutions such as those evaluated in this Forrester Wave, but they also provide deep knowledge and domain expertise in analyzing data and building predictive models.

Predictive Analytics And Machine Learning Evaluation Overview

To assess the state of the PAML market and see how the vendors compare, Forrester evaluated the strengths and weaknesses of the top commercial and open source PAML vendors. After examining past research, user requirements, and vendor interviews and expert interviews, we developed a comprehensive set of 23 evaluation criteria, which we grouped into three high-level buckets:

- › **Current offering.** We evaluated each product’s model training scalability and deployment options, workload scalability, data preparation, algorithms, model management, and tooling for data scientists to establish the differentiated capabilities of the vendor’s current offering. All evaluated products must have been publicly available by December 15, 2016.
- › **Strategy.** We reviewed each vendor’s strategy to assess its ability to compete and grow in the enterprise PAML market. Key criteria include Forrester’s confidence in the vendor’s ability to execute on its stated strategy and support current and future customers. We also assessed each vendor’s product road map to evaluate how this will affect the vendor’s competitive position relative to the other vendors in this evaluation.

- › **Market presence.** To determine each vendor's market presence, we evaluated the install base of paying customers, PAML-specific revenue, market awareness of the vendor's PAML solution, and partnerships with other technology and services firms.¹³

EVALUATED VENDORS AND INCLUSION CRITERIA

Forrester included 14 vendors in the assessment: Alpine Data, Angoss, Dataiku, Domino Data Labs, FICO, H2O.ai, IBM, KNIME, Microsoft, RapidMiner, Salford Systems, SAP, SAS, and Statistica. Each of these vendors has (see Figure 1):

- › **A comprehensive, differentiated PAML solution.** The evaluated vendors must offer a solution designed for data scientists that can operate on large data sets, provide capabilities for data acquisition and preparation, provide statistical and machine learning algorithms, provide a differentiated user interface to build models, and provide model management features. If a vendor offers a PAML that is based in whole or in part on open source, the vendor must have value-added differentiation.
- › **A standalone PAML solution.** We only included PAML solutions that are not technologically embedded in any particular application, BI, data prep, ETL, or middleware stacks.¹⁴ Vendors should offer a focused, standalone product as defined by Forrester for the PAML category. For example, a BI vendor that includes the ability to run R scripts would not be included in this Forrester Wave.
- › **Install base and revenue history.** Included vendors must have 10 or more named, paying enterprise customers using the PAML solution and at least \$4 million in revenue over the last 12 months.
- › **Motivated client inquiries.** Forrester clients often discuss the vendors and products through inquiries; alternatively, vendors may, in Forrester's judgment, warrant inclusion or exclusion in this evaluation because of technology trends, market presence, or lack of client interest.

FIGURE 1 Evaluated Vendors: Product Vendor Information And Selection Criteria

Vendor	Product name
Alpine Data	Chorus
Angoss	KnowledgeSTUDIO
Dataiku	Dataiku DSS
Domino Data Labs	Domino
FICO	FICO Decision Management Suite
H2O.ai	H2O
IBM	IMB SPSS Modeler
KNIME	KNIME Analytics Platform
Microsoft	Cortana Intelligence Suite, Microsoft R Server
RapidMiner	RapidMiner Platform
Salford Systems	The SPM Salford Predictive Modeler software suite
SAP	SAP BusinessObjects Predictive Analytics, SAP HANA SPS
SAS	SAS Analytics Suite
Statistica	Statistica

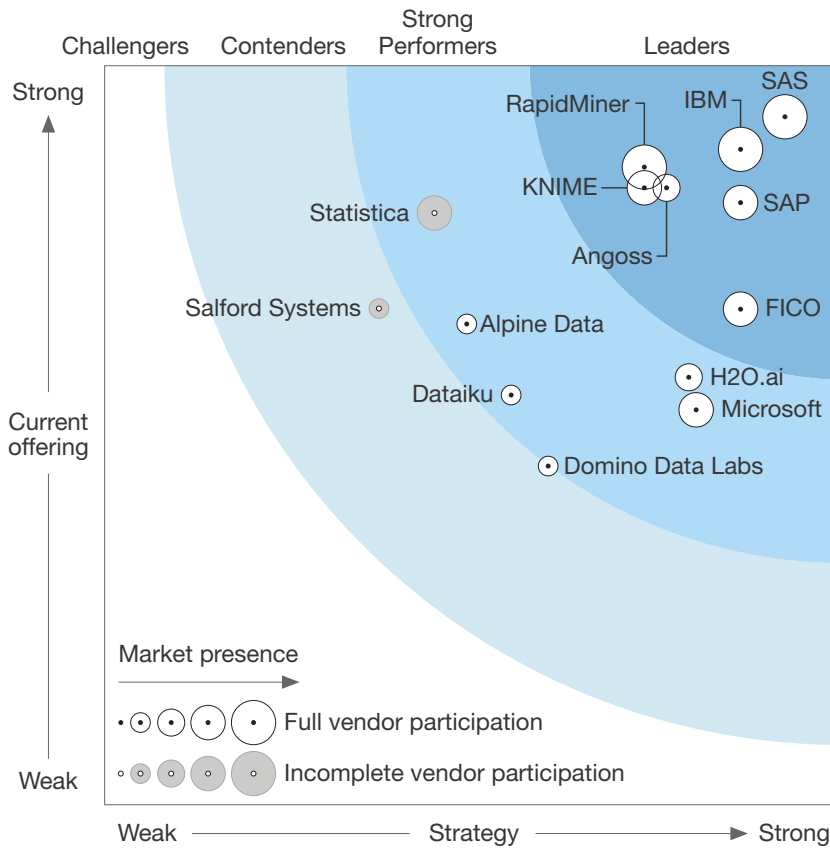
Vendor inclusion criteria

- A comprehensive, differentiated PAML solution.** The vendors included in this evaluation must offer a solution designed for data scientists that can operate on large data sets, provide capabilities for data acquisition and preparation, provide statistical and machine learning algorithms, provide a differentiated user interface to build models, and provide model management features. If a vendor offers a predictive analytics and machine learning (PAML) solution that is based in whole or in part on open source, the vendor must have value-added differentiation.
- A standalone PAML solution.** We only included PAML solutions that are not technologically embedded in any particular application, business intelligence (BI), data prep, ETL, or middleware stacks. Vendors should offer a focused, standalone product as defined by Forrester for the PAML category. For example, a BI vendor that includes the ability to run R scripts would not be included in this evaluation.
- Install base and revenue history.** Vendors included in this evaluation have 10 or more named, paying enterprise customers using the PAML solution and at least \$4 million in revenue over the last 12 months.
- Motivated client inquiries.** Forrester clients often discuss the vendors and products through inquiries; alternatively, vendors may, in Forrester's judgment, warrant inclusion or exclusion in this evaluation because of technology trends, market presence, or lack of client interest.

Vendor Profiles

This evaluation of the PAML market is intended to be a starting point only. We encourage clients to view detailed product evaluations and adapt criteria weightings to fit their individual needs through the Forrester Wave Excel-based vendor comparison tool (see Figure 2). Forrester clients can also schedule an inquiry with the author of this report to discuss their specific needs and how vendors may meet those needs.

FIGURE 2 Forrester Wave™: Predictive Analytics And Machine Learning Solutions, Q1 '17



FORRESTER RESEARCH
The Forrester Wave™
Go to Forrester.com to download the Forrester Wave tool for more detailed product evaluations, feature comparisons, and customizable rankings.

FIGURE 2 Forrester Wave™: Predictive Analytics And Machine Learning Solutions, Q1 '17 (Cont.)

	Forrester's weighting	Alpine Data	Angoss	Dataiku	Domino Data Labs	FICO	H2O.ai	IBM	KNIME	Microsoft	RapidMiner	SAP	SAS
Current Offering	50%	3.44	4.16	2.76	2.28	3.34	2.88	4.42	4.16	2.66	4.30	4.06	4.64
Architecture	20%	4.00	4.00	2.50	2.50	4.00	4.00	4.50	4.00	3.50	4.00	4.00	5.00
Data set preparation	20%	4.50	5.00	4.00	3.00	3.50	3.00	5.00	4.50	3.00	5.00	5.00	5.00
Algorithms	20%	3.00	3.40	2.20	2.20	2.20	3.80	4.20	4.60	3.40	4.60	3.00	3.80
Model management	20%	2.70	4.40	1.90	1.90	4.40	1.00	4.40	3.50	1.00	3.30	4.10	5.00
Data science tooling	20%	3.00	4.00	3.20	1.80	2.60	2.60	4.00	4.20	2.40	4.60	4.20	4.40
Strategy	50%	2.45	3.80	2.75	3.00	4.30	3.95	4.30	3.65	4.00	3.65	4.30	4.60
Acquisition and pricing	20%	3.00	3.00	3.00	3.00	3.00	5.00	3.00	5.00	3.00	5.00	3.00	3.00
Solution road map	30%	2.00	4.00	3.00	3.00	4.00	4.00	4.00	3.00	3.00	3.00	4.00	5.00
Ability to execute	25%	2.00	4.00	3.00	3.00	5.00	4.00	5.00	4.00	5.00	4.00	5.00	5.00
Implementation support	25%	3.00	4.00	2.00	3.00	5.00	3.00	5.00	3.00	5.00	3.00	5.00	5.00
Market Presence	0%	1.83	2.83	1.83	1.40	3.30	2.43	4.48	3.30	3.30	4.30	3.83	5.00
Vendor sizing	70%	1.75	2.75	1.75	2.00	3.00	1.75	4.25	3.00	3.00	4.00	3.75	5.00
ISV and services partners	30%	2.00	3.00	2.00	0.00	4.00	4.00	5.00	4.00	4.00	5.00	4.00	5.00

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

LEADERS

- › **SAS reimagines its data science portfolio.** SAS is unifying its comprehensive portfolio of data science solutions under SAS Visual Suite. It brings together world-class data prep, visualization, data analysis, model building, and model deployment. This unified tooling approach provides a consistent user experience that data scientists need to build even the most sophisticated models. SAS's vision for data science is not limited to innovation in tools. It has been quick to jump on new, promising analytical methods across multiple disciplines, such as statistics, econometrics, optimization, machine learning, deep learning, and natural language interaction. It recently introduced support for calling SAS analytics from Python, Java, and Lua, leveraging open source data science notebooks. A key challenge is that SAS has a target on its back by the open source zealots that summarily and wrongly dismiss SAS as old school. Customers complain about premium pricing compared with other solutions.
- › **IBM is loving open source.** SPSS is still the core of IBM's data science platform, but IBM is launching projects such as SystemML from its investments in its Spark Technology Center. IBM has also introduced the Data Science Experience for data science coders, which provides a quick cloud provisioning of open source Jupyter and/or RStudio notebooks with a Spark cluster on the back end to run data pipelines and train models. SPSS is a good fit for data scientists who want the productivity afforded by methods encapsulated in operators. IBM's perennial challenge with a comprehensive portfolio is that customers often misunderstand it because there are multiple products that seem to solve the same problem.
- › **SAP draws a straight-line from predictive models to business applications.** SAP offers comprehensive data science tools to build models, but it is also the biggest enterprise application company on the planet. This puts SAP in a unique position to create tools that allow business users with no data science knowledge to use data-scientist-created models in applications. SAP's solution offers the data tools that enterprise data scientists expect, but it also offers distinguished automation tools to train models. The solution has plenty of room to grow into its existing applications customer base, but its dependence on SAP's HANA data platform will limit its attractiveness to non-SAP customers.
- › **Angoss is ready to be your primary solution.** Angoss KnowledgeSEEKER is a must-have for data science teams that wish to use beautiful and comprehensive visual tools to build decision and strategy trees. That's what Angoss is known for, but its portfolio is actually very well-rounded: It also offers KnowledgeSTUDIO for building models, InsightOPTIMIZER for numerical optimization, and KnowledgeMANAGER for model management. Angoss recently added a coding environment that allows data scientists to use programming languages including R, Python, and the language of SAS. It also has some integration with Hadoop and Apache Spark. Angoss has impressive adoption for a small company, but its key challenge is to create market awareness in an increasingly crowded field of data science startups.

- › **RapidMiner wraps breadth and depth in a beautiful package.** RapidMiner invested heavily to revamp its visual interface, making it the most concise and fluid that we have seen in this evaluation. It also has a comprehensive set of operators that encapsulate a wide range of data prep, analytical, and modeling functionality to increase productivity of data scientists. RapidMiner is open source and has a community that contributes to its growing list of operators. RapidMiner Studio is free to download and use for up to 10,000 data rows, with tiered pricing for more than 10,000 rows of data.
- › **KNIME's vibrant open source community pays dividends in productivity.** KNIME is not a big company, but it has a big community of contributors who continually push the platform forward with capabilities such as bioinformatics and image processing. The KNIME Analytics Platform is free to download and use and includes over 1,000 analytical and model building operators. The vendor funds its ongoing operations by offering commercial extensions for team collaboration, such as the KNIME Server for sharing workflows, advanced security, and remote execution of model building workflows. Maybe KNIME was smart for not taking gobs of venture funding during the big data rush. It relies on the community for innovation. However, the result is that sometimes its enterprise features lag larger vendors as the community focuses on other areas, such as new analytical methods.
- › **FICO makes enterprise decisions smarter with models.** FICO's extensive real-world experience has led to a solution that focuses on the needs of the chief data scientists as well as the rank-and-file data scientists in a large organization. Chief data scientists at mature enterprises demand three key things from data scientists: 1) explainable models, 2) accurate models, and 3) decision management. FICO's Decision Management Suite encompasses the end-to-end-capabilities needed to create, deploy, and monitor models for use in complex, consequential enterprise decisions. FICO needs to expand the number of algorithms it supports to compete more broadly. The company struggles to explain to prospects that it is not just about the consumer FICO Score, especially in North America. The software division is strong, with an impressive customer base across multiple industries.

STRONG PERFORMERS

- › **H2O.ai puts algorithms first.** H2O.ai is best known for developing open source, cluster-distributed machine learning algorithms at a time (2011) when big data demanded them but no one else had them. To say H2O.ai is an algorithm company today is an understatement. It also offers Sparkling Water to create, manage, and run workflows on Apache Spark and Steam to deploy models. Further, it offers Flow — a notebook-like experience similar to Jupyter. The company recently announced Deep Water and amalgamated distribution of open source deep-learning libraries Caffe, MXNet, and TensorFlow. H2O.ai has seen significant adoption by large enterprises including Capital One, Comcast, and Macy's. However, H2O.ai does not often displace heavies such as SAS because it only recently added a data science workbench and lacks enterprise model management capabilities. The company's operations and growth are funded by Series B venture capital.

- › **Microsoft is much more than R for enterprises.** Microsoft offers Microsoft R for data scientists who wish to code in the R programming language supported by callable cluster-distributed algorithms. It also offers Azure Machine Learning to data scientists who want a more traditional visual development tool. There is no reason why enterprises cannot enjoy both. Microsoft's machine learning cloud services offer pretrained models for such tasks as image labeling, voice recognition, and natural language processing that allow developers with no data science knowledge to use them in applications. Microsoft invested heavily in R, thinking it was the go-to language for data science, only to find a few years later that Python was rising in popularity. Just like application development, data science coding has become polyglot. Microsoft needs to release its bias toward R and go everywhere data scientists want to go, just like the company has done with application development.
- › **Alpine Data focuses on collaboration.** Data scientists spend an inordinate amount of time preparing data and conversing with business stakeholders compared with the time they spend on building valuable models. Alpine Data's visual tool provides data engineers, data scientists, and business stakeholders with the capabilities they need to divide and conquer the work of building models. Data engineers can use the tool to prep data. Data scientists and business stakeholders can communicate using built-in collaboration features. Founded in 2010, Alpine Data was the first vendor to design a solution to work with Hadoop. That's not enough for Alpine to succeed, though. It must improve its capabilities across the board to stay ahead of the upstarts and stay in the game with the established heavies.
- › **Dataiku gets code or click right.** A haiku is a Japanese form of poetry of 17 syllables — concise and evocative if done well. That's Dataiku's guiding inspiration — to offer a data science platform that lets coders use a notebook when they must, but use visual tools to build workflows when productivity is at a premium. Dataiku is one of the new venture-funded startups that aim to be a well-rounded alternative to the long-time market competitors by offering a notebook experience embedded in a visual experience. With better model management capabilities, Dataiku is poised to challenge the leaders posthaste.
- › **Statistica finds a new home, again.** Statistica was founded in 1984 as Statsoft and acquired by Dell in 2014 as part of Dell's focus on building an enterprise software portfolio. It is now part of the newly relaunched Quest Software. The Statistica solution is based on a data science workbench that has a rich set of algorithms and data prep tools that are especially relevant for manufacturers and scientific use cases. Statistica was a nonparticipating vendor in this evaluation.

CONTENDERS

- › **Domino Data Labs wants coders to collaborate across open source.** The open source options for data science coders have never been richer. Data science notebooks such as Jupyter allow coders to be more productive and collaborative, and language libraries of algorithms continue to proliferate. The problem: Many of these tools are not integrated, leading to a disjointed

development experience that becomes a burden to maintain, especially for larger enterprise teams. To counter this, Domino Data Labs' solution aims to package the most popular open source coding tools and libraries and provide a unifying interface for teams of data science coders. But this approach also has a drawback: Many of the features critical to enterprises, such as model management and advanced workbench tools, lack open source options. We don't think Domino can wait for the open source community to add critical enterprise features.

- › **Salford Systems touts accuracy and automation.** Salford Systems is adored by its community of customers, large and small, for its implementation of specific methods including CART, MARS, Random Forests, and TreeNet. Most other vendor solutions have one or more of these methods, but Salford claims that its methods are the best because they are implemented by their inventors — including Jerome Friedman, a professor of statistics at Stanford University. Salford provides a workbench tool for modeling and has added automation scenarios to improve data science productivity. Its focus on creating the most accurate models has been at the expense of adding features like model management and big data analysis that a broader range of enterprise customers needs. Salford Systems was a nonparticipating vendor in this evaluation.

Engage With An Analyst

Gain greater confidence in your decisions by working with Forrester thought leaders to apply our research to your specific business and technology initiatives.

Analyst Inquiry

To help you put research into practice, connect with an analyst to discuss your questions in a 30-minute phone session — or opt for a response via email.

[Learn more.](#)

Analyst Advisory

Translate research into action by working with an analyst on a specific engagement in the form of custom strategy sessions, workshops, or speeches.

[Learn more.](#)

Webinar

Join our online sessions on the latest research affecting your business. Each call includes analyst Q&A and slides and is available on-demand.

[Learn more.](#)



Forrester's research apps for iPhone® and iPad®

Stay ahead of your competition no matter where you are.

Supplemental Material

ONLINE RESOURCE

The online version of Figure 2 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 three data sources to assess the strengths and weaknesses of each solution. We evaluated the vendors participating in this Forrester Wave, in part, using materials that they provided to us by December 15, 2016.

- › **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.
- › **Product demos.** We asked vendors to conduct demonstrations of their products' functionality. We used findings from these product demos to validate details of each vendor's product capabilities.
- › **Customer reference calls.** To validate product and vendor qualifications, Forrester also conducted reference calls with 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 for evaluation in this market. From that initial pool of vendors, we 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 evaluation — and then score the vendors based on a clearly defined scale. We intend these default weightings to serve only as a starting point and 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. For more information on the methodology that every Forrester Wave follows, go to <http://www.forrester.com/marketing/policies/forrester-wave-methodology.html>.

INTEGRITY POLICY

We conduct all our research, including Forrester Wave evaluations, in accordance with our Integrity Policy. For more information, go to <http://www.forrester.com/marketing/policies/integrity-policy.html>.

Endnotes

- ¹ Yogi Berra (1925-2015) was an American professional baseball catcher, manager, and coach. He was known for his “Yogi-isms” — witty nuggets of truth.
- ² Probabilistic predictive models can help enterprises make better decisions. See the Forrester report “[Prescriptive Analytics: The Black Belt Of Digital Decisions.](#)”
- ³ Our prior evaluation of this market, published on April 1, 2015 as “The Forrester Wave™: Big Data Predictive Analytics Solutions, Q2 2015,” emphasized in the title the ability of these solution to allow data scientists to build predictive models using the largest of data sets — big data. We decided to drop “big data” in the title since all of these PAML solutions have an ability to use big data albeit in varying degrees. For this evaluation, we also decided to add “machine learning” to the title to reflect the substantial use of the term by vendors and end user companies to describe these solutions. All of the solutions we evaluated in our Q2 2015 Forrester Wave also supported machine learning. There are often different terms used in the market that describe the same category. The PAML category can also be labeled as “predictive analytics,” “advanced analytics,” “machine learning,” and “data mining,” and some vendors are starting to use “artificial intelligence.”
- ⁴ In the predecessor to this evaluation (“The Forrester Wave™: Big Data Predictive Analytics Solutions, Q2 2015”), Forrester defined this category 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.” The definition in this report defines the same category as the prior definition — both define the categories as software used by data scientists to create and deploy predictive models. We’ve refined the newer definition to introduce the key users of the software — data scientists — and also call out the use of statistical and machine learning algorithms. In this latest definition, we also dropped the mention of hardware, since this is a software category. Finally, we dropped the emphasis on data and generic uses cases because we believe it is well understood that predictive models are created by analyzing data.
- ⁵ Source: Forrester Data: World Business Intelligence And Analytics Software Forecast, 2016 To 2021 (Global).
- ⁶ See the Forrester report “[Massive Machine-Learning Automation Is The Future Of Data Science.](#)”
- ⁷ Standards such as predictive modeling markup language (PMML) and portable format for analytics (PFA) specify a format to represent a model that can be embedded in platforms that support the standard.
- ⁸ See the Forrester report “[Prescriptive Analytics: The Black Belt Of Digital Decisions.](#)”
- ⁹ See the Forrester report “[Apache Spark Is Powerful And Promising.](#)”
- ¹⁰ See the Forrester report “[Artificial Intelligence: What’s Possible For Enterprises In 2017](#)” and see the Forrester report “[TechRadar™: Artificial Intelligence Technologies And Solutions, Q1 2017.](#)”
- ¹¹ See the Forrester report “[The Forrester Wave™: Customer Analytics Solutions, Q1 2016.](#)”
- ¹² See the Forrester report “[The Forrester Wave™: Customer Analytics Solutions, Q1 2016.](#)”
- ¹³ Some vendors in this evaluation also offer products in markets other than PAML. Our assessment for market awareness is based only on the market awareness of the vendors PAML product, not the overall awareness of the vendor.
- ¹⁴ ETL: extract, transform, load.

We work with business and technology leaders to develop customer-obsessed strategies that drive growth.

PRODUCTS AND SERVICES

- › Core research and tools
- › Data and analytics
- › Peer collaboration
- › Analyst engagement
- › Consulting
- › Events

Forrester's research and insights are tailored to your role and critical business initiatives.

ROLES WE SERVE

Marketing & Strategy Professionals

CMO
B2B Marketing
B2C Marketing
Customer Experience
Customer Insights
eBusiness & Channel Strategy

Technology Management Professionals

CIO
› **Application Development & Delivery**
Enterprise Architecture
Infrastructure & Operations
Security & Risk
Sourcing & Vendor Management

Technology Industry Professionals

Analyst Relations

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.



Angoss is a global leader in delivering advanced analytics to businesses looking to improve performance across risk, marketing and sales. With a suite of big data analytics software solutions and consulting services, Angoss delivers powerful approaches that provide you with a competitive advantage by turning your information into actionable business decisions. Many of the world's leading organizations in financial services, insurance, retail and high tech rely on Angoss to grow revenue, increase sales productivity and improve marketing effectiveness while reducing risk and cost. Headquartered in Toronto, Canada, with offices in the United States, United Kingdom and Singapore, Angoss serves customers in over 30 countries worldwide. For more information, visit www.angoss.com