**Overview**

AI has the potential to deliver unmatched economic value, but many organizations are challenged to consistently deliver returns on their AI investments. With the huge volume and variety of data being collected and the relatively small number of data scientists who are available to implement machine learning initiatives and drive advanced analytics, organizations need to find a more efficient way to turn raw data into usable information. In fact, experts spend up to 80% of their efforts on finding, cleaning, and shaping the data, rather than getting the real insights into business problems they need to solve.

**MANAGE DATA TO VALUE END-TO-END IN ONE PLATFORM**

With the seamless integration of DataRobot Paxata self-service data preparation into the DataRobot AI platform, citizen data scientists can rapidly explore, clean, combine, and shape data for training and deploying machine learning models and production data pipelines to accelerate innovation with AI. Data science teams can collaborate, reuse, and share data sources, datasets, and recipes with full enterprise governance and security to ensure compliance with organizational policies.

**SELF-SERVICE DATA PREP FOR EVERYONE**

DataRobot Paxata offers a purpose-built application for citizen data scientists and analysts to prepare data and deploy pipelines for machine learning initiatives with no need for deep, technical, code-centric experience.

- **Intuitive, point-and-click UI.** Visually and interactively work with your full datasets in a familiar spreadsheet style view. Easily spot patterns in your data and assess data quality with visual cues.

- **Powerful data cleansing, enrichment, and shaping capabilities.** Use intuitive point-and-click actions to rapidly transform raw data, clean inconsistencies, populate missing values, and enrich features. Aggregate, deduplicate, and pivot data on the fly with no coding required.

- **Multi-user project development with reusable recipes and collaboration.** Collaborate with data science team members to jointly develop datasets, share data sources and reuse data prep steps across multiple projects.

“Joining the dots across different data sources and extracting value out of data is an important element of the Bank’s strategy as a diverse global bank. Paxata is helping us transform raw data into an insightful information fabric and this can help increase our speed of decision-making. This investment is an important step in building our capabilities as a data-driven company where employees in all functions are able to harness data at their fingertips.”

— Michael Gorriz
Group Chief Information Officer
ACCELERATION THROUGH AI ASSISTANCE

Embedded AI and algorithms constantly assist the user throughout the data preparation process to make intelligent inferences and simplify complex data structures. Our AI assistance also provides guidance and recommendations for standardizing and enriching categorical variables and features.

- **Intelligent variable standardization.** Embedded NLP-based algorithms detect similar sounding or similarly spelled words or phrases and suggest which clusters could be standardized to most commonly used values.

- **Intelligent join detection.** Automatically detect exact or fuzzy single or multi-column join criteria to combine datasets for feature enrichment.

- **Intelligent data ingest.** Automatically detect file types at ingestion time and intelligently flatten complex structures into tabular representations for learning dataset creation.

ENTERPRISE-GRADE DATA PREP

DataRobot Paxata offers tight integration with the industry’s first and most complete enterprise AI platform, empowering novice users to work on any size dataset, collaborate with data science team members and stakeholders on dataset preparation, and implement end-to-end governance to securely manage and track data transparency.

- **Powered by Apache Spark™ for big data.** Interactively view, explore, shape, and standardize variables and features on full data volumes – not just arbitrarily small data samples.

- **Intelligent automation for production data pipelines.** Automate and schedule data prep projects to run on regular intervals with automatic detection and inclusion of preceding data prep steps to create an entire data flow.

- **Full data lineage.** Automatically record data preparation steps for data lineage tracking, project steps reuse, and explainability of the entire data preparation process.

“In building this platform we had one goal: empower researchers and oncologists to spend a fraction of their time in restructuring the data, and the majority of their time on mapping patients to our research database to come up with a treatment plan.”

— Dave Parkhill
CEO, Precision Profile

Contact Us

DataRobot
225 Franklin Street, 13th Floor
Boston, MA 02110, USA

[www.datarobot.com](http://www.datarobot.com)
[info@datarobot.com](mailto:info@datarobot.com)

© 2020 DataRobot, Inc. All rights reserved. DataRobot and the DataRobot logo are trademarks of DataRobot, Inc. All other marks are trademarks or registered trademarks of their respective holders.