

# Magic Quadrant for Data Science and Machine Learning Platforms

Published 11 February 2020 - ID G00385005 - 84 min read

---

Expert data scientists and other professionals working in data science roles require capabilities to source data, build models and operationalize machine learning insights. Significant vendor growth, product development and myriad competing visions reflect a healthy market that is maturing rapidly.

## Market Definition/Description

This Magic Quadrant evaluates vendors of data science and machine learning (DSML) platforms. Gartner defines a DSML platform as a core product and supporting portfolio of coherently integrated products, components, libraries and frameworks (including proprietary, partner and open source). Its primary users are data science professionals. These include expert data scientists, citizen data scientists, data engineers and machine learning (ML) engineers/specialists.

Coherent integration means that the core product and supporting portfolio provide a consistent “look and feel” and create a user experience where all components are reasonably interoperable in support of an analytics pipeline.

The DSML platform offers a mixture of basic and advanced functionality essential for building DSML solutions (primarily predictive and prescriptive models). The platform also supports the incorporation of these solutions into business processes, surrounding infrastructure, products and applications. It supports variously skilled data scientists in multiple tasks across the data and analytics pipeline, including all of the following areas:

- Data ingestion
- Data preparation
- Data exploration
- Feature engineering
- Model creation and training
- Model testing
- Deployment

- Monitoring
- Maintenance
- Collaboration

Not all organizations build their DSML models from scratch or entirely on their own. Some need assistance with getting started with or extending their DSML initiatives. Although this Magic Quadrant does assess the availability of some prepackaged content, such as templates and samples, it does not assess service providers that can help jump-start or extend DSML projects throughout an organization. Nor does this Magic Quadrant assess specialized vendors of industry-, domain- or function-specific solutions.

Readers of this Magic Quadrant should understand the following:

- This market features a diverse range of vendors: Gartner invited a wide range of DSML platform vendors to participate in the evaluation process for potential inclusion in this Magic Quadrant. Users of these platforms have different requirements and preferences for user interfaces (UIs) and tools (see “How to Choose the Right Data Science and Machine Learning Platform”). Expert data scientists prefer to code data science models (typically in Python or R) and often build and run data models in notebooks. Other users are most comfortable building models by using a drag-and-drop UI to create visual pipelines. Many members of emerging citizen data science communities favor a much more augmented approach (often called AutoML). This uses artificial intelligence/machine learning techniques to guide and accelerate practitioners through the model building and operationalization process (see “Augmented Machine Learning Is Democratizing Data Science”). Some expert data scientists prefer an augmented approach, which enables them to navigate the model building and operationalization process more efficiently and productively.
- Platforms aimed primarily at application developers and business analysts are explicitly excluded: The data science and traditional analytics spaces have been colliding at multiple key points for several years now. Additionally, many developers are getting educated around machine learning and incorporating it into their applications and software. The lines between roles and toolkits continue to blur. However, there remains clear distinction between market spaces based on intended user types. For platforms and services designed for these constituencies, see “Magic Quadrant for Cloud AI Developer Services” (forthcoming) and “Magic Quadrant for Analytics and BI Business Intelligence Platforms.”
- A Leader may not be the best choice: There’s a wide range of DSML products available, all of which offer a breadth and depth of capability and varied approaches to developing, operationalizing and managing models. It is therefore important to evaluate your specific needs when assessing vendors. A vendor in the Leaders quadrant, for example, might not be the best choice for you. Equally, a Niche Player might be the perfect choice. For an extensive review of the functional capabilities of each platform, see the (forthcoming) companion “Critical Capabilities for Data Science and Machine Learning Platforms.” Bear in mind that this Magic Quadrant includes only a small selection of the myriad vendors offering various DSML solutions.

- Only vendors with commercially licensable products are included: Pure open-source platforms are excluded from this Magic Quadrant. Only commercially licensed open-source platforms are included. We do, however, recognize the well-established trend of commercial platforms leveraging open-source components and libraries. Vendors take different approaches to including and supporting open source. Open-source solutions represent an opportunity for both users and vendors to get started with DSML with little upfront investment (see Note 1). Innovation is fast-paced within the open-source community and success of new technologies is highly democratic. In addition, many users of DSML platforms are either already proficient in or can easily learn and apply open-source technologies. Open-source technologies have also become ubiquitous in university data science curricula. Leveraging open source through collaborative or orchestrated integration with commercial offerings also reduces the need for vendors to re-create specific capabilities. Vendors can incorporate the best elements from a fast-changing landscape of algorithms and techniques, leaving more resources to focus on other areas of differentiation for their platforms. However, a platform's ease of use and coherence may suffer if its vendor does not account for the needs of all types of user.
- Platforms must support not only model building but also model operationalization: The full benefit — including business value — of DSML will not be achieved unless models are both:
  - Embedded in business processes and decision environments
  - Maintained, monitored and managed over time
- There have been numerous recent advances in technology, process and talent development. However, an alarming percentage of models developed with the intention of full deployment are never actually operationalized. There are many reasons for this, but a crucial one is a lack of tools to enable and facilitate operationalization. Operationalization (often referred to as MLOps) extends to ongoing review and adjustment of models to ensure their relevancy over time as the business and its objectives change. MLOps also includes key functionality such as drift detection, catalogs, governance, explainability and business impact analysis.
- AI is overhyped and inescapable: All DSML can be classified as artificial intelligence (AI), but not all AI concepts should be called DSML. Still, DSML platforms cannot avoid being swept up in the sustained hype around AI. The semantics are unlikely to ever be agreed upon and are not worth fighting over. AI hype brings undoubtedly valuable attention and enthusiasm to the data science space. But without education, discipline and reasonable expectations, that hype can do far more harm than good. The vendors in this Magic Quadrant have done a fine job embracing the hype while clearly communicating and delivering value and differentiating their platforms from other AI solutions. The diversity of DSML platforms largely reflects the wide range of people that use them. This Magic Quadrant is therefore aimed at a variety of audiences:

- Expert data scientists: These are the highly sought-after individuals who possess the skills and knowledge to understand and engage all stage of the data science life cycle. Most expert data scientists spend the largest share of their time and energy on model creation, with supporting roles such as data engineers and ML engineers taking on data pipelining and MLOps responsibilities. Tenured experts can move into data science manager roles, using platforms to gain visibility into a team’s full portfolio of projects and facilitating collaboration and timely delivery of value. Some expert data scientists work mostly independently on “point” solutions and rarely collaborate much with other data scientists or departments within their organization.
- Citizen data scientists: Increasingly, citizen data scientists are building DSML models. These are people who need access to DSML capabilities, but who do not have the advanced skills of expert data scientists. Citizen data scientists can come from roles such as business analyst, line of business (LOB) analyst, data engineer and application developer. They need to understand the nature of the DSML market and how it differs from, but complements, the analytics and business intelligence (BI) market (see “Magic Quadrant for Analytics and Business Intelligence Platforms”). Citizen data scientists do not replace expert data scientists but, instead, work in collaboration with them.
- Supporting roles: These include data engineers, developers, ML engineers and other roles. While not responsible for model building, training and testing, the supporting cast in data science teams are vital to scaling operations and ensuring data quality and consistent model accuracy.
- Line of business (LOB) data science teams: Typically, these are sponsored by their LOB executive and charged with addressing LOB-led initiatives in areas such as marketing, sales, finance and R&D. These teams focus on their own and their department’s priorities. Levels of collaboration with other LOB data science teams vary. LOB data science teams can include both expert and citizen data scientists. Supporting roles may reside in the LOB or be assigned from IT or other areas.
- Corporate data science teams: These have strong and broad executive sponsorship, and can take a cross-functional perspective from a position of enterprisewide visibility. In addition to model building support, they are often charged with defining and supporting an end-to-end process for building and deploying DSML models. They often work in partnership with LOB data science teams in multitier organizations. In addition, they might provide assistance for LOB teams that do not have their own data scientists. Corporate data science teams typically include expert data scientists. Supporting roles may reside in the corporate data science team or be assigned from IT or other areas.

## Magic Quadrant

Figure 1. Magic Quadrant for Data Science and Machine Learning Platforms

Source: Gartner (February 2020)



## Vendor Strengths and Cautions

### Altair

[Altair](#) is headquartered in Troy, Michigan, U.S. In December 2018, Altair completed its acquisition of Datawatch, which had acquired Angoss and its main data science product components the previous January. The Datawatch product portfolio is now known as Altair Knowledge Works and the core product considered for this Magic Quadrant evaluation is Knowledge Studio. The Knowledge Works suite also includes Knowledge Studio for Apache Spark, Knowledge Seeker, Knowledge Hub, Knowledge Manager, Panopticon and Monarch.

Knowledge Works' Online User Community provides customers with the ability to collaborate and learn with fellow end users and data science professionals using this suite of products. To this end, Altair offers a series of customer training days in several cities. Beyond normal customer support, on-site and premium support are available at a negotiated price. Altair also offers a variety of consulting and managed service offerings to support building and deploying models.

Altair enters this Magic Quadrant as a Niche Player (both Datawatch and Angoss were Niche Players in the previous two years). Scores for its Completeness of Vision and Ability to Execute are tempered by the risks and uncertainties associated with the product suite's second acquisition in as many years. Altair may be the ideal home for the evolving combined portfolios of Datawatch and Angoss, but a transition period is ongoing. Knowledge Studio has also fallen behind in several areas of high vendor activity, including augmented DSML, MLOps and cutting-edge use cases.

### Strengths

- Ease of use and balance of coder and noncoder appeal: Knowledge Studio's highly rated UI will attract visually oriented data science professionals. Coders will also find a flexible environment in which to exploit the best of Python and R. This dual appeal is needed in most modern data science teams. The interface has the opportunity to broaden and strengthen its appeal by adding augmented DSML capabilities.
- Customer satisfaction: Altair inherits a strong legacy of customer success with Knowledge Works. Customers are happy across multiple touchpoints with Altair. This includes account management, integration and deployment, and general service and support.
- Potential in new industries and development with other Altair products: Knowledge Works will continue to be known for its traditional strengths (decision trees, financial services use cases). Integration with Altair products and a holistic strategy will open a wider range of opportunities. Altair is positioned to introduce Knowledge Works to new markets and sales channels.

### Cautions

- Second acquisition of product portfolio in two years: All acquisitions — no matter how strategic and dovetailed — come with uncertainty and risk. Allowing for a transition period, Altair cannot delay important updates and roadmap deliveries that clients and prospects are demanding. The challenge facing Altair is to continue integrating and optimizing the Datawatch acquisition while positioning and advancing Knowledge Works in an accelerated and often brutally competitive market.
- Gaps in vision and current offering: Altair needs to catch up with thought leaders in key areas such as augmented DSML, MLOps and explainable AI. Stronger support for containerization and Kubernetes is needed. Altair will need to improve in these areas to maintain expert appeal and citizen data scientist mind share.

- Slower adoption compared to competitors: Market adoption of the core Knowledge Works products have been slow compared to competitor offerings. Particularly in the past two years, while Knowledge Works has been changing hands, several competitors have sustained hypergrowth and developed market-leading products.

## Alteryx

[Alteryx](#) is based in Irvine, California, U.S. It provides four software products that comprise its DSML platform: Alteryx Connect, Alteryx Designer, Alteryx Server and Alteryx Promote. Alteryx Designer is the core product.

Alteryx hosts an Alteryx Community that provides customers, partners and Alteryx associates with a platform to interact with one another. Alteryx organizes Alteryx Inspire Conferences in North America, Europe and Australia annually. The vendor provides a multilingual global support for customers out of U.S., U.K., Australia and Singapore offices. Alteryx has returned to its 2018 position of Leader, from being a Challenger in 2019, by demonstrating a solid company and product vision, especially in relation to augmented DSML and process automation. The vendor continues to outperform almost all other vendors in this Magic Quadrant from a revenue growth perspective and has significantly expanded its business internationally.

Alteryx has made significant progress in changing the perception that it is solely a data preparation provider. It made two strategic acquisitions (ClearStory Data and Feature Labs) in 2019 to expand its platform capabilities. ClearStory Data provides a solution that enables automation of analytics of complex data and unstructured data on large-scale data processing platforms like Apache Spark. Feature Labs automates feature engineering, the creation of AI applications and data preparation process to help improve model accuracy and overall process efficiency. The Feature Labs acquisition occurred after the cutoff for this research report and was not considered in scoring.

## Strengths

- Market perception and execution: Alteryx's focus on addressing the end-to-end analytic process clearly positions it as a vendor with a comprehensive platform. It is no longer simply viewed as a data preparation vendor by the market. Alteryx's value proposition is clear and resonating in the market, as demonstrated by its strong revenue and customer growth.
- No-code approach: Alteryx's no-code approach is attractive to organizations that want to use ML but need an easy-to-use platform built for business analysts and citizen data scientists. The vendor provides hundreds of automation building blocks for users to quickly build analytics, data science and process automation workflows. The platform can directly integrate with robotic process automation (RPA) solutions and enterprise applications to automate running of ML pipelines and business processes.
- A robust, growing community: Alteryx has built an engaged online community of users to share knowledge and answer questions. Its online forum offers challenge exercises to help

users improve their product knowledge. In addition, it offers three annual user conferences (U.S., Europe and APAC) that are growing quickly.

### Cautions

- **Pricing:** Alteryx's high price continues to concern some customers given open-source and other lower-cost options in the market, and this may limit broader use across organizations. Some customers report licensing complexity and being charged both by named user and by CPU core. Lack of volume discounts also limit broader use of the platform within organizations.
- **Linux support:** Of the four Alteryx products, only Promote runs on Linux OS; CentOS is the only supported distribution. Alteryx Connect and Server only run on the Microsoft Windows Server OS. This can be an issue for organizations that do not support Windows Server.
- **Streaming/Internet of Things (IoT) capabilities:** Alteryx has limited capabilities for supporting streaming/IoT, which is important for real-time decision making. Sample use cases include those related to IoT and logistics such as failure prediction, asset management and supply chain optimization.

### Anaconda

[Anaconda](#) is based in Austin, Texas, U.S. It offers Anaconda Enterprise (this analysis excludes the Anaconda Distribution version), a data science development environment based on the interactive notebook concept that supports users in using open-source Python- and R-based packages. Anaconda continues to provide a loosely coupled distribution environment, which offers access to a wide range of open-source packages and open-source libraries, primarily Python-based.

The company is committed to a very large and engaged community of open-source and enterprise users whose needs are an important driver of Anaconda's innovations.

AnacondaCON is the vendor's annual conference. The active community is an important source for user support, although enterprise customers can also rely on a support desk and professional services.

Anaconda remains a Niche Player in this year's Magic Quadrant. It is well-suited to data scientists who embrace Python or R and who are eager to explore a continuous stream of new open-source capabilities, while still benefiting from an environment more structured and integrated than a pure notebook environment. However, this flexibility comes at the price of a lack of automation and the complexities of finding one's way in a frequently changing range of open-source capabilities.

Recently, Anaconda has further improved its scalability to serve greater numbers of concurrent users. Better support for security and graphics processing unit (GPU) deployments are other noteworthy improvements.

### Strengths



- Python and open-source support: The popularity of Python among data scientists gives Anaconda great visibility to developers. Anaconda offers its users consolidated access to Python, R and other open-source tools, libraries and rapidly integrated innovations.
- Active community: Anaconda is the only data science vendor not just supporting but also securing the Python open-source community. The community fosters cutting-edge Python code libraries and integration with other open-source data science projects. Anaconda Cloud provides means of collaboration and code library exchanges for data scientists and developers to explore and accelerate model development production.
- Scalable development for open-source libraries: Anaconda's scalability takes several forms. Users can scale workloads with GPUs and can leverage Apache Hadoop, Apache Hadoop YARN or Kubernetes clusters, on-premises or in the cloud.

### Cautions

- Designed for experts: Anaconda targets experienced data scientists who are often familiar with Python, R and notebooks like Jupyter. Despite the flexibility of these environments, they are not conducive to fruitful discussions with business stakeholders — a capability that is increasingly valued by organizations wanting to improve the alignment between business and data science.
- Open-source shortcomings: Like many open-source promoters, Anaconda suffers from the usual drawbacks associated with large and flexible developer communities: compatibility issues between packages and versions; uncertain outcomes of new, sometimes overlapping projects; and, despite ongoing progress in terms of workbench homogeneity, challenges with overall coherence, especially for nonexpert users.
- Automation and operationalization: The Anaconda environment is lagging in capabilities such as autoML, MLOps and explainability. Novice Anaconda users will have difficulty finding their way through the Python “jungle.” Citizen data scientists will find themselves in uncharted territory within Anaconda's environment.

### Databricks

[Databricks](#)' headquarters are in San Francisco, California, U.S. The vendor's Unified Data Analytics Platform spans data science, ML and data engineering, emphasizing ease of infrastructure automation and scalability. Rooted in Apache Spark, Databricks keeps contributing to the open-source community by leading MLflow, Delta Lake and Koalas. Its commercial platform further enables customers to keep up with a variety of releases and fast-paced updates of popular open-source frameworks and libraries, which Databricks packages and optimizes under ML Runtime.

Databricks has been leading the Apache Spark community with two Spark+AI Summits in the U.S. and Europe, and with numerous meetups across the globe. The company constantly evolves its training and certification programs. It has a global support team and offers several tiers of support depending on size of deployment and required response times.

In the past year, Databricks has demonstrated strong execution of product development, revenue growth and customer satisfaction, especially in the enterprise sector, as well as international expansion in Europe and Asia/Pacific. Reference customers indicated that a key detail in their selection of the Databricks platform was the vendor's product roadmap and future vision. These two factors underpin Databricks' position as a Leader in this Magic Quadrant.

### Strengths

- Strong execution, expansion and partnerships: Databricks exhibited strong execution and consequent growth across multiple industries and geographies. Its vertical sales strategy works well for its customers. Databricks' growth also comes from a well-executed strategy on developing a partner ecosystem of over 500 companies worldwide.
- Commitment to customer success: Databricks aims to rapidly deliver value from the platform with consulting services, professional services and education. Reference customers especially praised its customer success engineers who work with them side by side.
- Databricks is pushing the envelope of scalability: Databricks' platform enables scalability of ML model deployments, notebook-based ML workflows, jobs automation, and management of data infrastructure and pipelines. Reference customers especially praised its easy and scalable cluster management.

### Cautions

- Heavy slant toward technical audience: Databricks' messaging to its core audience is technical. The vendor relies on developer relations and advocacy. A proprietary data science notebook solution, Databricks Notebooks, is the dominant method for data scientists to develop solutions on the platform. Databricks Unified Data Analytics Platform has a relatively low barrier to entry for data scientists and data engineers with coding backgrounds. However, adoption is harder for business analysts and other emerging citizen data scientists.
- No generally available on-premises offering: The Databricks offering is generally available in selected public clouds only. An on-premises Databricks platform can be procured as a one-off deployment from the vendor. It is also currently available from a third party, the company's OEM partner Booz Allen Hamilton, as part of its Open Data Platform offering.
- Tight association with Spark plays down Databricks as a DSML platform: Many customers chose Databricks not because they were looking for a DSML platform, but because they were interested in options to execute Apache Spark. These customers had Hadoop or a data lake on Amazon Web Services or on Microsoft Azure, and they see Databricks as a data processing platform, where ML is in the background. Companies that are not interested in Spark often do not consider Databricks as a DSML platform option.

### Dataiku

[Dataiku](#) is headquartered in New York City, U.S., and has a main office in Paris, France. Dataiku's core product is Data Science Studio (DSS), which has a focus on cross-discipline collaboration and ease of use.

Dataiku has an online user community that meets at several annual conferences both in the U.S. and Europe. Along with partners, this community can provide and share plug-ins through a marketplace to extend Dataiku's standard offerings. Dataiku has grown its ecosystem of alliances, partners and resellers to include over 80 organizations worldwide. It offers support according to license-specific SLAs, as well as professional services for strategy, development, implementation and training.

Dataiku is positioned as a Leader this year. This is mainly due to its ease of use, vision and ability to support multiple user types in collaboration — from data engineers and data scientists, to citizen data scientists. In 2019, Dataiku continued to strengthen its capabilities in relation to scalability, data governance and augmented ML. It has become a mature platform, supporting a variety of use cases. Dataiku could further improve its position by extending its advanced and real-time analytics functionalities to support IoT and decision automation use cases.

## Strengths

- Collaboration across data science roles: From its inception, teamwork has been at the core of Dataiku's DSS platform. All the principal roles involved in developing DSML models have their place on the platform. DSS offers various UI endpoints, all pointing to the same execution core, which makes it one of the most coherent offerings evaluated in this Magic Quadrant.
- Ease of use: The qualities of Dataiku most often highlighted by clients are that its platform is relatively easy to learn and provides a rapid path to productivity. Augmented ML capabilities and the explicit possibility of delivering interpretable or "white box" models also contribute to the platform's intuitive feel and short learning curve. Yet expert data scientists can also do development in notebook style.
- Data governance: In the past year, DSS has extended its capabilities to support governance and compliance with GDPR and other laws and regulations. This includes data pseudonymization, highlighting of sensitive data to remind users of data protection rules and processes, and restrictive access to sensitive projects and data sources. The framework approach allows the incorporation of multiple internal and external policies and regulations.

## Cautions

- Advanced analytics: Dataiku has improved its support for graph and time series analytics, but it does not seem to prioritize simulation and optimization capabilities, which are increasingly important in the context of decision automation. Support for rule processing is announced, but not available yet.

- Precanned solutions: Dataiku does not offer horizontal or vertical solutions, for which it is relying on plug-ins provided by partners or other third parties. Although it makes sense that the vendor wants to focus on developing and offering a generic platform to further democratize data science, Dataiku should continue to further strengthen its partner ecosystem to enrich its offering with solutions to catalyze data science adoption.
- Streaming and the IoT: Dataiku's DSS seems to be favored by service-centric organizations and, to a lesser extent, those in asset-intensive industries. This could be explained by the fact that Dataiku does not offer strong IoT capabilities and real-time analytics functions, which are often associated with asset-centric use cases.

## DataRobot

Headquartered in Boston, Massachusetts, U.S., [DataRobot](#) provides an augmented DSML platform, incorporating automation across the end-to-end analytics process. The platform enables citizen data scientists to productively access and use data science techniques and expert data scientists to gain efficiencies and access to new algorithms, as well as test biases. The AI Success Plan offered by DataRobot enables presale and postsale support via a team of field engineers, AI success managers, account executives and customer-facing data scientists (CFDSs). The team works with customers to identify and prioritize projects and provides guidance, support and training to get models into production and maintain them over time. DataRobot's broad network of partnerships and integrations continues to grow, with a new user community site launched in October 2019 to facilitate communications between users and the DataRobot community (after the cut-off for inclusion in the Magic Quadrant evaluation). The vendor conducts annual regional user conferences at locations around the globe.

DataRobot has maintained its position as a Visionary. It acquired Cursor, a data collaboration platform, in February 2019; ParallelM, an MLOps platform in June 2019; and Paxata, a data preparation provider, in December 2019 (after the cut-off for inclusion in the Magic Quadrant evaluation). In conjunction with the acquisition of Nutonian in 2017 for augmented time series modeling and Nexosis in 2018, DataRobot continues to build out its end-to-end capabilities for a diverse set of users.

## Strengths

- Augmented analytics thought leadership and appeal for broad mix of users: DataRobot leads the charge to incorporate augmented analytics within DSML. It continues to define and demonstrate the use of augmented analytics to engage new types of users in collaboration with traditional roles. DataRobot provides capabilities beneficial to a wide variety of roles including developers, data scientists, statisticians and business analysts.
- Increasing market traction: DataRobot maintains solid traction in the market with significant growth in revenue, number of users and global name recognition. This is attributable to effective marketing and sales, as well as excellent ongoing support via CFDSs and in-product assistance.

- Hands-on onboarding and support: The AI Success Plan and its associated team helps users develop in-house data science competency, recognize and prioritize projects, jump-start initiatives and achieve fast time to insight. DataRobot is also quick to add product enhancements and respond to customer requests.

### Cautions

- Perceived as introductory, elementary tool for nonexperts: Some customers indicate concern over DataRobot's ability to address difficult or complex models. Some have considered other platforms as their needs have matured and grown to be increasingly enterprisewide. DataRobot needs to change this market perception as augmented model building is quickly becoming commoditized within both the DSML and analytics and BI market.
- AI Success Plan approach is resource-heavy: Although the use of CFDSs has worked well, the approach will be difficult to maintain if growth sustains at current high levels and customers do not transition quickly from reliance on the CFDSs. DataRobot is scaling its AI Success Plan, but it must also put increased focus on its intent to make customers self-sufficient, especially given the platform's easy-to-use augmented approach.
- Cost concerns: While most customer references indicated that DataRobot provides good value for money, the perception is that the cost is still high. Licensing costs have eliminated DataRobot from consideration for some organizations. Cost concerns have also limited the ability for some customers to deploy the platform more broadly.

### Domino

[Domino](#) is based in San Francisco, California, U.S. Domino is the industrial-strength, feature-rich and tools-agnostic platform for end-to-end DSML in the cloud or on-premises. Domino can serve (and already serves) as a central DSML platform in enterprises with large, loosely coupled data science communities. It accommodates disparate tools and provides a transparent, scalable and collaborative environment.

The annual Domino-led conference, Rev, for data science leaders and teams doubled its attendance in 2019. The vendor's reference customer score for support and service was among the highest of all vendors evaluated. Domino aims to help data science teams get the most out of its software and enable customers to become self-sufficient in the deployment, monitoring and use of the Domino platform.

Domino is positioned as a Visionary in this Magic Quadrant largely because of its product progress and roadmap, both of which manifest the vendor's deep DSML market understanding. Domino successfully enables enterprises in their industrial-strength deployments. Moving forward, the vendor needs to continue its growth and improve its marketing, market visibility and international presence.

### Strengths

- Center of the enterprise ML ecosystem: Domino's open architecture helps consolidate all data science assets and workloads within a single platform. The platform truly helps data science

teams orchestrate and streamline the ML workflow, from environment management to reproducibility and final deployment, including further monitoring.

- Business enablement and collaboration: Reference customers chose Domino to improve business process outcomes and to support collaboration between business users and data science teams. The highly collaborative environment empowers valuable exchanges within data science teams and allows managers and business stakeholders to follow results and progress. Licenses for business users are free. Domino provides business performance tracking that shows endpoint engagement for APIs, apps and other published assets.
- Governance and audit: Governance and audit capabilities reinforce Domino as an enterprise-grade platform. Its new experiment tracking is comprehensive and especially attractive to regulated industries. Domino can orchestrate ML and data processing pipelines across internal and external parties through a governed framework.

### Cautions

- Partnership strategy: Domino has good partnerships with some cloud providers and system integrators. Still, a larger and broader partner network could deliver its product to the companies that need it, increase the vendor's visibility in the DSML market, and help drive its financial success.
- Domino is geared toward large teams: Domino's collaboration, experiment management and other enterprise features serve large teams very well, but would be excessive for teams with just a few data scientists. Companies with less mature data science programs might not immediately benefit from a platform geared toward a data science community, rather than individual data scientists.
- Inflexible pricing: Domino's pricing and contract negotiations inhibit product uptake. The vendor's licensing model needs greater flexibility to accommodate various implementation and expansion scenarios within the enterprise. This is especially important to the large and distributed enterprises where the Domino platform is most attractive.

### Google

[Google](#), headquartered in Mountain View, California, U.S., offers the Google Cloud AI Platform as its core DSML platform. The platform has an expanded suite of products that includes Cloud AutoML, BigQuery ML and TensorFlow.

Google has one of the largest ML stacks in the industry. This includes hardware, a top-notch deep learning framework (TensorFlow), powerful AI APIs and massive cloud execution infrastructure. With its commitment to Kubeflow, it is increasingly executing on a "run anywhere" paradigm.

However, the Google Cloud AI Platform is currently not a stand-alone product. For a full end-to-end ML pipeline, customers require the presence of a variety of other Google Cloud products (e.g., Cloud Data Fusion, Data Studio, Datalab, Deep Learning Containers, BigQuery, Anthos) and non-Google products (e.g., Kubernetes and Cloud Dataprep by Trifacta).

Google Cloud has a strong emphasis on supporting developers, data scientists and ML experts, especially at the cutting edge. It provides plenty of video tutorials, bleeding-edge meetups and conferences (e.g., Google Cloud Next). Collaboration is further fostered by its AI Hub and Kaggle community, which serves 3.5 million users. Google Cloud has a vast set of partnerships, including hundreds of system integrators, and is expanding its set of ML services.

Google is positioned as a Visionary in this Magic Quadrant. While it is an excellent choice for top-notch data science talent, Google Cloud is still improving capabilities for citizen data scientists. Going forward, making the platform more coherent and more accessible for less advanced data science teams should be top priorities.

### Strengths

- **Emphasis on cutting-edge data science problems:** For developers, the Google Cloud AI Platform offers a massive execution framework — Google Cloud Platform (GCP) — and many different paths to develop ML models (Cloud AutoML, TensorFlow, BigQuery ML, Kubeflow, Notebooks and a variety of open-source options). Google has a strong portfolio of AI APIs (ranging from text, audio and image to video processing) and numerous first-rate packages for large-scale data handling and integration (e.g., BigQuery and Data Fusion).
- **Innovation:** Google Cloud's AI Platform offers high-end scalability and built-in data labeling capabilities. Its AI Hub (still in public beta) is one of the most promising implementations of a global algorithm market that fosters collaboration and reuse capabilities. AI Hub can serve as a model repository and feature and pipeline store internally. Google Cloud also offers strong support for streaming data with Cloud Dataproc and Cloud Dataflow.
- **Scalability:** Google Cloud's AI Platform offers a managed cloud infrastructure for autoscaling different hardware configurations, from midsize to ultralarge. It includes hardware and cloud accelerators such as Google Kubernetes Engine (GKE) cluster autoscaling, and Cloud Tensor Processing Unit (TPU) pods.

### Cautions

- **Big footprint and learning curve:** Creating a full ML development pipeline with Google Cloud's AI Platform can be challenging, especially for small to midsize data science teams lacking GCP infrastructure knowledge. But even for experienced GCP developers, the AI Platform toolchain involves multiple interconnected components and has a significant learning curve.
- **Still largely incoherent ML pipeline and lack of project management support:** The toolchain is not yet homogeneously adapted to the needs of data science projects. Instead, Google Cloud tools are used for many other industry purposes. The reuse of partial experiment pipelines is difficult. Google's roadmap suggests that these issues will be addressed in future versions of the AI Hub.

- Persistent beta and limited on-premises capabilities: Like in last year's evaluation, many of Google Cloud's most differentiated capabilities are still in public beta (e.g., AI Hub, AI Platform Data Labeling Service, Cloud AutoML Tables). Moreover, some innovative components like AutoML and all AI APIs are still not available on-premises.

## H2O.ai

H2O.ai is based in Mountain View, California, U.S. It offers a commercial product called H2O Driverless AI and support for the open-source product H2O-3 (called Sparkling Water in a version that is integrated with Apache Spark). In addition, there is an open-source, GPU-accelerated ML package called H2O4GPU. Among the vendor's strengths are its high performance ML and its vision on important trends such as augmented data science and explainability.

H2O.ai provides global customer support, complemented by a community slack channel. For advice and consultation, H2O.ai provides access to its expert data scientists and Kaggle Grandmasters as part of all commercial engagements. H2O.ai has a broad array of partners and has started to work with some of them to develop its first solutions for specific business areas or industries.

H2O.ai has maintained its position as a Visionary from last year's Magic Quadrant. Its Completeness of Vision is the strongest among the vendors evaluated, while its Ability to Execute in terms of global market visibility and presence is likely to improve as a result of ongoing expansion efforts in Europe and Asia.

## Strengths

- Augmentation (automation): H2O Driverless AI eases the adoption of DSML by offering augmentation in multiple areas. It provides augmented feature engineering, including natural language processing (NLP) functions to convert textual attributes into features. Likewise, model selection and hyperparameter tuning can be automated. Augmentation is used to reduce the complexity of time series analysis or to create visualizations of the most relevant statistics. Driverless AI can be extended and customized with open-source components that are made available through a "recipe" catalog.
- Explainability: H2O.ai sets an example in providing rich explainability functionality, using diverse techniques such as K-LIME, LIME-SUP, Shapley, variable importance, decision tree surrogate, ICE, partial dependence plots, disparate impact analysis and "what-if analysis." The AutoDoc capability automatically generates a complete set of explanations in document format.
- High-performance ML components: H2O.ai's open-source ML components set a standard for the industry, with many other platforms integrating them. Both the commercial and open-source offerings are available on all major clouds. The vendor's components are highly optimized and parallelized for CPU multicore and multinode configurations. H2O4GPU offers a software layer for significant GPU acceleration.

## Cautions



- Limited cohesion between H2O Driverless AI and open-source platform: Although H2O.ai's open-source product line can share results through the new "recipe" capability with H2O Driverless AI, their separation still impedes potential collaboration by differently skilled enterprise users. Competitive offerings, with a larger share in the market of citizen data science, provide more integral and collaborative solutions.
- Lack of certain data access and preparation capabilities: In contrast to its strong capabilities in ML, H2O.ai has room for improvement in data access and aspects of data preparation. These include data refresh, data lineage, access governance, metadata management and data catalogs, data access through REST APIs, watermarking, labeling and annotation.
- Other advanced analytics: H2O.ai does not seem to prioritize complementary and alternative techniques needed for IoT or decision automation use cases, such as stream processing, graph analytics, geospatial analysis, discrete event or agent-based simulations, design of optimization experiments and decision management.

## IBM

IBM is based in Armonk, New York, U.S. Its core product considered for this Magic Quadrant is Watson Studio. The supporting portfolio of products for Watson Studio includes Watson Machine Learning, Watson Knowledge Catalog, Watson OpenScale, SPSS (Modeler and Statistics), IBM Decision Optimization for Watson Studio, and IBM Streams.

IBM holds numerous large user conferences, along with domain-specific and regional events and workshops. It also participates in several open-source ML communities, including TensorFlow, PyTorch, Apache Spark and Kubeflow.

IBM provides support to its global customer base using a variety of channels and models. Support plans range from free on-demand digital support to fully staffed models.

IBM is the lone Challenger in this year's Magic Quadrant. With Watson Studio and its supporting portfolio, IBM has proven that it can support customer success over time. The vendor demonstrates thought leadership in hybrid and multicloud data science with Cloud Pak for Data. Improving its product bundling and corresponding go-to-market approach will compound IBM's laudable efforts to revamp its offering and to keep pace with increasingly fierce competition from both large and small vendors.

## Strengths

- Execution of Watson Studio strategy: Watson Studio is fulfilling the original promise shown by the product in its early days when known as Data Science Experience. The best features from SPSS (e.g., canvassing and model management) have been finely integrated into the product. IBM has maintained strong fundamentals around data management and information architecture and flexibility with the Watson Studio portfolio.
- Product deployment and business value: Reference customers reported high satisfaction with IBM's service and support. Watson Studio is well-established in enterprise use cases that are visible and deliver value for organizations. IBM also made the choice to be one of the most supportive large vendors of hybrid and multicloud DSML. IBM's Cloud Pak for Data on Red

Hat OpenShift is designed to manage data engineering and MLOps flexibly across multiple deployment types and data sources.

- Collaboration and appeal to multiple levels of expertise: Watson Studio has several new and modified collaboration features, including catalogs, asset galleries, asset lineage and permission management. Watson Studio and its supporting portfolio offer strong user experiences for different analytics roles and levels of data science expertise.

### Cautions

- Product bundling and configuration: The number of IBM products needed for end-to-end data science is growing, and the portfolio's development rests with multiple product teams. Customers report challenges navigating multiple products, while exchanges between products are not seamless. The licensing costs of these tools also remain a concern.
- Ongoing need for innovation after catching up in key areas: IBM has established a strong beachhead in modern data science with Watson Studio. However, its recent improvements will need additional innovation to stay relevant. For example, within Watson Studio, AutoAI is a significant stride in augmented analytics, but much of this functionality is set to become "table stakes" soon and other vendors are debuting innovations.
- Brand restoration and keeping up with stiff competition: Watson Studio is indicative of the direction of modern data science platforms, but Watson is not yet a top brand in modern ML. Revamping Watson Studio has taken time and IBM's competitors have been busy. IBM's challenge is to restore strength to the Watson brand in advanced ML circles and keep pace with its traditional competitors, large cloud vendors and a growing group of smaller, agile vendors.

### KNIME

Based in Zurich, Switzerland, KNIME provides an open-source platform called KNIME Analytics Platform. In addition, the commercial extension, KNIME Server, provides advanced functions that include collaboration, automation and operationalization capabilities.

An active end-user community is fully integrated with KNIME Hub, which provides an open, searchable marketplace of blueprints and best practices for KNIME workflows, nodes, components and extensions. KNIME conducts two user conferences annually — one in Europe and one in the U.S. — as well as many meetups and "learnathons" across the world. A large global partnership network provides customer support for KNIME. Support is also available through the KNIME community forum (in English). KNIME Server commercial customers have a full support service contract.

KNIME is positioned as a Visionary, moving out of the Leaders quadrant from last year largely due to its lower visibility and slow revenue growth relative to other vendors evaluated. However, it continues to have a strong focus on innovation, an excellent product and deep connections within the data science community.

### Strengths

- Open-source with commercial support for enterprise-grade capabilities: Combined with the commercial KNIME Server, the open-source KNIME Analytics Platform is a robust, comprehensive and end-to-end DSML solution. The breadth of capabilities provided is enhanced by the contributions of the open-source community, and this lends itself to quick and easy scaling. The KNIME Server facilitates orchestration across the end-to-end analytic process, including a simplified operationalization of models.
- Enablement and support for multiple skill levels: KNIME's traditional data science capabilities continue to be extended not only with its open approach, but also with the ability to create and share — through the KNIME Hub — reusable components with multiple types of users. Capabilities are available for coding and leveraging APIs for programming-centric developers, expert data science model development using a workflow for seasoned data scientists, and augmented capabilities for citizen data scientists.
- Mature augmented analytics vision: KNIME's vision for augmented analytics is well-defined, providing flexibility between a more guided or automated approach, depending on the user's needs. Continued focus on the abstraction of data science — through additional augmented capabilities and support of components that can be shared — further enhances the augmented vision.

### Cautions

- Subdued visibility and slow relative growth: KNIME has less visibility and slow growth in the market compared to other entrants. Although there are thousands of users of the open-source version of KNIME, far fewer have upgraded to the commercial platform. As such, many are not able to take advantage of the end-to-end capabilities provided by the full offering. There is still significant opportunity in both nascent and maturing data science teams for KNIME to grow.
- Training shortcomings: Although both self-training and in-person training are available for KNIME, training users — especially those with less data science experience — is challenging. In addition, difficulty in onboarding and training can compromise the acceptance and use of the platform across the enterprise.
- Performance issues: When working with large datasets, long workflows or complex use of multiple extensions, which sometimes require significant data movement or conversions, performance can be an issue. Although the vendor has recently made some efforts to address certain performance issues (such as big data and streaming nodes), customers still indicate performance as an area of concern.

### MathWorks

MathWorks is headquartered in Natick, Massachusetts, U.S. Its two major products are MATLAB and Simulink, but only MATLAB met the inclusion criteria for this Magic Quadrant (Simulink is considered part of the supporting portfolio).

The company benefits from a large community of users, thanks to its longevity and recognition among the engineering and scientific community. Referenced users praised the

support they receive from MathWorks and often find themselves self-sufficient when it comes to implementing the technology; MathWorks users' peer community participation is among the highest of any vendor in the Magic Quadrant.

MathWorks is positioned as a Leader. Its move from last year's position in the Visionaries quadrant is a tribute to the vendor's adaptability and engagement with cutting-edge technologies and long-standing Ability to Execute. As the demand for ML and AI techniques mounts within asset-centric domains, MathWorks is well-positioned to capitalize on that demand. Along with the integration of an increasing number of ML techniques, MathWorks tackles new business issues for a community eager to quickly productize its ML efforts.

### Strengths

- **Production-focused environment:** Aligned with the group of industries that it serves, MathWorks offers a carefully engineered environment; from data preprocessing and models development (through a fully open environment) to production, including the ability to simulate the system to be deployed, with the possibility to automatically generate code. The code can then be embedded in smart assets, edge or any transactional systems, on-premises, in the cloud or through hybrid platforms. From inception, every model on the company's platform is destined to production.
- **Platform coherence and integration:** MathWorks offers a fully integrated platform where every new technique (e.g., reinforcement learning, transfer learning) is seamlessly integrated within its environment. Throughout the preparation, development, simulation, deployment and production steps, MATLAB provides a polished experience.
- **Prebuilt solutions:** Through the numerous implementations of its users in solving particular problems, MathWorks has packaged the experience and the domain expertise in turnkey applications such as predictive maintenance.

### Cautions

- **MLOps:** Given MATLAB's proficiency in getting models into production, the next logical step is to provide users with a means to monitor, manage and govern those models. MLOps capabilities will also need to handle highly distributed models. These are capabilities MathWorks needs to significantly enhance in its upcoming version.
- **Center for data science and project collaboration:** Notebooks model development is a must for most engineers and data scientists. However, if MathWorks wants to widen its user scope and improve the collaboration between subject matter experts or citizen engineers and the model builders, it will have to provide an easier communication medium than notebooks and provide analytics workflow (canvas) capabilities.
- **Interpretable AI:** As the production of ML models spreads across its users, an increasing number of model consumers (and regulators) will start pressuring the ML builders for better interpretability of their models. Complex engineering systems are particularly exposed to this demand, and MATLAB's new deployments using deep learning and reinforcement learning

techniques will amplify that need. This will require MathWorks to significantly enhance its interpretability capabilities.

## Microsoft

Microsoft is based in Redmond, Washington, U.S. Its core product considered for this Magic Quadrant is Azure Machine Learning (Azure ML). The supporting portfolio of products for Azure ML includes Azure Machine Learning Studio, Azure Data Factory, Azure HDInsight, Azure Databricks, Power BI and other components. For on-premises workloads, Microsoft offers Machine Learning Server.

Microsoft is a sponsor of the Global AI Community and has achieved significant growth in its global bootcamps. The Azure AI community page offers broad forums and tutorials for users of various backgrounds and skills. Microsoft also has ongoing partnerships with hardware providers and various elements of the OSS community. The vendor hosts a number of large user conferences throughout the year, including Microsoft Ignite, Microsoft Inspire and Microsoft Build.

Microsoft's co-selling, industry, deployment and consulting partners number in the thousands. Microsoft Unified Support offers several support options for Azure ML and its portfolio.

Microsoft remains a Visionary after a strong year, with advances in Completeness of Vision and Ability to Execute. However, its Ability to Execute is limited by a cloud-first approach and nagging coherence issues. Microsoft is among the few vendors in this Magic Quadrant that cater to the full spectrum of data science talent.

## Strengths

- Appeal and mature functionality for multiple skill levels and analytics personas: Microsoft is achieving its aim of appealing to users along a broad spectrum of sophistication and requirements. Azure ML has plenty to offer citizen data scientists in the form of drag and drop and augmented analytics, without losing its identity as a platform for experts. The larger portfolio offers products and functionality that will draw in developers (the larger Azure stack), data engineers (Azure Data Factory, Azure Databricks) and analysts (Power BI).
- Pace of development: Customers appreciate Microsoft's ability to update and improve the cloud product at a rapid cadence. Users feel heard when updates are released, and their experience is better for it. Expert users especially enjoy always having new pieces and the latest from the open-source community to incorporate into their work.
- Compute power and control: Microsoft's scores for performance and scalability continue to be among the highest out of all vendors evaluated. Azure ML boasts stand-out features for cost control and visibility. Numerous clients use Azure ML and its supporting portfolio to manage agile, large-scale and mature data science operations.

## Cautions

- On-premises, hybrid and multicloud are not first-class citizens: The vast majority of Azure ML customers are deployed in pure cloud environments. Many capabilities within the Azure ML portfolio become weaker or more complicated in hybrid, multicloud or on-premises environments. Much like Google and Amazon Web Services (AWS), multicloud support is limited.
- Coherence: Although the Azure ecosystem offers diverse tools and approaches for data science, many users find the number of components overwhelming and are frustrated by the overall experience. Data science teams in hybrid environments face licensing and total cost of ownership concerns.
- Solid but not leading augmented DSML: Microsoft continues to invest to keep pace with the augmented DSML capabilities of smaller, solely data-science-focused vendors. New features like augmented data preprocessing and feature engineering are steps in the right direction, but augmented DSML is one of the fastest-moving areas of development in this space. Microsoft's long-term vision for augmented analytics is sound, but in the near term other vendors primarily focused in this area are winning new accounts.

## RapidMiner

RapidMiner is based in Boston, Massachusetts, U.S. RapidMiner Studio is the vendor's primary model development tool and is available as a free edition, as well as a commercial edition. RapidMiner Server is an enterprise extension designed for deploying and maintaining models and facilitating collaboration. RapidMiner's supporting portfolio also includes RapidMiner Real-Time Scoring and RapidMiner Radoop. RapidMiner Turbo Prep, RapidMiner Auto Model and RapidMiner Automated Model Ops are augmented features of the platform, while the RapidMiner AI Cloud offers cloud-based deployment options. RapidMiner's massive user community is active and bolstered by the vendor's enablement efforts, including skills certifications. The vendor hosts one of the largest data science communities and conducts two user conferences annually in the U.S. and Germany. RapidMiner offers 24/7 enterprise support for paying customers with SLAs based on issue severity. It also offers a Center of Excellence program that includes white glove service. RapidMiner is positioned as a Visionary, moving from the Leaders quadrant last year largely due to its slower growth relative to other vendors in this Magic Quadrant. It maintains a strong focus on innovation and product excellence, which will serve the vendor well as it fights for market share in a crowded and competitive space. RapidMiner's new mission statement and new go-to-market strategy targeted at enterprise clients may help accelerate growth.

## Strengths

- End-to-end augmented data science life cycle management: RapidMiner makes it easy for expert data scientists and budding citizen data scientists to manage their end-to-end data science pipeline, from inception to model building to production. The vendor specifically offers a certification program through its RapidMiner Academy for nondata scientists to

understand the product, model development and operationalization. Turbo Prep, Auto Model and Model Ops form a compelling vision of end-to-end augmented data science.

- Flexibility, agility and modern scalability: RapidMiner strongly supports a wide range of technologies (e.g., R, Python, Scala, Java, MATLAB, Octave, HiveQL, Pig, SQL and Groovy). For deep learning, Keras, TensorFlow, Eclipse Deeplearning4j and Theano are integrated, and all algorithms from the Weka library are available as part of the RapidMiner platform. RapidMiner also allows containerization with Docker and Kubernetes to transparently run and scale models.
- Model governance: With its new Automated Model Ops solution, RapidMiner enables compliance and auditing of models both on-premises and in the cloud. This governance covers not only data lineage, but also model training, model usage, model retraining, model versioning and model access.

### Cautions

- Slow relative growth and increased competition: RapidMiner has slow relative growth compared to other vendors with comparable value propositions, especially in the last two years. This is partially the result of a business strategy prioritizing profitability over growth. While thousands of users have downloaded RapidMiner, far fewer have upgraded to the commercial platform. RapidMiner remains a foundational brand in data science, but will need increased revenue and market share to maintain product development and relevance.
- Sales innovation: RapidMiner's strategy for sales and expansion is different from most competitors, with revenue growing at approximately the rate of overall market growth. Its sales team and average deal size are small compared to other similarly sized vendors. To grow faster, RapidMiner needs a new approach to convert the many young data science professionals that try the product into paying enterprise customers.
- Licensing and service customizations: Prospective customers find it difficult to navigate through the pricing model of RapidMiner. Total cost of ownership may be difficult to calculate. Customers would also like to see more customized service offerings.

### SAS

SAS is based in Cary, North Carolina, U.S. It offers a variety of software products for analytics and data science supporting statistics, ML, text analytics, forecasting, time series analysis, econometrics and optimization. SAS Visual Data Mining and Machine Learning (VDMML) was the core product evaluated for this Magic Quadrant. Aimed at business analysts and citizen/expert data scientists, VDMML incorporates multiple products including Visual Analytics and Visual Statistics.

SAS provides a dedicated resource center for analytics that offers a series of webinars, events, fact sheets, webcasts, white papers, and more. The SAS Resource Center acts as both a library and a support community for users looking for feature information, tips and tricks and case studies in analytics. SAS hosts a SAS Global Forum along with various customer connection

events. It offers 24/7 standard technical support in addition to SAS Premium Support and SAS Elite Support.

SAS is again positioned as a Leader this year. Its DSML products have a high degree of enterprise readiness and consistently deliver high business value to customers. While customer adoption of open-source alternatives to the SAS platform remains a competitive threat, the vendor retains a strong and adaptive presence in the market. SAS's Ability to Execute continues to be impacted by high license costs, which cause existing and prospective customers to explore other options. It recently launched life cycle product bundles called Unified Insights to reduce licensing complexity.

### Strengths

- **Recognized and trusted brand:** SAS's long-standing market presence and trusted brand have earned much customer respect. Customers choose SAS for its enterprise-grade platform capabilities and support for the entire analytics life cycle — from exploration to modeling and deployment.
- **Model ops:** SAS offers one of the best model operationalization and management platforms. It includes performance monitoring for model decay, automated retraining of models when thresholds are exceeded, governance via a centralized model repository with templates and version control, and lineage for both SAS and open-source models.
- **Ease of use and augmentation:** SAS VDMML's drag-and-drop interface appeals to citizen data scientists wanting ease of use, while also supporting traditional data scientists who prefer to code. The product provides automated suggestions for data quality and preparation, and suggests visualizations based on variable distribution and other measures.

### Cautions

- **Pricing and sales execution:** SAS's pricing remains a concern for customers who frequently investigate less-costly alternatives. Open-source tools are often used along with SAS products as a way of controlling costs, especially for new projects. SAS has introduced more flexible pricing, such as bursting and unlimited capacity, which is more customer-friendly.
- **Coherence:** SAS's full complement of products are still complex and often confusing to customers. Some use cases require licensing additional products such as SAS Model Manager for model management or SAS/ACCESS to connect to different data sources. SAS recently announced life cycle product bundles to help address licensing complexity.
- **Marketing strategy:** SAS has made great strides in modernizing its core platform with SAS Viya, together with offering easier to learn/use products like VDMML. But to win over traditional data scientists, the vendor will need to counter the perceived lack of support for open-source languages and tools. SAS also needs to increase marketing of SAS Viya as a proven enterprise-grade platform to operationalize and scale ML initiatives.

### TIBCO Software



TIBCO Software is based in Palo Alto, California, U.S. In recent years, TIBCO has built a well-rounded and powerful analytics platform (TIBCO Data Science) through the acquisition of enterprise reporting and modern BI platform vendors (Jaspersoft and Spotfire); descriptive and predictive analytics platform vendors (Insightful, Statistica and Alpine Data); a streaming analytics vendor (StreamBase Systems); a metadata management capability (Orchestra Networks); and an in-memory data platform (SnappyData).

The company nurtures a diversified user community that comes together in various forums, including TIBCO Exchange, where users can share and reuse analytical assets. TIBCO is also active in many open-source efforts through a significant amount of contributions. Reference customers spoke highly about TIBCO support, both for implementation and ongoing support. TIBCO Software is once again a Leader in this Magic Quadrant as it continues to integrate the functionalities from its broad portfolio. TIBCO is simplifying and streamlining TIBCO Data Science while keeping the platform open and supportive of the fast moving ML landscape. TIBCO's origins in the integration middleware space give the vendor an edge with data engineering and deployment and drive its success in a range of use cases across many industries.

## Strengths

- Edge analytics and streaming: TIBCO continues to stand out in IoT — from edge integration using TIBCO Data Science, to TIBCO Streaming and the open-source Project Flogo, to execution of deep learning models on-device. TIBCO is still one of just a few ML vendors leading the production and execution of ML on the edge, especially for asset-centric organizations.
- Openness: TIBCO received strong scores for the flexibility and openness of its platform. For example, TIBCO Data Science can integrate proprietary developed models using a wide range of open-source capabilities while managing these models within its workflow environment.
- End-to-end platform: TIBCO's end-to-end experience is built on two foundations that many of its competitors lack. First, from the data management and preparation process end of the spectrum, TIBCO benefits from its BI and visualization heritage from Spotfire. At the other end of the spectrum, in deployment and production, TIBCO can leverage its expertise from its application integration platform technology (e.g., TIBCO BusinessWorks, Project Flogo and TIBCO Cloud Mashery).

## Cautions

- MLOps: Despite its lead in deployment and production, TIBCO still has some work to do when it comes to model operationalization management. As competitive differentiation and market demand shift toward the ability to productize ML models, TIBCO MLOps — currently in its early version — will have to catch up with the leading edge in its next incarnation.

- **Interpretable AI:** TIBCO needs to continue investing in interpretable AI. As ML models find their way into production applications, business users and engineers require greater transparency to interpret insights delivered by black boxes. TIBCO will have to focus part of its upcoming development environment to keep pace with the market on explainable AI features.
- **Augmented ML:** TIBCO has shown progress in its adoption of augmented ML capabilities, especially through visualization techniques and entry-level capabilities. However, as the market moves quickly on this set of features, the vendor will have to consolidate its offering to stay on par with the market.

## Vendors Added and Dropped

We review and adjust our inclusion and exclusion criteria for Magic Quadrants as markets change. As a result of these adjustments, the mix of vendors in any Magic Quadrant may change over time. A vendor's appearance in a Magic Quadrant one year and not the next does not necessarily indicate that we have changed our opinion of that vendor. It may be a reflection of a change in the market and, therefore, changed evaluation criteria, or of a change of focus by that vendor.

### Added

- Altair

### Dropped

- SAP, which has launched a new DSML offering.
- Datawatch (Angoss), which was acquired by Altair in December 2018.

## Inclusion and Exclusion Criteria

We made some changes to the inclusion criteria for this edition of the Magic Quadrant. The inclusion process included requirements for vendors to meet a revenue threshold and identify reference customers. A stack ranking process assessed how well products support the typical use-case scenarios for DSML, namely:

- **Business exploration:** This is the classic scenario of exploring the unknown and forming hypotheses that requires extensive data preparation, exploration and visualization capabilities. This scenario could also include the incorporation of augmented capabilities to guide the data preparation, use of visualization and analysis.
- **Advanced prototyping:** This scenario describes when data science, especially novel ML techniques, is used to significantly improve traditional analytic approaches. Traditional approaches can be the use of human judgment, exact solutions, long-established heuristic approaches or legacy data mining approaches. Advanced prototyping utilizes some or all of the following:

- Many more data sources
- Novel analytic approaches (such as deep neural nets, transfer learning and reinforcement learning)
- Large-scale computing infrastructure
- Specialized computer science and ML skills
- Production refinement: In this scenario, the organization has numerous data science solutions implemented and delivered into the business, but the focus now shifts to making refinements, improvements and updates to the existing models. The activities and functionality of MLOps also fall primarily within the production refinement use case, wherein the platform supports the release activation, monitoring, performance tracking, management, reuse, maintenance and governance of models.
- Augmented data science and machine learning: This use case has been renamed from “nontraditional data science.” This is the scenario where ML/AI is used to automate and accelerate key aspects of data science, such as feature engineering and model selection, as well as model operationalization, model explanation, model tuning and management. This reduces the requirement for specialized skills to generate, operationalize and manage an advanced analytics model. It opens up DSML content creation to citizen data scientists and (application) developers. Highly skilled data scientists can also be more productive and have more time to focus on other tasks.

We used the following 15 critical capabilities to score the vendors’ capabilities across the four use-case scenarios:

- Data access: How well does the product support data access across many types of data (such as tables, images, graphs, logs, time series, audio and texts)?
- Data preparation: Does the product have a significant array of noncoding or coding data preparation features?
- Data exploration and visualization: Does the product allow for a range of exploratory steps, including interactive visualization?
- Automation and augmentation: Does the product facilitate the automation of feature generation, algorithm selection, hyperparameter tuning and other key data science tasks?
- User interface (UI): Does the product have a coherent “look and feel” and have an intuitive interface, ideally one supporting a visual pipelining component or visual composition framework (VCF)?
- Machine learning (ML): How broad are the ML approaches that are easily accessible and shipped (prepackaged) with the product, along with support for modern ML approaches like ensemble techniques (boosting, bagging and random forests) and modern dimension reduction schemes?
- Other advanced analytics: How are other methods from the fields of statistics, optimization, simulation, and text and image analytics integrated into the development environment?

- Flexibility, extensibility and openness: How can various open-source libraries be integrated into the platform? How can users create their own functions? How does the platform work with notebooks?
- Performance and scalability: How can desktop, server and cloud deployments be controlled? How are multicore and multinode configurations utilized?
- Delivery: How well does the platform support the ability to create APIs, or container (for example, code, Predictive Model Markup Language [PMML], packaged apps) that can be utilized for faster deployment into various business scenarios?
- Platform and project management: What management capabilities does the platform provide (such as for security, compute resource management, governance, project or experiment organization, auditing lineage and reproducibility)?
- Model management: What capabilities does the platform provide to monitor and recalibrate hundreds or thousands of models? This includes model testing capabilities such as K-fold cross-validation, train-validation-test-splits, area under the curve (AUC), receiver operating characteristic (ROC), loss matrices, as well as testing models side by side (for example, champion-challenger [A/B] testing).
- Precanned solutions: Does the platform offer “precanned” solutions (for example, cross-selling, social network analysis, fraud detection, recommender systems, propensity to buy, failure prediction and anomaly detection) that can be integrated and imported via libraries, marketplaces or galleries?
- Collaboration: How can users of various skills work together on the same workflows and projects? How can projects be archived, commented on and reused later?
- Coherence: How intuitive, consistent and well-integrated is the platform to support an entire data analytics pipeline? The platform itself must provide metadata and integration capabilities to take the preceding 14 capabilities and provide a seamless end-to-end experience, making data scientists more productive across the whole data analytics pipeline. This metacapability includes ensuring data input/output formats are standardized wherever possible, so that components have a similar and consistent look and feel, and ensuring unified terminology across the platform.

The subcriteria aligned with each critical capability have been reviewed and modified to realign subcapabilities with the appropriate overall capabilities and to reflect new developments and key subcapabilities that differentiate solutions.

The weighting of customer experience has been lowered as Gartner methodology has reduced the number of customer references required from vendors for surveying and interviews. The lower weight reflects the smaller sample size and reduced pool of interviewees.

As mentioned above, we have renamed the “nontraditional data science” case to “augmented data science and machine learning.” This is to align with other Gartner research on the theme of augmented analytics and to resolve confusion over the meaning of nontraditional data science.

The weighting for the four use cases remained the same as in last year’s Magic Quadrant.

To qualify for inclusion in this Magic Quadrant, each vendor had to pass the following assessment inclusion criteria.

## Inclusion Criterion No. 1: Data Science and Machine Learning Platform Offering

Vendors' DSML platforms needed to:

- Offer a mixture of basic and advanced functionality essential for building DSML solutions (primarily predictive and prescriptive models)
- Support the incorporation of these solutions into business processes, surrounding infrastructure, products and applications
- Support variously skilled data scientists in all of the following tasks across the data and analytics pipeline:
  - Data ingestion
  - Data preparation
  - Data exploration
  - Feature engineering
  - Model creation and training
  - Model testing
  - Deployment
  - Monitoring
  - Maintenance
  - Collaboration

Vendors also needed to be able to provide technical support for their DSML platform directly and/or via commercial support partners.

## Inclusion Criterion No. 2: Revenue and Number of Paying Customers

Three common license models were assessed, and revenue (and/or customer adoption) from each were combined (if applicable) and evaluated against the criteria below for each core product under consideration:

- Perpetual license model: Software license, maintenance and upgrade revenue (excluding revenue from hardware and professional services) for either 2018 calendar year or fiscal year.
- SaaS subscription model: Annual contract value (ACV) for either 2018 calendar year or fiscal year, excluding any professional services included in annual contracts. For multiyear contracts, only the contract value for the first 12 months was used for this calculation.
- Customer adoption: The number of active paying client organizations using the vendor's DSML platform (excluding trials).

Each vendor's core product needed to have:

- At least \$75 million combined perpetual license revenue and ACV in either 2018 calendar year or fiscal year, *or*
  - At least \$10 million combined perpetual license revenue and ACV in either 2018 calendar year or fiscal year *and* at least 19% combined revenue growth when compared with the immediately previous calendar or fiscal year, *or*
  - At least \$5 million combined perpetual license revenue and ACV in either 2018 calendar year or fiscal year *and either*
  - At least 38% combined revenue growth when compared with the immediately previous calendar or fiscal year, *or*
  - 50 new corporate customers gained during 2018 calendar or fiscal year
- Only vendors with core products that passed this inclusion criterion were considered for the third inclusion criterion.

### Inclusion Criterion No. 3: Customer Counts

Vendors that satisfied the requirements of inclusion criterion No. 1 were next evaluated on the basis of the reference customers they identified. Vendors had to show significant cross-industry and cross-geographic traction for each core product under consideration. Counts included only active unique customer organizations using the latest version of the core product or a version released in the 12 months prior to August 2019.

#### *Cross-Industry Reference Customers*

Each vendor had to identify active unique customer organizations using each of the DSML platforms under consideration in production. For a core product to be considered, 10 unique organizations were required. These had to have data science solutions in production and come from at least four of the following major industry segments:

- Banking and securities
- Communications, media and services
- Education
- Government
- Healthcare
- Insurance
- Manufacturing and natural resources
- Retail
- Transportation
- Utilities
- Wholesale trade

#### *Cross-Region Customer Count*

Among the reference customers for each vendor, there had to be at least two active customer organizations in each of the three following major geographic regions:

- North America
- European Union (including the U.K.) and Switzerland
- Rest of the world

Only vendors that passed inclusion criterion No. 2 progressed to inclusion criterion No. 3.

## Inclusion Criterion No. 4: Product Capability Scoring

Vendors that successfully passed the first three criteria were then assessed by Gartner analysts measuring how well their product(s) met the 15 critical capabilities.

Because the number of vendors that can be included in a Magic Quadrant is limited, only vendor products with the highest 16 to 18 critical capability scores continued to the detailed evaluation phase.

If two or three vendors' products tied, we included each of them, bringing the maximum number of vendors to 18. If more than three had platforms that tied, we would have used a metric incorporating internet search, Gartner search and Gartner client inquiry data to determine which vendors' products had greater market traction and break the tie on that basis. In no case would more than 18 vendors appear in the Magic Quadrant.

Approximately 70 vendors were considered for inclusion. Sixteen vendors were selected for final inclusion.

## Exclusion Criteria

Vendors whose DSML platforms are primarily marketed to and used by application developers or business analysts do not meet the inclusion criteria for this Magic Quadrant.

## Honorable Mentions

The following list includes notable vendors that either did not meet the inclusion criteria or whose eligibility for inclusion we were unable to verify due to a lack of information:

- Amazon Web Services (AWS), which provides Amazon SageMaker (aimed primarily at developers and data scientists) and a robust portfolio of supporting ML products and components (including Amazon SageMaker Studio, an integrated development environment for machine learning, announced at AWS re:Invent in December 2019).
- Cloudera, which provides Cloudera Machine Learning (the next generation of Cloudera Data Science Workbench) for data science teams to accelerate, manage and scale ML workflows, and for enterprise IT to secure and govern ML workflows.
- FICO, which specializes in decision management and is an especially good fit for teams building and operationalizing ML models for financial services use cases.

- Iguazio, whose Data Science Platform enables enterprises to develop, deploy and manage AI applications at scale and in real time.
- Oracle, which provides Oracle Machine Learning and a broad suite of data and analytics products that support data preparation, visualization, augmented analytics, model development and deployment, and other stages of the data science life cycle, with SQL, R and Python APIs.
- SAP, which has revamped its DSML platform (SAP Data Intelligence) with a focus on enterprise readiness, data management and governance, and integration with SAP's numerous data, analytics and AI solutions.
- Teradata, which offers Teradata Vantage, facilitates a unified view of enterprise data and empowers users to perform predictive and prescriptive analytics as well as autonomous decision making and machine learning at scale.
- World Programming, which provides WPS Analytics, a flexible platform that supports analytics development, governance and deployment with visual and programming tools for Python, R, SQL and SAS.

## Evaluation Criteria

### Ability to Execute

Product/service: Core goods and services that compete in and/or serve the defined market. This criterion assesses current product and service capabilities, quality, feature sets, skills and so on. These may be offered natively or through OEM agreements and partnerships, as defined in the market definition and detailed in the subcriteria.

Overall viability (business unit, financial, strategy and organization): This criterion includes an assessment of the organization's overall financial health, as well as the financial and practical success of the business unit. The criterion also assesses the likelihood of the organization continuing to offer and invest in the product, as well as the product's position in the current portfolio.

Sales execution/pricing: This criterion assesses the organization's capabilities in all presales activities and the structure that supports them. Included are deal management, pricing and negotiation, presale support and overall effectiveness of the sales channel.

Market responsiveness and track record: This criterion assesses a vendor's ability to respond, change direction, be flexible and achieve competitive success as opportunities develop, competitors act, customers' needs evolve and market dynamics change. It also considers a vendor's history of responsiveness to changing market demands.

Marketing execution: This criterion assesses the clarity, quality, creativity and efficacy of programs designed to deliver the organization's message in order to influence the market, promote a brand, increase awareness of products and establish a positive identification in the minds of customers. This "mind share" can be driven by a combination of publicity, promotional, thought leadership, social media, referrals and sales activities.



Customer experience: This criterion assesses products, services and/or programs that enable customers to achieve anticipated results with the products evaluated. Specifically, it considers the quality of supplier-buyer interactions, technical support and account support. Ancillary tools, customer support programs, availability of user groups and SLAs may also be evaluated, among other things.

Operations: This criterion assesses the organization’s ability to achieve its goals and fulfill its commitments. Factors considered include the quality of the organizational structure, skills, experiences, programs, systems and other vehicles that enable the organization to operate effectively and efficiently.

**Table 1: Ability to Execute Evaluation Criteria**

Enlarge Table

Evaluation Criteria	Weighting
Product or Service	High
Overall Viability	Medium
Sales Execution/Pricing	Low
Market Responsiveness/Record	Medium
Marketing Execution	Low
Customer Experience	Medium
Operations	Medium

Source: Gartner (February 2020)

## Completeness of Vision

Market understanding: This criterion assesses a vendor’s ability to understand customers’ needs and to use that understanding to create products and services. Vendors that have a

clear vision of their market, and that listen to and understand customers' demands, can shape or enhance market changes.

Marketing strategy: This criterion looks for clear, differentiated messaging that is consistently communicated internally, and externalized through social media, advertising, customer programs and positioning statements.

Sales strategy: This criterion looks for a sound strategy for selling that uses appropriate networks, including direct and indirect sales, marketing, service, and communication networks. It also considers partners that extend the scope and depth of a vendor's market reach, expertise, technologies, services and customer base.

Offering (product) strategy: This criterion looks for an approach to product development and delivery that emphasizes market differentiation, functionality, methodology and features as they map to current and future requirements.

Innovation: This criterion looks for direct, related, complementary and synergistic layouts of resources, expertise or capital for investment, consolidation, defensive or preemptive purposes.

Table 2: Completeness of Vision Evaluation Criteria

Enlarge Table

Evaluation Criteria	Weighting
Market Understanding	Medium
Marketing Strategy	Low
Sales Strategy	Low
Offering (Product) Strategy	High
Business Model	Not Rated
Vertical/Industry Strategy	Not Rated
Innovation	High

Evaluation Criteria

Weighting

Geographic Strategy

Not Rated

Source: Gartner (February 2020)

## Quadrant Descriptions

### Leaders

Leaders have a strong presence and significant mind share in the DSML market. They demonstrate strength in depth and breadth across the full data exploration, model development and operationalization process. While providing outstanding service and support, Leaders are also nimble in responding to rapidly changing market conditions. The number of expert and citizen data scientists using Leaders' platforms is significant and growing.

Leaders are in the strongest position to influence the market's growth and direction. They address the majority of industries, geographies, data domains and use cases, and therefore have a solid understanding of, and strategy for, this market. Not only can they focus on executing effectively, based on current market conditions, but they also have solid roadmaps to take advantage of new developments and advancing technologies in this rapidly transforming sector. They provide thought leadership and innovative differentiation, often disrupting the market in the process.

Leaders are suitable vendors for most organizations to evaluate. They should not be the only vendors evaluated, however, as other vendors might address an organization's unique needs more precisely. Leaders provide a benchmark of high standards to which others should be compared.

### Challengers

Challengers have an established presence, credibility, viability and robust product capabilities. They may not, however, demonstrate thought leadership and innovation to the same degree as Leaders.

There are two main types of Challenger:

- Long-established DSML vendors that succeed because of their stability, predictability and long-term customer relationships. These vendors need to revitalize their vision to stay abreast of market developments and become more broadly influential and innovative. If they simply continue doing what they have been doing, their growth and market presence may be impaired.
- Vendors established in adjacent markets — such as the analytics and BI, data and analytics service provider, and developer tool markets — that are entering the DSML market with

solutions that extend their current platforms. These vendors provide a reasonable option not only for existing customers but also for new customers. As these vendors prove they can influence this market and provide clear direction and vision, they may develop into Leaders. But they must avoid the temptation to introduce new capabilities quickly but superficially. Challengers are well-placed to succeed in this market as it is currently defined and are operating effectively within current market conditions. Their vision and roadmap, however, may be impaired by a lack of market understanding, excessive focus on short-term gains, strategy- and product-related inertia, and a lack of innovation. Equally, their marketing efforts, geographic presence and visibility may not be on a par with that of Leaders.

## Visionaries

Visionaries are often relatively small vendors or newer entrants representative of trends that are shaping, or have the potential to shape, the market. There may, however, be concerns about these vendors' ability to keep executing effectively and to scale as they grow. They are typically not well-known in this market, and therefore often don't have the momentum of Challengers and Leaders.

Visionaries not only have a strong vision, but also have a solid supporting roadmap. They are innovative in their approach to addressing the market's needs. Although their offerings are typically innovative and solid in terms of the capabilities they do provide, there are often gaps in these offerings' completeness and breadth.

Visionaries are worth considering because they may:

- Represent an opportunity to jump-start an innovative initiative
- Provide some compelling, differentiating capability that offers a competitive advantage as either a complement to, or a substitute for, existing solutions
- Be more easily influenced with regard to their product roadmap and approach

Visionaries, however, also pose a potentially riskier choice for buyers. In today's highly competitive DSML market, Visionaries may also struggle to gain momentum, develop a presence, increase their market share, fulfill their vision and execute on their roadmap. They may also be targets for acquisition.

As Visionaries mature and prove their Ability to Execute, they may eventually become Leaders.

## Niche Players

Niche Players demonstrate strength in a particular industry or approach, or pair well with a specific technology stack. They should be considered by buyers in their particular niche. Some Niche Players demonstrate a degree of vision, which suggests they could become Visionaries. Often, however, they are struggling to make their vision compelling, relative to others in the market. They are considered more followers than leaders in terms of driving and defining the market. They may also be struggling to develop a track record of innovation and thought leadership that could give them the momentum to become Visionaries.

Other Niche Players could become Challengers if they continue to execute in a way that increases their momentum and traction in the market.

## Context

The DSML market is simultaneously more vibrant and messier than ever. Vendors weave together rapidly evolving proprietary solutions with numerous open-source components and increasingly complex partnership networks. Data and analytics leaders seek to satisfy expert data scientists demanding cutting-edge capabilities, ambitious citizen data scientists seeking fast and accessible model builders, and developers tackling the immense challenge of ML engineering. Movement in this market is rapid and multidirectional. Snapshots become out of date alarmingly quickly.

The once-sleeping giants of Google and AWS are now awake and thundering across the landscape (though Amazon SageMaker was excluded from this Magic Quadrant as it was primarily marketed to and used by application developers at the time of evaluation). The traditional powers of the space, IBM and SAS, are charging into the new decade with modern offerings and adaptive strategies. Numerous smaller and midsize vendors are in sustained periods of hypergrowth. Other long-standing and admired brands in the space demonstrate exciting innovations and healthy financials.

As with last year's Magic Quadrant, vendors are heavily focused on innovation and differentiation, rather than pure execution. Innovation remains key to survival and relevance. Even with the large number of established and thought-leading vendors, countless new DSML startups with diverse products and value propositions continue to emerge. The select group in this Magic Quadrant have all established strong customer bases, financial performance and technology. Still, a new generation of vendors is striving for visibility, relevance and perhaps someday significant market share. Given the competitive and vibrant nature of this market, vendors that might have competed for placement on the left side of the quadrant have a difficult time breaking in.

Merger and acquisition activity in this market has been regular but moderate. Vendors in this Magic Quadrant will likely continue to acquire interesting companies to round out their platforms, while transformative major acquisitions (in the vein of Salesforce's acquisition of Tableau in 2019) are always possible.

Data and analytics leaders need to work hard just to keep up with this market. End-user organizations need to increase their engagement just to stay reasonably up to date. Leaders should focus on developing new use cases and applications for DSML — ones that are highly visible, deliver real business value and build momentum for future initiatives. In addition, they should look to extend access to the market's technologies to nontraditional roles and develop significant internal education programs.

Whether end-user organizations are just getting started with predictive and prescriptive analytics or have mature capabilities in advanced analytics, they must monitor changes in the market. This includes how vendors are developing their offerings and providing new

capabilities to different kinds of user and extended use cases. Organizations should start by identifying gaps in their own portfolios and monitor vendors' offerings in light of their business needs. They should first assess whether their existing analytics vendors are stepping up to the new challenges. They should consider not only DSML vendors, but also analytics and BI vendors, which are increasingly extending their capabilities to perform more advanced analytics. As DSML capabilities are increasingly adopted across enterprises, cross-departmental work is important to avoid excessive fragmentation and a lack of common standards. Otherwise, individual departments may adopt different platforms and processes — a situation that leads to operational and maintenance-related problems.

To achieve fully mature, advanced analytic capabilities, organizations must plan for and invest in the end-to-end data science life cycle. The life cycle includes processes for accessing and transforming data, conducting analysis and building analytic models, operationalizing and embedding models, managing and monitoring models over time to reassess their relevancy, and adjusting models to reflect changes in the data and business environment. Whether beginning or extending their journey in the field of DSML, organizations need not travel alone. Data and analytics service providers offer guidance, a structured approach and reduced risk of failure. Service providers also help ameliorate the common challenge of data science talent recruiting and retention (see “Market Guide for Data and Analytics Service Providers”).

## Market Overview

The DSML market is beyond healthy and thrillingly innovative. The broad mix of vendors offer a granular range of capabilities, with solutions appropriate for most levels of maturity. The definitions and parameters of data science and data scientists continue to evolve, and the space is dramatically different from this Magic Quadrant's inception in 2014.

Many vendors are now aiming for a sweet spot with their platforms to simultaneously appeal to and delight both expert data scientists and citizen data scientists. Vendors are adding more capabilities designed for data engineers, developers and ML engineers, as participation from a supporting cast in the data science life cycle becomes more common. Vendors that previously only catered to expert data scientists are adding augmented capabilities and improved interfaces to appeal to citizen data scientists. Vendors want to expand the footprint and availability of their solutions to maximize customer return on platform investments.

There remains a glut of compelling innovations and visionary roadmaps, as indicated by the positioning of many vendors to the right of the Completeness of Vision axis. Though many elements of vendors' vision and value proposition overlap, key areas of differentiation continue to emerge. These include user interface, augmented DSML (autoML), MLOps, performance and scalability, and cutting-edge use cases and techniques (e.g., deep learning, large-scale IoT, reinforcement learning).

Many organizations are starting DSML initiatives using free or low-cost open-source and public cloud service provider offerings to build up their knowledge and explore possibilities.

They are then likely to adopt commercial software to tackle broader use cases and requirements for team collaboration, and to operationalize their deployment and management of models. While enterprise data science success with a purely open-source stack is possible, the vast majority of mature and impactful data science teams have invested in a commercial platform.

Overall revenue from DSML platform software grew by 19% in 2018 (up from 17% in 2017) to represent the second-fastest-growing segment of the analytics and BI software market (behind only modern BI platforms). The segment's revenue for 2018 was \$3.2 billion (up from \$2.6 billion in 2017). Its share of the overall analytics and BI market grew from 14.1% in 2017 to 15.1% in 2018. Several of the smaller and younger vendors in this space are now sustaining hypergrowth. Growing at the rate of the market is actually a slow growth rate compared to many vendors in this Magic Quadrant. Those interested in this market should monitor and regularly assess the following developments:

- The user mix of DSML platforms grows increasingly heterogeneous. Expert data scientists remain the primary users, but the citizen data scientist is close to representing the greater source of demand and potential revenue. There is no such thing as a standard citizen data scientist. The persona includes not only business and BI analysts, but also people from the traditional data space, such as data analysts and data engineers, as well as application developers and application engineers. The ability to collaborate and share is more crucial than ever as more users — in different roles — adopt DSML platforms.
- Both the analytics and BI platform space and the DSML platform space continue to collide and influence one another. More vendors in the analytics and BI sector are offering predictive and prescriptive capabilities, often through augmented vendors. For their part, data science vendors are adding more robust data transformation and data visualization capabilities to their platforms while making their environment more hospitable to individuals without traditional data science backgrounds.
- Although new vendors are entering the market, “legacy” vendors are highly relevant. Many traditional vendors in the DSML space have firmly established new products or are revamping and modernizing their approach or expanding through strategic partnerships, mergers or acquisitions. Big names continue to offer new capabilities and approaches. At the same time, they are enabling existing customers to continue benefiting from investments they have already made and a technology stack they are used to working with.
- The open-source ecosystem and community is as vibrant as ever. Python is firmly established as the dominant language in DSML and the R community continues to grow. Open-source software enables organizations to jump-start or extend DSML initiatives with little upfront or additional investment. Additionally, the ecosystem is open to — and supported by — vendors that additionally provide commercial platforms in the DSML market.
- Algorithm building blocks are often used to create models. This trend will continue as models continue to be abstracted and packaged for specific domain and industry problems.

- Packaged models are increasingly available through APIs that can easily be integrated with, and consumed in, applications (see “Magic Quadrant for Cloud AI Developer Services”). Many cloud service APIs are highly focused on specific domain and industry problems. This approach can reduce or even eliminate the need for organizations to build models themselves.
- Data science teams have become much better at developing models than operationalizing them. As a result, business value is often not measured or realized. MLOps capabilities within data science platforms go beyond mere deployment to the ongoing management and maintenance of models in production. Whereas many models are developed, few are operationalized in a way that leads not only to deployment but to ongoing management and maintenance. In addition, models that are not properly managed and monitored are at risk of becoming irrelevant or inaccurate as business conditions change.

The DSML market’s pace of change and innovation will only continue to accelerate for the foreseeable future.

Over the past year, several trends in DSML platform offerings have gained momentum.

Modern platforms incorporate or accommodate the following:

- **Componentization:** Platforms composed of multiple components have become the norm as vendors develop their own components, use open-source software or partner with other vendors to expand their offerings. Vendors increasingly provide a heterogeneous collection of tools, as opposed to native integrations within a single product. The definition of a DSML platform has been significantly updated this year to reflect this reality.
- **Open-source acceptance:** All DSML platforms use and incorporate open-source software, although to varying degrees. Some provide APIs to access common open-source libraries. Some build open-source technologies into capabilities accessible within their own platforms. Others include the ability to use analytic artifacts created within the platform within the open-source ecosystem. Still others provide more of a wrapper for working natively with open-source tools in a consistent environment that also enables operationalization. Supporting open-source platforms and frameworks through various collaborative and orchestrated approaches has become the standard. These adaptive platforms increase support for new capabilities and increased workloads while reducing the need for users to switch platforms for different contexts. Using open-source software enables vendors to keep pace with new developments and tap into the expertise of contributors to the open-source community.
- **Multiple user types:** Organic developments in data science demand that tools address the needs of users with different skills and different levels of DSML knowledge. Some platforms are still primarily designed for a specific user type (usually expert data scientists). However, components or capabilities to enable a broad range of users — from citizen data scientists to expert data scientists to application developers — are increasingly the norm.
- **Platform coherence:** Increased componentization and open-source incorporation creates more potential for fragmented, awkward solutions. The need to access multiple components



and platforms for full, robust capabilities must be balanced against the desirability of accessing all functionality in a seamless and cohesive manner. As offerings embrace a heterogeneous environment, cohesion becomes increasingly important. As offerings expand to provide more capabilities and keep pace with emerging technologies, it is crucial that they support the ability not only to manage multiple components, but also to access them easily and seamlessly from within the platform.

- MLOps: Operationalization capabilities not only deploy, but also manage and maintain models over time. MLOps functionality is crucial for encouraging ongoing reevaluation of the relevance and validity of analysis over time as business needs, priorities and conditions change. As DSML moves out of the lab and into the mainstream, it must be operationalized with seamless integration and carefully designed architecture and processes. MLOps capabilities should also include explainability, versioning of models and business impact analysis, among others.
- Model and data repositories: There is a trend for providing a means of tracking and sharing both the data and the analytic artifacts generated as part of the model development and deployment process. This is vital for deduplication of efforts, governance and enterprise scalability of data science initiatives. It also supports the ongoing freshness of analytical assets in use and provides critical transparency into data science operations.
- Collaboration: As access to DSML platforms becomes democratized and more types of user work together across the analytic pipeline, the need to be able to collaborate easily and seamlessly increases significantly. As platforms become more accessible to new types of user, these products must enable people to work together and share in real time throughout the data science life cycle. DSML platforms are also facilitating vital collaboration between data science teams and IT, and between data scientists and line-of-business leaders.
- Extension into decision management: Increasingly, DSML platforms are extending beyond operationalization to support collaboration, which, in turn, fuels interest in decision management capabilities as analytics tools move beyond prediction to explicitly drive business decisions.

## Evidence

Gartner's assessments and commentary in this Magic Quadrant draw on the following sources:

- Instruction manuals and documentation of selected vendors. We used these to verify platform functionality.
- An online survey of vendors' reference customers, conducted from September through October 2019. This survey elicited 175 responses about the reference customers' experience with vendors' platforms. The list of survey participants derived from information supplied by the vendors.
- Phone interviews conducted with customer references provided by the vendors.

- A questionnaire completed by the vendors.
- Vendor briefings, including product demonstrations, about individual vendors' strategy and operations.
- An extensive RFP inquiring how each vendor delivers specific features that correspond to our 15 critical capabilities (see "Toolkit: RFP for Data Science and Machine Learning Platforms").
- A prepared video demonstration of how well vendors' DSML platforms address specific functionality requirements across the 15 critical capabilities.
- Interactions between Gartner analysts and Gartner clients deciding their evaluation criteria, and Gartner clients' opinions about how successfully vendors meet these criteria.

## Note 1 Definition of an Open-Source Platform

The open-source approach is becoming more common throughout the DSML platform market. It enables people to innovate collaboratively, each contributing their own perspective in a way that shortens time to market.

The open-source approach is quickly becoming a mainstream way to introduce new capabilities. Many such capabilities are evaluated in this Magic Quadrant.

The most common examples of open source in the DSML platform market are components. Open-source components include:

- Open-source programming languages such as Python and R
- Open-source libraries and frameworks such as scikit-learn and TensorFlow
- Open-source visualizations such as D3 and Plotly
- Open-source notebooks such as Jupyter and Zeppelin
- Open-source data management platforms such as Apache Spark and Hadoop

A platform is considered open — but not open source — if it offers flexibility and extensibility for accessing open-source components. In addition, a platform can itself be open source, which means that its source code is made available for use or modification.

Open-source software is usually developed as a public collaboration and made freely available. However, only open-source platforms that also have commercially licensable products were eligible for inclusion in this Magic Quadrant.

## Evaluation Criteria Definitions

### Ability to Execute

Product/Service: Core goods and services offered by the vendor for the defined market. This includes current product/service capabilities, quality, feature sets, skills and so on, whether offered natively or through OEM agreements/partnerships as defined in the market definition and detailed in the subcriteria.

**Overall Viability:** Viability includes an assessment of the overall organization's financial health, the financial and practical success of the business unit, and the likelihood that the individual business unit will continue investing in the product, will continue offering the product and will advance the state of the art within the organization's portfolio of products.

**Sales Execution/Pricing:** The vendor's capabilities in all presales activities and the structure that supports them. This includes deal management, pricing and negotiation, presales support, and the overall effectiveness of the sales channel.

**Market Responsiveness/Record:** Ability to respond, change direction, be flexible and achieve competitive success as opportunities develop, competitors act, customer needs evolve and market dynamics change. This criterion also considers the vendor's history of responsiveness.

**Marketing Execution:** The clarity, quality, creativity and efficacy of programs designed to deliver the organization's message to influence the market, promote the brand and business, increase awareness of the products, and establish a positive identification with the product/brand and organization in the minds of buyers. This "mind share" can be driven by a combination of publicity, promotional initiatives, thought leadership, word of mouth and sales activities.

**Customer Experience:** Relationships, products and services/programs that enable clients to be successful with the products evaluated. Specifically, this includes the ways customers receive technical support or account support. This can also include ancillary tools, customer support programs (and the quality thereof), availability of user groups, service-level agreements and so on.

**Operations:** The ability of the organization to meet its goals and commitments. Factors include the quality of the organizational structure, including skills, experiences, programs, systems and other vehicles that enable the organization to operate effectively and efficiently on an ongoing basis.

## Completeness of Vision

**Market Understanding:** Ability of the vendor to understand buyers' wants and needs and to translate those into products and services. Vendors that show the highest degree of vision listen to and understand buyers' wants and needs, and can shape or enhance those with their added vision.

**Marketing Strategy:** A clear, differentiated set of messages consistently communicated throughout the organization and externalized through the website, advertising, customer programs and positioning statements.

**Sales Strategy:** The strategy for selling products that uses the appropriate network of direct and indirect sales, marketing, service, and communication affiliates that extend the scope and depth of market reach, skills, expertise, technologies, services and the customer base.

**Offering (Product) Strategy:** The vendor's approach to product development and delivery that emphasizes differentiation, functionality, methodology and feature sets as they map to current and future requirements.

Business Model: The soundness and logic of the vendor's underlying business proposition.

Vertical/Industry Strategy: The vendor's strategy to direct resources, skills and offerings to meet the specific needs of individual market segments, including vertical markets.

Innovation: Direct, related, complementary and synergistic layouts of resources, expertise or capital for investment, consolidation, defensive or pre-emptive purposes.

Geographic Strategy: The vendor's strategy to direct resources, skills and offerings to meet the specific needs of geographies outside the "home" or native geography, either directly or through partners, channels and subsidiaries as appropriate for that geography and market.