

The Forrester Wave™: Notebook-Based Predictive Analytics And Machine Learning, Q3 2020

The 12 Providers That Matter Most And How They Stack Up

by Kjell Carlsson, PhD and Mike Gualtieri
September 10, 2020

Why Read This Report

In our 26-criterion evaluation of notebook-based predictive analytics and machine learning (PAML) providers, we identified the 12 most significant ones — Amazon Web Services, Anaconda, Civi Analytics, Cloudera, Databricks, Domino Data Lab, Google, MathWorks, Microsoft, OpenText, Oracle, and RStudio — and researched, analyzed, and scored them. This report shows how each provider measures up and helps application development and delivery (AD&D) professionals select the right one for their needs.

Key Takeaways

Microsoft, Cloudera, Domino Data Lab, Google, And MathWorks Lead The Pack

Forrester's research uncovered a market in which Microsoft, Cloudera, Domino Data Lab, Google, and MathWorks are Leaders; Amazon Web Services, RStudio, and Oracle are Strong Performers; and Anaconda, OpenText, Databricks, and Civi Analytics are Contenders.

Operationalizing Trusted AI Apps And A Strategy For Success Are Key Differentiators

The top notebook-based PAML offerings will drive your AI transformation with new capabilities for developing and explaining AI apps. More important, they offer ModelOps capabilities to help you rapidly deploy and manage your AI models in production. But it's not just about existing tech. Enterprises should choose platforms with the resources, roadmap, and partners that can help them now and into the future.

The Forrester Wave™: Notebook-Based Predictive Analytics And Machine Learning, Q3 2020

The 12 Providers That Matter Most And How They Stack Up

by [Kjell Carlsson, PhD](#) and [Mike Gualtieri](#)
with [Srividya Sridharan](#) and [Robert Perdoni](#)
September 10, 2020

Table Of Contents

- 2 Notebook-Based PAML Will Operationalize Open Source AI
- 3 Evaluation Summary
- 6 Vendor Offerings
- 7 Vendor Profiles
 - Leaders
 - Strong Performers
 - Contenders
- 12 Evaluation Overview
 - Vendor Inclusion Criteria
- 14 Supplemental Material

Related Research Documents

- [The Forrester New Wave™: Automation-Focused Machine Learning Solutions, Q2 2019](#)
- [The Forrester Wave™: Multimodal Predictive Analytics And Machine Learning, Q3 2020](#)
- [Introducing ModelOps To Operationalize AI](#)
- [Now Tech: Predictive Analytics And Machine Learning, Q2 2020](#)



Share reports with colleagues.
Enhance your membership with
Research Share.

The Forrester Wave™: Notebook-Based Predictive Analytics And Machine Learning, Q3 2020

The 12 Providers That Matter Most And How They Stack Up

Notebook-Based PAML Will Operationalize Open Source AI

The power of open source communities and their tech-giant benefactors continues to drive an unprecedented expansion of machine learning (ML) use cases, model accuracy, explainability, and many other formerly magic capabilities. Enterprise data science teams need Python, R, TensorFlow, PyTorch, and a host of other open source ML languages and frameworks to get the latest and greatest in ML algorithms and attract top data science talent. That said, ML algorithms are only one component of a much fuller lifecycle that includes data preparation, feature engineering, training, evaluation, experiments, and model operations (ModelOps). That's why code-first data science teams need a notebook-based PAML platform to support this full lifecycle. These platforms bring software development discipline to otherwise unruly data science processes, increasing productivity, boosting collaboration, and improving time-to-value. As a reminder, [Forrester segments PAML into three segments: multimodal, notebook-based, and automation-focused](#). This Forrester Wave™ evaluates vendors in the notebook-based segment.

Notebook-based PAML customers should look for providers that help you:

- › **Amp up code-first model development and team collaboration.** Notebook-based PAML offerings take notebook-based data science to the next level by making individual data scientists dramatically more productive, especially by providing collaboration capabilities for teams. These platforms accelerate the agile development of successful models — with capabilities to manage ML experiments — but also rapidly prototype interactive end-user-facing AI applications. They also drive adoption with features that provide granular explanations for both models and their predictions, and they help users test different models to see which ones suit their business needs.
- › **Operationalize AI models at scale with ModelOps.** A top complaint of data science teams and, increasingly, line-of-business leaders is the challenge of deploying machine learning models in production. Frantic handoffs, manual processes, and loose governance impede organizations' ability to deploy more business-worthy AI use cases faster. Most PAML vendors have traditionally focused their product development efforts on analytical tools and ML methods. That's still important, but now that AI has gone more mainstream, enterprises want more ML, faster. ModelOps capabilities within PAML products enable cross-functional AI teams to efficiently deploy, monitor, retrain, and govern AI models in production systems.¹
- › **Support your broader AI needs, now and into the future.** As enterprise AI workloads have skyrocketed, tech titans have poured development dollars into expanding and upgrading their PAML offerings while investors double down on the most promising startups.² An increasingly important part of notebook-based PAML is the rich ecosystem of AI capabilities that these vendors offer, spanning AI infrastructure to pretrained AI models offered as a service. Enterprises need to take these considerations into account when they choose their notebook-based platforms. Choose vendors with a rich roadmap of critical capabilities and innovations, and back it up with a dense network of partners that can support an ever-expanding catalog of AI use cases.

The Forrester Wave™: Notebook-Based Predictive Analytics And Machine Learning, Q3 2020

The 12 Providers That Matter Most And How They Stack Up

Evaluation Summary

The Forrester Wave evaluation highlights Leaders, Strong Performers, Contenders, and Challengers. It's an assessment of the top vendors in the market and does not represent the entire vendor landscape. You'll find more information about this market in our reports on [predictive analytics and machine learning](#).

We intend this evaluation to be a starting point only and encourage clients to view product evaluations and adapt criteria weightings using the Excel-based vendor comparison tool (see Figure 1 and see Figure 2). Click the link at the beginning of this report on Forrester.com to download the tool.

The Forrester Wave™: Notebook-Based Predictive Analytics And Machine Learning, Q3 2020

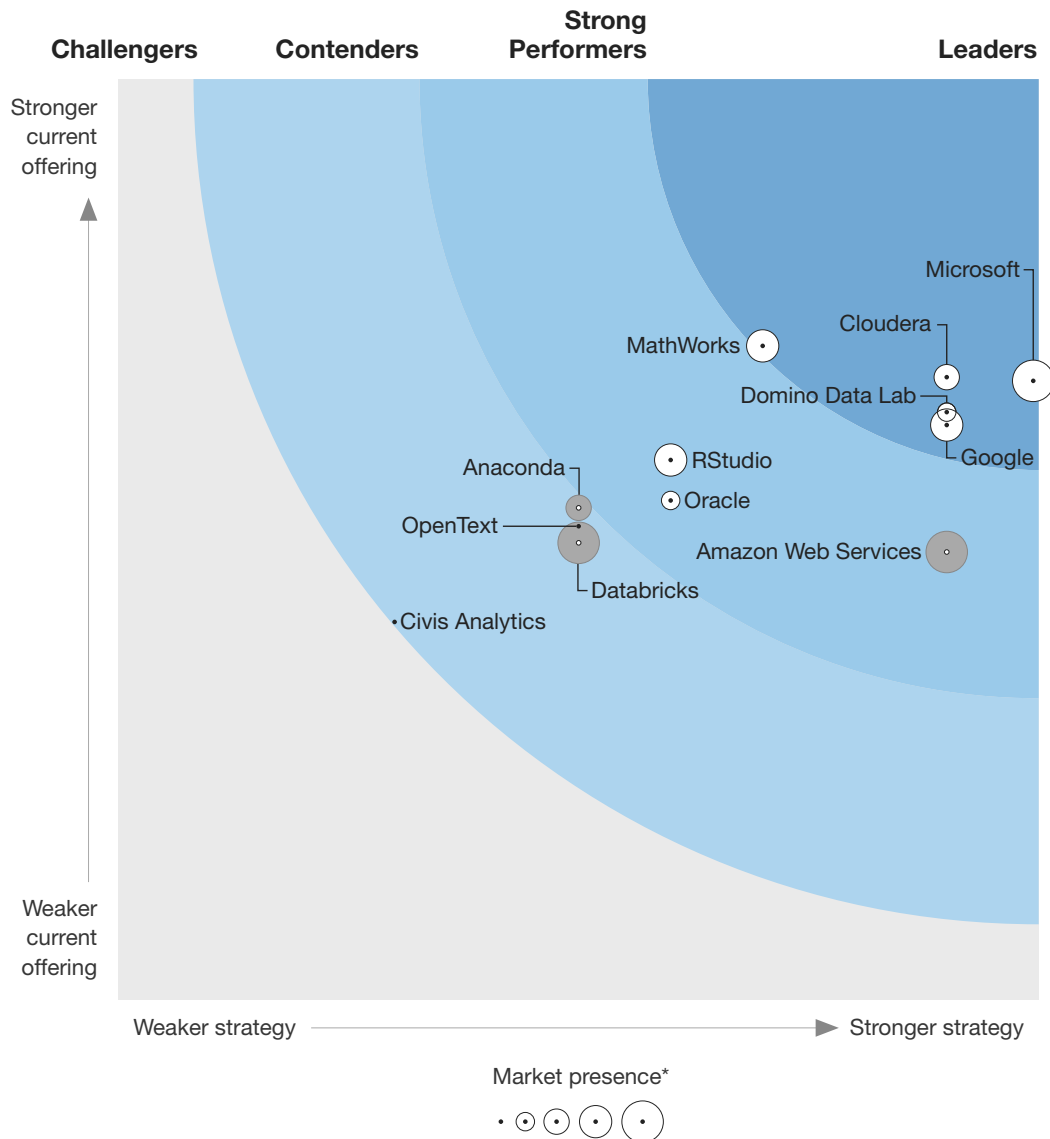
The 12 Providers That Matter Most And How They Stack Up

FIGURE 1 Forrester Wave™: Notebook-Based Predictive Analytics And Machine Learning, Q3 2020

THE FORRESTER WAVE™

Notebook-Based Predictive Analytics And Machine Learning

Q3 2020



*A gray bubble indicates a nonparticipating vendor.

The Forrester Wave™: Notebook-Based Predictive Analytics And Machine Learning, Q3 2020

The 12 Providers That Matter Most And How They Stack Up

FIGURE 2 Forrester Wave™: Notebook-Based Predictive Analytics And Machine Learning Scorecard, Q3 2020

	Forrester's weighting	Amazon Web Services*	Anaconda*	Civis Analytics	Cloudera	Databricks*	Domino Data Lab	Google	MathWorks	Microsoft	OpenText	Oracle	RStudio
Current offering	50%	2.43	2.67	2.05	3.38	2.48	3.19	3.12	3.55	3.36	2.57	2.71	2.93
Data	14%	2.00	1.00	2.00	2.00	1.00	1.00	2.50	3.50	3.50	3.00	2.00	1.50
Modeling	14%	2.33	2.33	1.67	3.00	2.33	2.33	3.00	3.00	3.67	2.33	3.00	1.67
Collaboration	14%	3.00	4.33	3.67	5.00	4.33	5.00	3.00	3.67	4.33	3.00	3.00	4.33
Model evaluation	14%	2.33	1.67	1.00	1.00	1.67	2.33	4.33	4.33	3.00	1.00	4.33	1.67
Model operations (ModelOps)	14%	2.33	2.33	1.00	3.67	1.00	3.67	3.00	2.33	3.00	1.67	1.67	2.33
Methods and algorithms	14%	4.00	4.00	4.00	4.00	4.00	3.00	5.00	3.00	5.00	4.00	4.00	4.00
Platform infrastructure	14%	1.00	3.00	1.00	5.00	3.00	5.00	1.00	5.00	1.00	3.00	1.00	5.00
Strategy	50%	4.50	2.50	1.50	4.50	2.50	4.50	4.50	3.50	5.00	2.50	3.00	3.00
Ability to execute	25%	5.00	1.00	1.00	3.00	3.00	5.00	5.00	5.00	5.00	1.00	3.00	3.00
Solution roadmap	25%	5.00	1.00	1.00	5.00	1.00	5.00	3.00	3.00	5.00	3.00	3.00	3.00
Enablement	25%	3.00	3.00	3.00	5.00	3.00	5.00	5.00	5.00	5.00	3.00	3.00	3.00
Partners	25%	5.00	5.00	1.00	5.00	3.00	3.00	5.00	1.00	5.00	3.00	3.00	3.00
Market presence	0%	4.33	2.33	1.00	2.33	4.33	1.67	3.67	3.67	4.33	1.00	1.67	3.67
Customer adoption	33%	5.00	1.00	1.00	3.00	5.00	1.00	3.00	5.00	5.00	1.00	3.00	5.00
Evaluated product revenue	33%	3.00	1.00	1.00	1.00	3.00	1.00	3.00	3.00	3.00	1.00	1.00	1.00
Market awareness	33%	5.00	5.00	1.00	3.00	5.00	3.00	5.00	3.00	5.00	1.00	1.00	5.00

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

*Indicates a nonparticipating vendor.

The Forrester Wave™: Notebook-Based Predictive Analytics And Machine Learning, Q3 2020

The 12 Providers That Matter Most And How They Stack Up

Vendor Offerings

Forrester included 12 vendors in this assessment: Amazon Web Services, Anaconda, Civis Analytics, Cloudera, Databricks, Domino Data Lab, Google, MathWorks, Microsoft, OpenText, Oracle, and RStudio (see Figure 3).

FIGURE 3 Evaluated Vendors And Product Information

Vendor	Product evaluated	Product version evaluated
Amazon Web Services	AWS SageMaker	
Anaconda	Anaconda Enterprise Edition	
Civis Analytics	Civis Platform	
Cloudera	Cloudera Machine Learning (CML)	CML Public Cloud May 2020 Release; Cloudera Data Science Workbench (CDSW) 1.7.2
Databricks	Data Science Workspace	
Domino Data Lab	Domino Data Science Platform	4.2
Google	AI Platform Notebooks	
MathWorks	MATLAB	R2020a
Microsoft	Azure Machine Learning	
OpenText	OpenText Magellan	20.2
Oracle	Oracle Cloud Infrastructure Data Science	May Release
RStudio	RStudio Server Pro (RSP); RStudio Connect (RSC); RStudio Package Manager (RSPM)	RSP 1.3; RSC 1.8.2; RSPM 1.1.4

The Forrester Wave™: Notebook-Based Predictive Analytics And Machine Learning, Q3 2020

The 12 Providers That Matter Most And How They Stack Up

Vendor Profiles

Our analysis uncovered the following strengths and weaknesses of individual vendors.

Leaders

- › **Microsoft provides coding data scientists with all the bells and whistles.** From what was a collection of disparate PAML offerings — Azure ML Workbench, Azure ML Studio, and Azure Batch AI — Microsoft has forged a new unified offering, Microsoft Azure Machine Learning. The result is transformational. Microsoft Azure Machine Learning offers a full suite of enterprise PAML capabilities, from centralized model registries to hyperparameter tuning and modular model training and deployment pipelines. Microsoft has paid particular attention to collaboration — e.g., making it possible for users to work simultaneously in the same notebook, and integrations with Jenkins and GitHub interactions to enable MLOps capabilities as well as fairness and responsible machine learning — e.g., building in capabilities to test models on sensitive variables like age and gender, recommending mitigation models, and protecting data by adding noise or enabling eyes-off training. Microsoft Azure Machine Learning also has AutoML wizards, drag-and-drop tools for building ML pipelines, and integrations to build models within SQL editors to support developers, data analysts, and other non-data scientists who want to build and deploy models.

The major cloud vendors have long had a gap in offering a comprehensive PAML platform that meets the full set of enterprise data science team needs, to the detriment of bewildered customers who have had to build or find their own solutions. Microsoft has filled that gap and then some. Between the strength of its sales teams, size of its existing customer base, and Microsoft's own massive internal usage of Microsoft Azure Machine Learning, the success of Microsoft's PAML strategy is a near certainty. Indeed, its future is azure'd.

- › **Cloudera is a machine learning — machine.** Machine learning is the engine, but data . . . sweet, sweet data is the fuel. Data science projects rise and fall on the availability of timely, clean, and especially complete data sets. As part of the Cloudera Data Platform, Cloudera Machine Learning automatically inherits a comprehensive set of data management capabilities designed to operate at enterprise scale. These world-class, integrated data management capabilities provide data scientists using Cloudera Machine Learning with all the horsepower and governance they need to build and deploy models using open source.

Cloudera Machine Learning offers strengths in open source, collaboration, governance, applications, and platform infrastructure. Enterprises that want a fully integrated data platform and open source machine learning that runs on-premises, in the public cloud, or in the private cloud will find Cloudera Data Platform (together with Cloudera Machine Learning) compelling. To be more competitive, Cloudera should invest in better modeling tools, including automation, and while it offers an increasingly robust set of ModelOps capabilities, it would benefit from easier-to-use model monitoring tools.

The Forrester Wave™: Notebook-Based Predictive Analytics And Machine Learning, Q3 2020

The 12 Providers That Matter Most And How They Stack Up

- › **Domino Data Lab unifies the tools your data science teams want to use.** Getting a data scientist to consistently use a single PAML tool is hard. Getting your teams to agree on a single PAML tool is impossible. It is also self-defeating, since the tools have different strengths and you will always want to take advantage of the latest ML innovations. Domino tames the chaos, bringing all your different PAML tools together and binding them in a common, governed platform. It lets your data scientists use the tools they want, whether it be Domino's own Jupyter, RStudio, and Zeppelin notebooks and integrated development environments (IDEs) or third-party tools like AWS SageMaker, DataRobot, MATLAB, and SAS. Domino drives productivity by abstracting away infrastructure provisioning, managing clusters, tracking experiments, maintaining version control, and deploying and monitoring models. And it drives collaboration with built-in knowledge management tools and shared repositories for data, code, model artifacts, and apps, irrespective of where they were developed.

Domino provides an enterprise data science platform that supports the diversity of ML options that users need in today's rapidly expanding PAML ecosystem, with repeatability discipline and governance. To remain competitive, Domino Data Lab must continue to be quick at making popular open source tools available within its platform and maintain relationships with the larger data management and ML ecosystem.

- › **Google offers one-stop AI shopping on Google Cloud Platform.** Google's AI Platform Notebooks offering, made generally available in March 2020, lets data scientists rapidly spin up a JupyterLab notebook environment — preconfigured for a range of open source ML frameworks — that has built-in integrations with Google's AI Platform. These related services include BigQuery (for data storage), Dataprep (for data preparation), Dataproc (for large scale data processing), Data Labeling (for labeling data), AI Platform Training (for training jobs), AI Platform Prediction (for deploying models), Kubeflow (for deploying models on-premises), and the What-If Tool (for explainability).

Google has services to support the full AI lifecycle, and it develops a host of AI innovations in both hardware and software that it often shares with the open source community, such as TensorFlow. Google's AI Platform Notebooks service is a convenient, scalable tool for data scientists looking to leverage Google Cloud Platform for training or deploying models, especially deep learning models. However, to be more competitive, it needs more modeling, collaboration, and ModelOps capabilities.

- › **MathWorks' MATLAB is machine learning for polymaths.** These days machine learning isn't just for data scientists. It is fast becoming an essential tool for engineers and scientists in many fields. Similarly, data scientists need to know a lot more than machine learning. MathWorks' MATLAB fits the bill by offering a broad range of data analysis techniques, including machine learning, math, optimization, statistics, and visualizations. The MATLAB programming language is the unifying layer underpinning all of these platform capabilities. However, MATLAB includes a plethora of other graphical user interfaces (GUIs), such as wizards and visualization tools, that are lacking in most other notebook-based PAML solutions. Want to use AutoML to build a deep learning module? There's a module for that. Need a workbench for annotating audio files? There's a module for that.

The Forrester Wave™: Notebook-Based Predictive Analytics And Machine Learning, Q3 2020

The 12 Providers That Matter Most And How They Stack Up

MathWorks' MATLAB offers strengths in the breadth of its data analysis methods, visualization, and modeling interfaces. Engineering and science teams that already use MATLAB will find ML capabilities in the box, including deep learning. To be more competitive, MathWorks should do more to support users who primarily want to use open source programming languages, but in the interim these users can leverage the MATLAB engine using Python, and MATLAB users can call Python functions within the MATLAB environment.

Strong Performers

- › **Amazon Web Services weaves a web of sagacious ML services.** From its not-so-humble beginnings in late 2017 as a collection of ML algorithms offered as a cloud service, AWS SageMaker has developed into a more complete PAML offering that covers the PAML lifecycle. Indeed, it is starting to outpace competitors by introducing innovative capabilities to support the broader lifecycle of an AI application. These include Ground Truth (a data labeling service), the Step Functions Data Science SDK (for rapidly building data and ML deployment pipelines), Model Monitor (for monitoring ML models in production), Augmented AI (human review for low-confidence predictions), and, to the delight of anyone training deep learning models, SageMaker Debugger.³ For model development, SageMaker Studio offers an increasingly comprehensive and integrated notebook environment, and SageMaker Autopilot distinguishes itself as an AutoML capability by creating fully transparent notebooks for each model it trains.

Given AWS's popularity for data storage and application development, it always had a head start when it comes to attracting cloud ML workloads, and it has built a widening set of frequently innovative PAML capabilities. To be more competitive, AWS needs to further integrate these services into a unified offering that can more seamlessly support the end-to-end workflow of enterprise data science teams. Amazon Web Services declined to participate in the full Forrester Wave evaluation process.

- › **RStudio is good for the R world.** RStudio is the public benefit corporation dedicated to providing free, open source tools for the R programming language (and Python, for those who use both) and supporting its ongoing development. It offers the most popular IDE for R, RStudio, for free, and it shepherds the "Tidyverse" of packages that are designed to make R easier to learn and use, and for data scientists to use easily together. To support these efforts, it offers the commercial RStudio Team bundle, which includes Server Pro (providing collaboration, security, and shared infrastructure for data scientists), Package Manager (for governing the R packages that data science teams use), and Connect (for securely hosting and sharing models, apps, and visualizations created using R and Python).

RStudio Team provides the core functionality for enterprise teams that primarily want to use R and some Python to develop and deploy their models, and it is offered at a rock-bottom price. For any recently established teams, teams doing advanced data science, teams that aspire to do more machine learning, or teams with very small budgets for tools (and, of course, R and Python enthusiasts), RStudio is their best friend.

The Forrester Wave™: Notebook-Based Predictive Analytics And Machine Learning, Q3 2020

The 12 Providers That Matter Most And How They Stack Up

- › **Oracle brings enterprise data science to its cloud.** Oracle acquired one of the most promising notebook-based startups, Datascience.com, in 2018 and completed the transfer to Oracle Cloud Infrastructure (OCI), making it available to customers, in February 2020. The offering retains its many features that were cutting-edge at the time — e.g., supporting multiple open source notebooks and IDEs, self-service provisioning of ML environments, and one-click deployment of models — though many have become widely available since then. However, OCI Data Science still has plenty of unique tricks up its sleeves, thanks to its proprietary Accelerated Data Science (ADS) SDK, which brings a host of easy-to-use functions to simplify (i.e., accelerate) each step in a data scientist’s workflow. For example, ADS will generate descriptive statistics, histograms, correlation maps, and identifying data quality issues with a single command. It will suggest and apply data cleaning actions and feature transformations and has an AutoML function to automatically test multiple methods and hyperparameter combinations. It also has evaluation and explainability functions to automatically validate, test, explore, and explain models and their results.

OCI Data Science blends the flexibility and extensibility of open source with enterprise support and security, as well as an ease of use expected from many multimodal PAML tools within a notebook environment. It needs more ModelOps capabilities to make it easier to deploy, monitor, and manage models in production as well as additional, sustained investment in ADS if Oracle wants to retain this as a differentiator.

Contenders

- › **Anaconda is the powerhouse behind Python data science.** Anaconda popularized the use of Python for data science in 2009 by creating the most well-known open source distribution of Python packages for data science. Data scientists’ use of Python as a programming language is as strong as ever, and Anaconda continues to offer the Conda distribution that includes more than 250 commonly used open source packages for data science. For enterprise customers, the company offers Anaconda Enterprise Edition to support data science teams with collaboration, governance, and deployment.

Anaconda’s strengths are in its breadth of notebook support, open source packages, applications, and security. Particularly noteworthy are its capabilities for governing the open source libraries and packages available to data science teams, which are increasingly being licensed by other data science platforms. Anaconda declined to participate in the full Forrester Wave evaluation process.

- › **OpenText excels at exploring and wrangling data.** OpenText’s Magellan includes integration with Jupyter notebooks to provide data science teams with full access to open source packages. Where Magellan really shines is data exploration and data preparation, thanks to an integrated set of specialized tools. Magellan includes sophisticated tools for both business analysts and data scientists alike to connect, visualize, and prepare data not only for data science projects but for business intelligence (BI) as well. Magellan also includes proprietary text analytics algorithms and is evaluated in “The Forrester Wave™: AI-Based Text Analytics Platforms (Document Focused), Q2

The Forrester Wave™: Notebook-Based Predictive Analytics And Machine Learning, Q3 2020

The 12 Providers That Matter Most And How They Stack Up

2020.”⁴ While Magellan’s use of Jupyter notebooks to train machine models presents itself as a notebook-based PAML solution, its GUI tools for data make that part of the product accessible to noncoders as well.

OpenText offers strengths in data exploration, data preparation, and text analytics. To be more competitive, OpenText has to either invest more in notebook-based features, ModelOps, and support for a wider array of open source frameworks or go the opposite route and invest in GUI tools for training models to appeal to noncoding data science teams — or both.

- › **Databricks unifies Spark with machine learning for the cloud.** Databricks was founded in 2013 by the original creators of Apache Spark to create a unified platform for all things data. Apache Spark has seen tremendous adoption by enterprises as a general data processing platform that supports the most basic to the most complex data processing workloads, including machine learning. Specifically, Databricks offers Data Science Workspace for coding data scientists who wish to leverage the horsepower of in-memory Apache Spark data processing and open source machine learning. Databricks offers solutions in AWS (as Databricks Unified Data Analytics on AWS) and on Microsoft Azure (as Azure Databricks).

Databricks’ strengths are in data processing workloads, collaboration capabilities, and open source. To be more competitive, Databricks should focus more on differentiated tools that make data science teams more productive rather than code-centric frameworks such as MLflow, and it should build out its ModelOps capabilities for deploying, monitoring, and governing ML models. Databricks declined to participate in the full Forrester Wave evaluation process.

- › **Civis Analytics is machine learning for people, about people.** Although Civis Platform can be used by data science teams for most use cases, it differentiates by focusing on customer intelligence use cases. This is thanks to both its proprietary data sets on US customers and additional integrated products, including a community outreach and engagement hub, census intelligence center, identity resolution, and others focused on reaching audiences. Civis Platform integrates Jupyter for coding and includes differentiated data tools to prepare person-level data sets for data science tasks.

Civis Analytics offers strengths in data preparation, security, and differentiated algorithms for person-level data sets. Civis’ sweet spot is helping data science teams that wish to leverage open source packages and need data quality, data privacy protection, and feedback loops geared toward audience engagement. The platform is available in cloud only, but it is both SOC 2 Type II-certified and HIPAA-compliant for data protection and privacy.⁵

The Forrester Wave™: Notebook-Based Predictive Analytics And Machine Learning, Q3 2020

The 12 Providers That Matter Most And How They Stack Up

Evaluation Overview

We evaluated vendors against 26 criteria, which we grouped into three high-level categories:

- › **Current offering.** Each vendor's position on the vertical axis of the Forrester Wave graphic indicates the strength of its current offering. Key criteria for these solutions include data, modeling, collaboration, model evaluation, model operations, methods and algorithms, and platform infrastructure.
- › **Strategy.** Placement on the horizontal axis indicates the strength of the vendors' strategies. We evaluated their ability to execute, solution roadmap, enablement, and partners.
- › **Market presence.** Represented by the size of the markers on the graphic, our market presence scores reflect each vendor's customer adoption, evaluated product revenue, and market awareness.

Vendor Inclusion Criteria

Forrester included 12 vendors in the assessment: Amazon Web Services, Anaconda, Civis Analytics, Cloudera, Databricks, Domino Data Lab, Google, MathWorks, Microsoft, OpenText, Oracle, and RStudio. Each of these vendors has:

- › **A notebook-based PAML solution, as identified by Forrester.** Vendors included in this evaluation must offer a notebook-based PAML solution as defined above and as represented in Forrester's Now Tech report on this market.⁶ Vendors identified by Forrester as offering multimodal PAML or automation-focused PAML solutions were not included in this Forrester Wave evaluation.
- › **A comprehensive, differentiated notebook-based PAML solution.** The vendors included in this evaluation must offer a solution that can operate on large data sets; provide capabilities for data acquisition and preparation; and provide statistical and ML algorithms, programming languages commonly used for data science, a differentiated user interface to build models, and model management features. If a vendor offers a PAML based in whole or in part on open source, the vendor must have value-added differentiation; for example, a BI vendor that includes the ability to run R scripts would not be included in this Forrester Wave, nor would a vendor that offers a cloud service for undifferentiated Jupyter notebooks.
- › **A standalone notebook-based PAML marketed to enterprise data science/AI teams.** Forrester included only solutions that are marketed toward enterprise data science and/or AI teams that use ML algorithms and other analytical algorithms to build predictive models and surface insights. PAML solutions that Forrester deemed to be technologically embedded into any particular application, BI, data prep, applications, ETL, or middleware stacks were not included in this evaluation.⁷

The Forrester Wave™: Notebook-Based Predictive Analytics And Machine Learning, Q3 2020

The 12 Providers That Matter Most And How They Stack Up

- › **Minimum install base and revenue requirements.** The vendor must have at least three paying, named enterprise customers using its PAML solution. The vendor must have provided Forrester with three customer references that were willing to speak with us and fill out a survey. Please note that reference calls and surveys will be kept under nondisclosure. Included vendors must also have a trailing 12-month revenue of at least \$4 million.
- › **Sparked client inquiries 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 or exclusion in this evaluation because of technology trends, market presence, lack of client interest, or loss of market momentum.

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 iOS and Android.

Stay ahead of your competition no matter where you are.

The Forrester Wave™: Notebook-Based Predictive Analytics And Machine Learning, Q3 2020

The 12 Providers That Matter Most And How They Stack Up

Supplemental Material

Online Resource

We publish all our Forrester Wave scores and weightings in an Excel file that provides detailed product evaluations and customizable rankings; download this tool by clicking the link at the beginning of this report on Forrester.com. We intend these scores and default weightings to serve only as a starting point and encourage readers to adapt the weightings to fit their individual needs.

The Forrester Wave Methodology

A Forrester Wave is a guide for buyers considering their purchasing options in a technology marketplace. To offer an equitable process for all participants, Forrester follows [The Forrester Wave™ Methodology Guide](#) to evaluate participating vendors.

In our review, we conduct primary research to develop a list of vendors to consider for the evaluation. From that initial pool of vendors, we narrow our final list based on the inclusion criteria. We then gather details of product and strategy through a detailed questionnaire, demos/briefings, and customer reference surveys/interviews. We use those inputs, along with the analyst's experience and expertise in the marketplace, to score vendors, using a relative rating system that compares each vendor against the others in the evaluation.

We include the Forrester Wave publishing date (quarter and year) clearly in the title of each Forrester Wave report. We evaluated the vendors participating in this Forrester Wave using materials they provided to us by June 12, 2020, and did not allow additional information after that point. We encourage readers to evaluate how the market and vendor offerings change over time.

In accordance with [The Forrester Wave™ Vendor Review Policy](#), Forrester asks vendors to review our findings prior to publishing to check for accuracy. Vendors marked as nonparticipating vendors in the Forrester Wave graphic met our defined inclusion criteria but declined to participate in or contributed only partially to the evaluation. We score these vendors in accordance with [The Forrester Wave™ And The Forrester New Wave™ Nonparticipating And Incomplete Participation Vendor Policy](#) and publish their positioning along with those of the participating vendors.

Integrity Policy

We conduct all our research, including Forrester Wave evaluations, in accordance with the [Integrity Policy](#) posted on our website.

The Forrester Wave™: Notebook-Based Predictive Analytics And Machine Learning, Q3 2020

The 12 Providers That Matter Most And How They Stack Up

Endnotes

- ¹ See the Forrester report [“Introducing ModelOps To Operationalize AI.”](#)
- ² For example, Domino Data Lab received \$43 million in series E funding in June 2020, Civis Analytics raised \$15 million in May 2020, and Databricks received \$400 million in series F funding in October 2019.
- ³ SDK: software development kit.
- ⁴ See the Forrester report [“The Forrester Wave™: AI-Based Text Analytics Platforms \(Document Focused\), Q2 2020.”](#)
- ⁵ HIPAA: Health Insurance Portability and Accountability Act.
- ⁶ See the Forrester report [“Now Tech: Predictive Analytics And Machine Learning, Q2 2020.”](#)
- ⁷ ETL: extract, transform, load.

We work with business and technology leaders to drive customer-obsessed vision, strategy, and execution that accelerate growth.

PRODUCTS AND SERVICES

- › Research and tools
- › Analyst engagement
- › Data and analytics
- › Peer collaboration
- › Consulting
- › Events
- › Certification programs

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.