Solutions Review

DATA SCIENCE & MACHINE LEARNING
BUYER'S GUIDE

MARKET OVERVIEW

The process for evaluating data science and machine learning platforms can be complex. These complexities continue to grow broader due to the convergence of product capabilities offered by traditional business intelligence vendors and their data science counterparts. Augmentation of the analytic process has been proliferating in the marketplace for several years now, making the research and discovery phase of buying data science and machine learning software a tricky one. Therefore, buyers are recommended to first begin product evaluations with a focus on current use cases and branch out from there.

Multipersona is a major trend in this marketplace. This refers to the ability of a data science and machine learning platform to serve users with different skill levels. Some vendors choose to design the user interface in a way that supports both citizen data scientists and expert data scientists, while others distinctly target non-technical or advanced users. Given that the lines between traditional BI and data science products are increasingly blurring, it remains important to attain references and proofs of concept while buying to ensure a potential tool works with your environment and talent.

In addition to multipersona, there are several other key differentiators between data science and machine learning providers, including the degree of augmentation and automation, the ability to scale, and support for hybrid and multi-cloud environments. Data science platforms also commonly include key capabilities of products in horizontal markets like data integration, preparation and exploration, and data governance.

The marketplace for data science and machine learning platforms is mature and crowded with a variety of solution providers, tools and products to meet even the most advanced needs. There are companies we refer to as 'mega-vendors' who offer enterprise-class tools while others tout cutting-edge capabilities designed for niche verticals. Still there are also data science and machine learning vendors that integrate with popular open-source technologies and feature robust user communities.

The BI and analytics marketplace is mature and crowded with excellent software tools for a variety of use cases, verticals, deployment methods, and budgets. There are very large providers we refer to as 'megavendors', like Microsoft, Tableau Software and Qlik. There are also lesser-known innovators with interesting products that play in niche areas, such as Looker, Pyramid Analytics and Yellowfin BI.

Cue the process of seeking out, evaluating, choosing, purchasing, and deploying a BI and data analytics solution. Solutions come in a variety of flavors—ranging from traditional enterprise reporting to data discovery and augmented analytic options. Each features a particular set of capabilities, strengths, and drawbacks. Choosing the right vendor and solution is a complicated process—one that requires in-depth research and often comes down to more than just the solution and its technical capabilities.

Solutions Review has developed this buyer's guide to assist buyers in search of the best possible tool to fit the needs of their organization. This resource features 10 important questions to ask during the buying process, and full, one-page vendor profiles that provide a solution overview, three key features, contact information, and our own 'Bottom Line' analysis. Companion research, including our popular vendor comparison matrix, can be found at solutionsreview.com.

Timothy King, *Senior Editor*
Solutions Review

5 Questions You Should Ask When Evaluating A Data Science & Machine Learning Solution

Who are my data science and machine learning users?

Are your users mainly non-technical business users that require self-service functionality, or experienced data scientists that require advanced tools to do cutting-edge work? The platform or product(s) you ultimately decide to go with should serve all your users as best as possible, or provide training modules to help different employees get up to speed in the shortest possible time.

Which key features are most important?

This will be largely dependent on who your users will be and what your organization already has for other data technologies within its environment. For example, many data science and machine learning tools offer capabilities commonly included in data integration and data management software. If you already have a way of preparing data for example, then you will want to consider that during vendor evaluations.

Do I require augmented data science?

You probably don't require augmented data science if this is your organization's first go at evaluating data science and machine learning software. However, if you plan to analyze large data volumes or have expert data science users running a variety of machine learning models, this may be an important consideration.

Is BI and data analytics enough?

Data science and machine learning are popular buzzwords in the technology sphere, and oftentimes organizations bite off a bit more than they can chew. It may be that an advanced analytics tool or platform will work better given you current needs. Don't buy into the hype to the point where you wind up with a tool that is too complex for your environment or users. If you don't need specialization in a specific area or for a particular use case, don't overpay for the functionality.

Do I need a unified platform?

The answer to this question will largely depend on what other data technologies you currently have in your environment. As we've noted, there is a great deal of overlap between data science and machine learning platforms and their more traditional BI counterparts. A unified platform is more of an investment and will require buy-in across various departments and users. There are also plenty of great open-source technologies in the space that can help you deal with common data science use cases.

5 Questions You Should Ask Your Potential Data Science & Machine Learning Provider

Can you provide customer recommendations?

You'll want to ask prospective vendors to provide references that can show how their software helped other customers achieve positive ROI. In-vertical references are best, so we recommend seeking first-hand accounts from similar size organizations that do business in related industries and analyze similar data sources. It's not about securing the best overall tool; it's about selecting the best solution for your specific set of circumstances.

Is the product user community-enabled?

Community support, especially among data science and machine learning vendors whom interface with open-source technologies, is common. Some of the providers listed in this report are known for offering robust user communities that can help individual users and organizations solve their problems. Community features often include training and online learning, ticketed support, user forums, and more.

What is your biggest differentiator?

In a mature software market like this one, you want to find out which capabilities or packages potential partners are most likely to market. As we noted, the hyper-specialization in this space makes it so that a potential partner may match up with your use cases and environment perfectly. Beyond learning about the key features, you should aim to understand a vendor's go-to-market strategy.

Can your data science software scale with my business?

Your current data sources, KPIs, business questions, and overall situation may very well change. It would be foolhardy to select a platform based only on how it can serve you at the current time. Very little remains static, so what may seem like the best possible solution now could be virtually useless in two, three or five years down the road. Weigh the vendor's current capabilities with its ability to scale as your organization grows.

Is there a support package included?

The rise in self-service has been a difference maker for many organizations. However, this requires solution providers to offer support services to keep non-technical users on their feet. You'll want to find out what types of services are offered with the purchase of a license. Popular support add-ons range from deployment assistance and dedicated customer service representatives to user training, certification, and even community forums.

Solution Provider Profiles

ALTAIR

Altair (formerly Datawatch) offers a suite of solutions through its Knowledge Works portfolio and is headlined by an advanced data mining and predictive analytics workbench called Knowledge Studio. The product features patented Decision Trees, Strategy Trees, and a workflow and wizard-driven graphical user interface. It also includes capabilities for data preparation tasks, visual data profiling, advanced predictive modeling, and in-database analytics. Users can import and export using common languages like R and Python, as well as data types like SAS, RDBMS, CSV, Excel, and SPSS.

Altair
www.altair.com

Key Features

No-Code Machine Learning Modeling

Altair features an interactive and intuitive interface that enables users to connect to a wide range of data sources, transform disparate data formats into usable datasets, and generate insights by using a range of modeling techniques and algorithms.

Explainable AI

Altair details how a model is configured and what the model's output means via Explainable AI. In conjunction with AutoML, this approach means all users of a model's output can be confident in the decision-making process.

Predictive to Prescriptive Analytics

Knowledge Studio offers patented Strategy Trees that combine your knowledge of the business with predictive modeling results. Users can also optimize the strategies and campaigns by scoring the impact of controllable change to multiple scenarios.

Bottom Line

Altair Knowledge Works is made up of Altair Knowledge Studio, Altair Knowledge Seeker (data visualization, segmentation, and strategy development), and Altair Knowledge Studio for Apache Spark. The solution provider is known for offering an easy-to-use platform for technical and non-technical users. Altair also touts strong features for augmented data preparation through integration with its advanced data preparation tools Altair Monarch and Altair Knowledge Hub. The provider updated its entire line of analytics products in mid-2020.

alteryx

Alteryx offers data science and machine learning functionality via a suite of software products. Headlined by Alteryx Designer which automates data preparation, data blending, reporting, predictive analytics, and data science, the self-service platform touts more than 260 drag-and-drop building blocks. Alteryx lets users see variable relationships and distributions quickly, as well as select and compare algorithm performance with ease. No coding is required while the software can be deployed in the cloud, behind your own firewall, or in a hosted environment.

Alteryx
3345 Michelson Dr
Irvine, CA
United States
+1 (888) 836 4274
www.alteryx.com

Key Features

Alteryx Server

Alteryx Server uses your organization's Gallery to identify and access relevant reports, dashboards, apps, and workflows. Server offers a flexible architecture that enables you to schedule analytic processing workflows for delivering outcomes automatically to BI dashboards, enterprise systems, data lakes, RPA systems, Excel, and documents.

Alteryx Connect

Alteryx Connect offers powerful search to find and reuse information contained in data files, databases, visualizations, dashboards, workflows, and analytic apps. Users can create a common repository for everyone to find, share, and collaborate on data and analytics.

Alteryx Promote

Alteryx Promote lets users deploy custom R and Python models directly from their preferred development environments. Promote also enables the simplified management of deploying multiple models from test to production.

Bottom Line

Alteryx offers simple connection to data sources while in-database analytic building blocks utilize the cloud to run analytics against big data sources. Integrated data preparation and data quality during model creation is a major value-add. Alteryx is unique due to its speed text mining and natural language processing, and the ability to enrich insights with location intelligence. The vendor's no-code approach is a major consideration for organizations that need tools for distinct user personas as well. Alteryx made several high-profile feature updates across its analytic product line in 2020.

ANACONDA.

Anaconda offers its data science and machine learning capabilities via a number of different product editions. Its flagship product is Anaconda Enterprise, an open-source Python and R-focused platform. The tool enables you to perform data science and machine learning on Linux, Windows, and Mac OS. Anaconda allows users to download more than 1,500 Python and R data science packages, manage libraries, dependencies, and environments, and analyze data with Dask, NumPy, pandas, and Numba. You can then visualize results generated in Anaconda with Matplotlib, Bokeh, Datashader, and Holoviews.

Anaconda
815 Brazos St
Austin, TX
United States
+1 (512) 776-1066
www.anaconda.com

Key Features

Development

Anaconda lets you work with your tool of choice so users can access Jupyter notebooks, RStudio, and popular IDEs within the Enterprise Edition. Customers can also choose from samples and preconfigured projects. AE projects are automatically containerized so they can be moved between environments easily.

Governance

The solution enables customers to maintain enterprise security and data protection by mirroring Anaconda's repository onto an organization's infrastructure. Users can also screen packages and monitor utilization through an admin dashboard, controlling user access and tracking package, project, and model deployment.

One-Click Deployment

Anaconda Enterprise Edition hastens time to production with deployment to pre-provisioned resources and takes care of load-balancing, DNS configuration, and centralized administration of deployed applications.

Bottom Line

Anaconda's product line includes Anaconda Individual Edition (open-source distribution), Commercial Edition (Commercial Package Manager), Team Edition (Package Repository) and the Enterprise option. Anaconda is popular among coders and in the financial services, energy, healthcare, manufacturing and retail industries. The product provides access to all R and Python libraries as well, and the ability to use Hadoop, Hadoop YARN and Kubernetes clusters are major value-adds.

databricks

Databricks offers a cloud and Apache Spark-based unified analytics platform that combines data engineering and data science functionality. The product leverages an array of open-source languages, and includes proprietary features for operationalization, performance, and real-time enablement on Amazon Web Services. A Data Science Workspace enables users to explore data and build models collaboratively. It also provides one-click access to preconfigured ML environments for augmented machine learning with popular frameworks.

Databricks
160 Spear St
San Francisco, CA
United States
+1 (866) 330-0121
www.databricks.com

Key Features

Collaborative Notebooks

Databricks notebooks natively support Python, R, SQL, and Scala so practitioners can work together with the languages and libraries of their choice. Users can also discover insights with build-in interactive visualizations or any library like matplotlib or ggplot. Notebooks can be scheduled to automatically run machine learning and data pipelines at scale.

Machine Learning Runtime

Databricks offers one-click access to preconfigured machine learning clusters that are powered by a scalable distribution of the most popular frameworks. The product also includes built-in AutoML and optmizations. The Machine Learning Runtime is build on top and updated with every Databricks Runtime release.

Managed MLflow

Managed MLflow is build on top of MLflow, an open-source platform developed by Databricks to help manage the complete machine learning lifecycle. It features experiment tracking with any machine learning library, model management using one central place to discover and share models, and model deployment.

Bottom Line

Databricks is a leading voice in the Apache Spark community and serves customers in a variety of industries. Its unified platform offers spans capabilities from data lake and ETL to machine learning and an enterprise cloud service. Recent enhancements include SQL Analytics, a cloud data warehouse that integrates with BI tools and enables users to query recent data in data lakes. Databricks raised $1 billion in pre-IPO funding in February to move ahead with product innovations and scale support for its new lakehouse data architecture.

dataiku

Dataiku offers an advanced analytics solution that allows organizations to create their own data tools. The company's flagship product features a team-based user interface for both data analysts and data scientists. Dataiku's unified framework for development and deployment provides immediate access to all the features needed to design data tools from scratch. Users can then apply machine learning and data science techniques to build and deploy predictive data flows.

Dataiku
902 Broadway
New York, NY
United States
www.dataiku.com

Key Features

Data Preparation

The Dataiku visual flow allows coders and non-coders to easily build data pipelines with datasets, recipes to join and transform datasets, as well as the ability to build predictive models. The visual flow also has code and reusable plugin elements for customization and advanced functions.

DataOps

Dataiku projects are the central place for all work and collaboration for users. Each Dataiku project has a visual flow, including the pipeline of datasets and recipes associated with the project. Users can view the project and associated assets (like dashboards), check the project's overall status, and view recent activity.

MLOps

The Dataiku unified deployer manages project files' movement between Dataiku design nodes and production nodes for batch and real-time scoring. Project bundles package everything a project needs from the design environment to run on the production environment. Data scientists can see all the deployed bundles as well.

Bottom Line

Dataiku is unique in that houses all data science and machine learning functionality in a single platform. Users enjoy the product's ease of use and operation, as well as features for enabling collaboration across teams. The Dataiku 7 update (released in March 2020) was highlighted by deeper integrations for technical professionals who work on machine learning project development. There is also new row-level explainability for white-box AI and Kubernetes-powered web apps that extend the product. Dataiku raised $100 million in Series D funding in August.

DataRobot

DataRobot offers an enterprise AI platform that automates the end-to-end process for building, deploying, and maintaining AI. The product is powered by open-source algorithms and can be leveraged on-prem, in the cloud or as a fully-managed AI services. DataRobot includes several independent but fully integrated tools (Paxata Data Preparation, Automated Machine Learning, Automated Time Series, MLOps, and AI applications), and each can be deployed in multiple ways to match business needs and IT requirements.

DataRobot
225 Franklin St
Boston, MA
United States
www.datarobot.com

Key Features

Data Preparation and Exploration
DataRobot features intuitive self-service data preparation via Paxata Data Prep to help users interactively explore, combine, and shape diverse datasets into ones that are ready for machine learning models and AI applications in enterprise environments. Paxata tools are deeply integrated into the platform.

Automated Machine Learning
DataRobot can automate the entire data science lifecycle and includes built-in guardrails at each step to ensure best practices are consistently followed. The solution even automates the model deployment, monitoring, and management processes for accurate predictions.

MLOps
DataRobot offers a single place to centrally deploy, monitor, manage, and govern production machine learning models no mater how they were created and where they are deployed. MLOps monitoring agents let you add monitoring in place to any existing production model already deployed. DataRobot also includes embedded governance, humility, and fairness.

Bottom Line

DataRobot's augmented data science platform can be leveraged by all user types, from AI creators like data scientists, data engineers, software developers, and business analysts, to AI operators in DevOps and IT, as well as AI consumers within lines of business. Additional product highlights include Visual AI for working with images just like any other data type, a trusted AI framework that incorporates best practices and recent research developments, and feature discovery. DataRobot can be deployed as managed, private, hybrid or on-prem.

DOMINO

Domino Data Lab offers an enterprise data science platform that allows data scientists to build and run predictive models. The product helps organizations with the development and delivery of these models via infrastructure automation and collaboration. Domino provides users access to a data science Workbench that provides open source and commercial tools for batch experiments, as well as Model Delivery so they can publish APIs and web apps or schedule reports. The company has raised more than $120 million in funding since its founding in 2013.

Domino Data Lab
548 Market St
San Francisco, CA
United States
+1 (41) 570-2425
www.dominodatalab.com

Key Features

Model Ops

Model Ops enables production model export to any infrastructure so you can see production assets at a glance. Users can turn models into on-demand REST APIs, export models as Docker images, and add CI/CD pipelines to ensure model deployment. Model Ops also allows interactive app creation, templatized analyses, and proactive monitoring.

Knowledge Center

Knowledge Center makes research reproducable and reusable while enabling governance best practices by serving as the centralized system of record for all data science work. Users can find, reuse and discuss work, see the health and status for all projects, and set goals for via custom stages.

Enterprise Infrastructure Foundation

Domino enables users to orchestrate elastic workload scaling across all the major cloud platforms and on-prem hardware. This feature also provides self-service access to compute resources and enables standards and governance across tools, languages, and teams. Users can add new tools and packages as they become available as well.

Bottom Line

The Domino Data Science Platform features an open architecture that is excellent for collaboration. The Workbench component lets users utilize the tools they want and run computationally intensive experiments simultaneously. Domino is also unique through its results and progress tracking capabilities which lets you find and reproduce past results, avoid inconsistent versions, and link Domino projects to a Jira ticket. Domino Model Monitor is available as an add-on to track data drift, prediction quality, and analyze failure conditions.

Google Cloud

Google Cloud AI offers one of the largest machine learning stacks in the space and offers an expanding list of products for a variety of use cases. The product is fully managed and offers excellent governance with interpretable models. Key features include a built-in Data Labeling Service, AutoML, model validation via AI Explanations, a What-If Tool which helps you understand model outputs, cloud model deployment with Prediction, and MLOps via the Pipeline tool.

Google Cloud
1600 Amphitheatre Pkwy
Mountain View, CA
United States
+1 (650) 253-0000
cloud.google.com

Key Features

Data Preparation

Google Cloud AI lets you prepare and store your datasets with BigQuery and Cloud Storage, then use the built-in Data Labeling Service to label training data for classification, object detection, entity extraction, and other objectives for image, video, tabular, and text data.

Model Building

The platform enables model validation with AI Explanations and What-If Tool so you can better understand model outputs, verify model behavior, identify bias, and find ways to improve your model and training data. The Vizier block-box optimization service tunes hyperparameters and optimized model performance.

Model Deployment

Customers can deploy models at scale to get predictions in the cloud with Prediction, which hosts models for online and batch prediction requests. You can also use AutoML Vision Edge to deploy models at the edge and trigger real-time actions based on local data.

Bottom Line

The Google Cloud AI platform touts an expansive set of tools including Google Cloud Data Fusion, Cloud AutoML, BigQuery ML, AI Platform Notebooks and TensorFlow. The technology mega-vendor is set to launch a unified offering with AutoML tables, XAI, AI platform pipelines and more in 2021. Google's data science and machine learning capabilities are impressive and growing, and an entrenched user base have it on pace to become a major player in the spacer in short order.

H₂O.ai

H2O.ai offers a number of AI and data science products, headlined by its commercial platform H2O Driverless AI. Driverless AI is a fully open-source, distributed in-memory machine learning platform with linear scalability. H2O supports widely used statistical and machine learning algorithms including gradient boosted machines, generalized linear models, deep learning and more. H2O has also developed AutoML functionality that automatically runs through all the algorithms to produce a leaderboard of the best models.

H2O.ai
475 Park Ave South
New York, NY
United States
+1 (650) 227-4572
www.h2o.ai

Key Features

Automatic Feature Engineering

Driverless AI automates the entire feature engineering process by detecting relevant features in a given dataset, finding the interactions within those features, handling missing values, deriving new features from data, and comparing the existing and newly generated features. Users can also see the relative importance of each feature.

Bring-Your-Own-Recipes

Data scientists can bring their own recipes or leverage the open-source recipes available by the community and curated by H2O.ai data science experts. This adds flexibility, extensibility and customization options for the platform. New data science recipes are curated by Kaggle Grand Masters at H2O.ai.

Automatic Visualization

The solution automatically generates visualizations and creates data plots that are most relevant from a statistical perspective based on the most relevant data statistics to help users gain a hastened understanding of their data prior to starting the model building process.

Bottom Line

The H2O.ai Driverless AI platform is buoyed by add-on modules in MLOps and AutoDoc. The company also offers a collection of open-source tools with commercial support options like the H2O 3 platform, AutoML for ML, Sparkling Water for Spark integration, and Wave for application development. The Bring-Your-Own-Recipes feature is a major value-add to working with this provider. End-user segment is positive surrounding all of the automation that H2O includes in its products as well.

IBM Watson Studio enables users to build, run, and manage AI models at scale across any cloud. The product is a part of IBM Cloud Pak for Data, the company's main data and AI platform. The solution lets you automate AI lifecycle management, govern and secure open-source notebooks, prepare and build models visually, deploy and run models through one-click integration, and manage and monitor models with explainable AI. IBM Watson Studio offers a flexible architecture that allows users to utilize open-source frameworks like PyTorch, TensorFlow, and scikit-learn.

IBM
1 New Orchard Rd
Armonk, NY
United States
+1 (800) 426-4968
www.ibm.com

Key Features

AutoAI

The AutoAI graphical tool in Watson Studio automatically analyzes your data and generates candidate model pipelines customized for your predictive modeling problem. These model pipelines are created iteratively as AutoAI analyzes your dataset and discovers data transformations, algorithms, and parameter settings that work best for your problem setting.

Data Refining

Use Data Refinery to cleanse and shape tabular data with a graphical flow editor. You can also use interactive templates to code operations, functions, and logical operators. When you cleanse data, you fix or remove data that is incorrect, incomplete, improperly formatted, or duplicated. When you shape data, you customize it by filtering, sorting, combining or removing columns, and performing operations.

Decision Optimization

IBM Decision Optimization for Watson Studio gives you access to IBM's solution engines for mathematical programming and constraint programming. You can build Decision Optimization for Watson Studio models either with notebooks or by using the Decision Optimization for Watson Studio model builder.

Bottom Line

IBM Watson Studio's visual workflow interface (graphic canvas) make it well suited to serve a number of different user personas. The technology mega-vendor also pays special attention to responsible and explainable AI and governance. Additional capabilities of note include data preparation, simplified modeling with the SPSS Modeler, the ability to monitor quality, fairness and drift metrics, and examining key model metrics via a side-by-side comparison.

Open for Innovation
KNIME

KNIME Analytics is an open-source platform for creating data science. It enables the creation of visual workflows via a drag-and-drop-style graphical interface that requires no coding. Users can choose from more than 2000 nodes to build workflows, model each step of analysis, control the flow of data, and ensure work is current. KNIME can blend data from any source and shape data to derive statistics, clean data, and extract and select features. The product leverages AI and machine learning and can visualize data with classic and advanced charts.

KNIME
Hardturmstr. 66
8005 Zurich
Switzerland
www.knime.com

Key Features

Tool and Data Blending

KNIME lets you blend tools from different domains with native nodes in a single workflow, including scripting in R & Python, machine learning, or connectors to Apache Spark. Data blending features are headlined by open and combine (simple text formats, unstructured data types, or time series data), database and data warehouse connection to a host, and the ability to access and retrieve data from sources like Salesforce and SharePoint.

Data Shaping

Users can derive statistics like mean, quantiles, and standard deviation, or apply statistical tests to validate hypothesis. KNIME also enables you to aggregate, sort, filter, and join data on a local machine, in-database, or in distributed big data environments. Users can then extract and select features or construct new ones to prepare datasets for machine learning.

Scale Execution with Demands

KNIME enables you to build workflow prototypes to explore various analysis approaches, as well as inspect and save intermediate results to ensure fast feedback. Users can also scale workflow performance through in-memory streaming and multi-threaded data processing.

Bottom Line

KNIME offers an advanced platform with deep data science and machine learning capabilities. It touts nearly 4,000 nodes for connecting to different data types and sources and can solve for virtually all of the major use cases in enterprise settings. KNIME's open-source nature is a major strength of the product and lets customers explore different components at will. KNIME unveiled Integrated Deployment in April, a new approach aimed at eliminating the gap between the creation of models and their use in production.

MathWorks®

MathWorks MATLAB combines a desktop environment tuned for iterative analysis and design processes with a programming language that expresses matrix and array mathematics directly. It includes the Live Editor for creating scripts that combine code, output, and formatted text in an executable notebook. MATLAB toolboxes are professionally developed, tested, and fully documented. MATLAB apps let you see how different algorithms work with your data as well.

MathWorks
1 Apple Hill Dr
Natick, MA
United States
+1 (508) 647-7000
www.mathworks.com

Key Features

MATLAB Graphics

MathWorks lets you use built-in plots to visualize your data, gain insights, and identify underlying patterns and trends. Users can explore function syntax and available chart options using integrated documentation as well. The tool also lets you choose from relevant plots presented based on the data selected so you can select the best-fit visualization type.

Algorithm Development

MATLAB lets you develop algorithms quickly and validate concepts, explore design alternatives, and distribute algorithms in the form that best suits the application. MathWorks touts thousands of core mathematical, engineering, and scientific functions, application-specific algorithms, and development tools for editing, debugging, and optimizing algorithms.

App Designer

App Designer lets you create professional apps without having to be a professional software developer. The feature includes drag-and-drop visual components to lay out design of your graphical interface. There is also an integrated editor for quickly programming its behavior. Apps can be shared using MATLAB Drive.

Bottom Line

MathWorks is best for large use cases with real-world applications. The vendor offers an impressive collection of AI features, and verifiable machine learning is a major strength of the platform. In addition to MATLAB, MathWorks also offers its Simulink product for designing and simulating systems before moving to hardware. Simulink lets you explore and implement designs that you may not otherwise consider without having to write C, C++, or HDL code.

Microsoft

The Azure Machine Learning service lets developers and data scientists build, train, and deploy machine learning models. The product features productivity for all skill levels via a code-first and drag-and-drop designer, and automated machine learning. It also features expansive MLops capabilities that integrate with existing DevOps processes. The service touts responsible machine learning so users can understand models with interpretability and fairness, as well as protect data with differential privacy and confidential computing. Azure Machine Learning supports open-source frameworks and languages like MLflow, Kubeflow, ONNX, PyTorch, TensorFlow, Python, and R.

Microsoft
One Microsoft Way
Redmond, WA
United States
+1 (425) 882-8080
www.azure.microsoft.com

Key Features

MLOps
Customers can use Microsoft's central registry to store and track data, models, and metadata. Users are enabled to automatically capture lineage and governance data as well. Azure Machine Learning also enables the use of Git to track work and GitHub to implement workflows.

Reinforcement Learning
Azure Machine Learning lets you scale reinforcement learning to powerful compute clusters, support multi-agent scenarios, and access open-source reinforcement learning algorithms, frameworks, and environments.

Responsible Machine Learning
Users can obtain model transparency at training and inferencing with interpretability capabilities. Azure Machine Learning also lets you assess model fairness through disparity metrics as a way to mitigate unfairness. There is data protection via differential privacy and confidential machine learning pipelines as well.

Bottom Line

While Microsoft's data science and machine learning portfolio is expansive and well-suited to meet the needs of enterprise organizations, multipersona support is actually a strength of the product. Microsoft offers a number of related components in addition to Azure Machine Learning, including Azure Data Factory (cloud ETL), Azure Data Catalog, Azure HDInsight (managed cluster service for open-source analytics), Azure Databricks, Azure DevOps, and Power BI, the company's flagship business intelligence and data analytics toolset.

rapidminer

RapidMiner offers a data science platform that enables people of all skill levels across the enterprise to build and operate AI solutions. The product covers the full lifecycle of the AI production process, from data exploration and data preparation to model building, model deployment, and model operations. RapidMiner provides the depth that data scientists need but simplifies AI for everyone else via a visual user interface that streamlines the process of building and understanding complex models.

RapidMiner
One Boston Pl
Boston, MA
United States
+1 (855) 472-7434
www.rapidminer.com

Key Features

Data Exploration, Preparation and Blending

RapidMiner offers a visual platform that can quickly and automatically prepare and blend data no matter the format, origin or narrative. Users can extract, join, filter, and group data across any number of sources and create repeatable data prep and ETL processes that can be scheduled for automation or shared with other users.

Business-based modeling

Once data is prepared, users can apply that data to hasten the creation of predictive models without using any code. Users can set up business constraints, such as cost and potential revenue gain, to help guide the predictions and prescribed actions. RapidMiner features a library of more than 1500 machine learning algorithms, and automated machine learning features can construct and test hundreds of models and give insight into which model might be the best fit.

Universal model operations

Once you've built a model that can deliver predictions, the RapidMiner platform can be used to quickly deploy the model to score new data in real-time. You can turn models into production web services, deploy a scoring agent, or even deploy scored data to support ad-hoc decision-making in spreadsheets or visualization tools. The model ops leaderboard lets you monitor the performance of models over time and even set up alerts and rules to turn off models.

Bottom Line

RapidMiner is a consideration for organizations seeking to utilize AI and machine learning to support and automate business decisions. The extensibility of an open-source core helps ensure that RapidMiner can be used to support nearly any use case, and a shallow learning curve makes it easy for anyone to build and deploy effective models. Those in the manufacturing vertical may find RapidMiner particularly interesting considering the considerable adoption of AI by organizations in that space.

SAS offers a suite of advanced analytics and data science products which is headlined by SAS Visual Data Mining and Machine Learning. The product provides access to data in any format and from any source, as well as automated data preparation and data lineage and model management. SAS Visual Data Mining and Machine Learning automatically generates insights for common variables across models. It also features natural language generation for creating project summaries. The companion SAS Model Manager enables users to register SAS and open-source models within projects or as standalone models.

SAS
100 SAS Campus Dr
Cary, NC
United States
+1 919-677-8000
www.sas.com

Key Features

Intelligent automation

SAS offers a public API to automate manual or complex modeling steps to build machine learning models. There is also an Automatic Feature Engineering node for automatic cleansing, transformation, and selecting features for models.

Embedded Python and R support

SAS lets you embed open-source code within an analysis and call open-source algorithms within Model Studio. The Open Source Code node in Model Studio is agnostic to Python or R versions. Model Studio also enables managing Python models in a common repository.

Analytical data preparation

The feature engineering best practice pipeline includes best transformations while distributed data management routines are offered via a visual front end. SAS also includes large-scale data exploration and summarization, cardinality profiling, and sampling.

Bottom Line

SAS Visual Data Mining and Machine Learning is offered as a part of the SS Viya product line, which also includes SAS Visual Machine Learning, SAS Visual Data Science, SAS Data Science Programming, and SAS Visual Data Decisioning. SAS is unique due to its cloud-native architecture and the ability to utilize Viya functionality in a container-based setting. The company also touts integrations with an array of open-source tools and languages for a variety of use cases.

TIBCO offers an expansive product portfolio for modern BI, descriptive and predictive analytics, and streaming analytics and data science. TIBCO Data Science lets users do data preparation, model building, deployment and monitoring. It also features AutoML, drag-and-drop workflows, and embedded Jupyter Notebooks for sharing reusable modules. Users can run workflows on TIBCO's Spotfire Analytics and leverage TensorFlow, SageMaker, Rekognition and Cognitive Services to orchestrate open source.

TIBCO
212 Elm St
Somerville, MA
United States
+1 (617) 702-1600
www.tibco.com

Key Features

Full spectrum analytics

TIBCO offers a number of analytic types, including machine learning, graph, predictive, text analytics, regression, clustering, time-series, decision trees, neural networks, data mining, multivariate statistics, statistical process control, and design of experiments are all accessible via built-in nodes.

Machine learning for big data

Automated analytical models with big data machine learning algorithms learn from data and optimize performance of the platform. TIBCO enables customers to find new patterns and insights without explicitly programming them where to look.

Security, governance and auditability

TIBCO Data Science interacts with secured clusters for advanced analytics on Hive and Spark using IT data security policies. Customers can also utilize role-based security for any asset within the system. TIBCO offers built-in version control, audit logs, and approval processes as well.

Bottom Line

TIBCO's Connected Intelligence approach is the result of integrating together a number of different data analytics products, whether developed or acquired. In addition to TIBCO Data Science, the company also offers business intelligence and analytics functionality through TIBCO Spotfire and TIBCO Streaming. TIBCO is a top consideration for organizations that deploy data and analytics across a range of functions. The company most recently released new analytics features in September 2020. TIBCO acquired Information Builders in early 2021.

ABOUT
SOLUTIONS REVIEW

Solutions Review is a collection of technology news sites that aggregates, curates, and creates the best content within leading technology categories. Solutions Review's mission is to connect buyers of enterprise technology with the best solution sellers.

Over the past four years, Solutions Review has launched ten technology buyer's guide sites in categories ranging from cybersecurity to wireless 802.11, as well as mobility management, business intelligence and data analytics, data integration, and cloud platforms.