The DataRobot machine learning automation platform captures the knowledge, experience and best practices of the world’s leading data scientists, to deliver unmatched levels of automation and ease-of-use for machine learning initiatives. DataRobot enables users of all skill levels – from business people to analysts to data scientists – to build and deploy highly-accurate machine learning models in a fraction of the time of traditional modeling methods.

“When we’re able to get rid of the repetitive and laborious tasks in a modeling process, we can free our minds to think about more high-level data science questions. We can spend our most productive hours answering the business question rather than trying to remember how to code certain models or researching how to answer certain technical questions.”

– Andrew Greenhut
Senior Data Scientist
LogMeIn

Ideal for business users
Domain expertise is the key to developing effective predictive models. DataRobot enables business users to generate accurate models quickly and efficiently. With DataRobot, business users perform sophisticated data science functions directly.

Built-in guardrails
With DataRobot, modeling projects follow a consistent methodology based on best practices so users can’t “forget” to perform a critical step, such as model validation.

Speeds model evaluation
DataRobot builds a leaderboard so you can see which models perform best with your data, and provides the tools you need to explore and compare individual models.

Advanced machine learning techniques
The DataRobot platform incorporates the techniques advanced data scientists use: boosting, bagging, random forests, kernel-based methods, generalized linear models, deep learning, and many others.

Builds the workflow for you
DataRobot creates the predictive modelling workflow for you. It knows what to do at each step of the process, and does it automatically, without prior programming or manual input from users.

Automates feature engineering
Data is prepared automatically, with DataRobot performing operations like one-hot encoding, missing imputation, text mining, and standardization to transform features for optimal results.

Leverages innovative open source engines
To harness the most advanced techniques, DataRobot uses open source machine learning libraries like R, scikit-learn, TensorFlow, Vowpal Wabbit, and XGBoost.

Supports advanced tuning
DataRobot automates model tuning, but supports manual tuning so data scientists can tune and adjust machine learning algorithms for even better results.
Notes:
1. Minimum 10TB HDFS storage across the Data Nodes is required.
2. Minimum 1TB storage required on the Edge Node.
3. The maximum dataset size supported with the above configuration is 10GB.

Built for the Enterprise
Operating at enterprise scale requires blazing performance, strict adherence to controls, and relentless focus on data protection. DataRobot is an enterprise-ready platform, delivering the governance, training, and world-class support your organization needs to get up-and-running quickly.

Use on-premise or in the cloud
On-premise: You can deploy DataRobot on-premise on standalone servers, existing Hadoop infrastructure, or in a Virtual Private Cloud (VPC).
Cloud: DataRobot offers a variety of cloud packages to meet your needs, including the DataRobot Cloud or installation on the cloud vendor of your choice.

Leverages distributed processing
DataRobot leverages modern distributed processing, running experiments in parallel to radically reduce the time needed to run a complete data science project.

Enables rapid collaboration
With DataRobot, business users, data scientists, and stakeholders work together on machine learning projects to deliver better results with less wasted effort.

Eliminates model deployment bottlenecks
There are multiple options for deploying your finished models with DataRobot, including native scoring, exportable prediction code, and prediction APIs for real-time and batch scoring.

Integrates with Hadoop
DataRobot uses your Hadoop distribution’s application management services to distribute runtime libraries to Hadoop Data nodes. Working directly with HDFS, DataRobot does not require a proprietary storage layer or moving data to an edge node. The DataRobot workload runs in YARN containers, so you do not need to partition your cluster to prevent resource conflicts.

Works with enterprise data
No matter where your data resides – relational databases, Hadoop clusters, text files or other sources – DataRobot quickly and easily connects to your data source.

Explainable models
Users can download DataRobot’s diagnostic charts, data, and documentation to share them with executives, stakeholders, and regulators.

Supports advanced security
DataRobot offers native security for fine-grained role-based authorization and supports Kerberos and LDAP protocols. In Hadoop, it works with your existing encryption policies. DataRobot meets the requirements of the CIA’s C2S cloud.

Contact Us
DataRobot | One International Place, 5th Floor | Boston, MA 02110
www.datarobot.com | info@datarobot.com