

# DataRobot Trusted AI

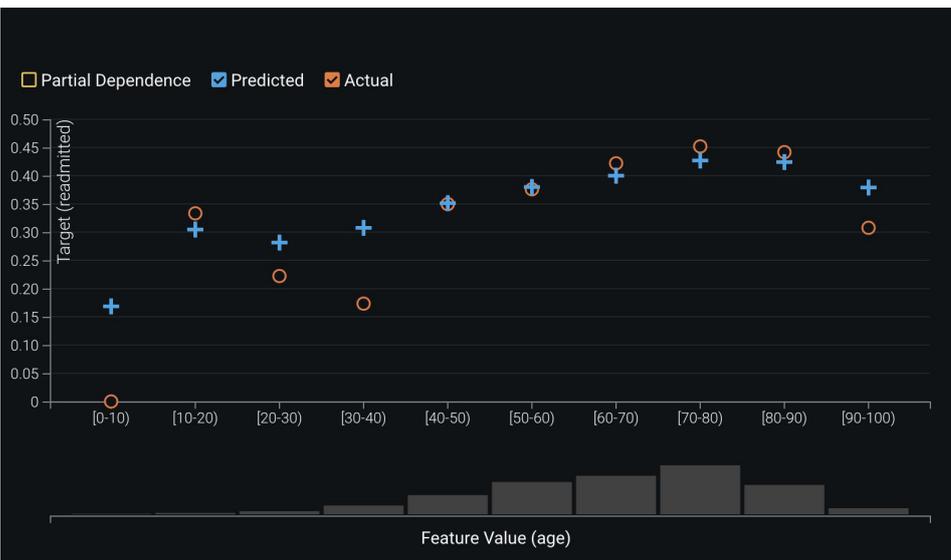
## Overview

At DataRobot, we understand that building AI you can trust is vital to your enterprise. The benefits of AI are well-known: AI systems are a powerful tool for value creation and augmented productivity and, ultimately, to better meet the needs of your customers. However, organizations have real concerns about trust, bias, and fairness related to AI systems embedded within their enterprise processes. And stories about lost public trust in AI among the biggest leaders of tech and industry are numerous.

Trust is never a given, but earned. This is the only path to trust for a company, for an individual, and for a machine learning model. At DataRobot, we're providing you built-in expertise and automated guardrails to both test your systems across multiple dimensions of trust throughout the machine learning lifecycle, and to design AI that is more robust, accurate, interpretable and stable, and therefore more trustworthy.

## What is DataRobot Trusted AI?

For DataRobot, trust in AI is established end-to-end, data to consumer. Understanding your data is the first step of model building, and the first requirement to eventually trusting your outcomes. Determining the optimal modeling approach and then having insight into what factors are driving the outcomes further bolsters your ability to ensure that the AI solution matches the needs of your process. Proper governance monitors the model in production to guarantee that it continues to perform as it is supposed to, and establishes a framework of accountability and ownership for decision-making. We should hold our AI systems to the same expectations as the greater business processes they inform--that they perform exceptionally, maintain operational excellence, and reflect our values. This is Trusted AI.



The Feature Effects panel can provide insight similar to the above. This model is performing less accurately for 30-40 year-olds. This might be a source of bias.

**42% of C-level execs, VPs and IT leaders are “very” to “extremely” concerned about AI bias.** The top two biggest concerns being “compromised brand reputation” and “loss of customer trust”

— DataRobot Report: The State of AI Bias in 2019

Executives are keen to operationalise artificial intelligence, but only 1/3 are ensuring that their data and data models are up to scratch.

— PwC Report, “AI priorities 2020”

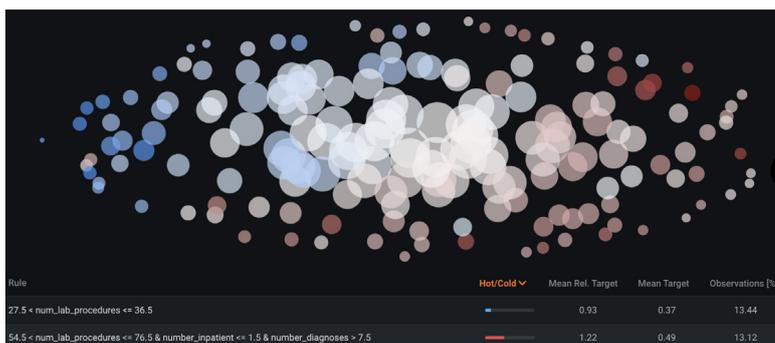


## Trust That Exceeds Your Needs

Trust in an AI system means different things to an AI creator, business operator, or an internal or external consumer. Various functionalities are required to support the ability of each persona to establish their trust in an AI system.

**AI Creators** need Trusted AI that is accurate, robust, interpretable, and tunable. At DataRobot, we support those needs through:

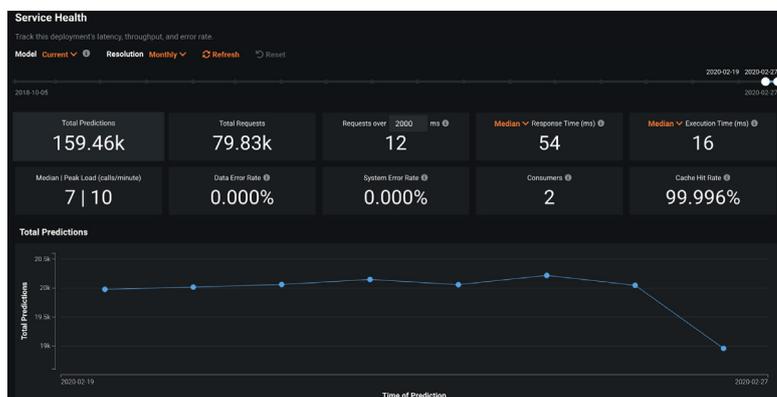
- **Data Quality Checks** - DataRobot will review the data to identify inliers, disguised missing values, an excess of zeroes and outliers upon upload and project creation, in addition to providing a project-level report of data quality with assigned severity levels to issues.
- **Breadth of Modeling Approaches** - DataRobot dynamically generates modeling approaches for a given dataset, encapsulating machine learning algorithms from classic linear or logistic models through the most advanced gradient boosted tree classifiers and deep learning approaches. It then pairs them with a variety of advanced feature engineering techniques to identify the optimal end-to-end approach.
- **Robust Target Leakage Detection** - DataRobot will identify and surface a clear warning for features at risk of target leakage.
- **Resource monitoring and reporting** - Deeper understanding into a model can be derived from tools like Feature Importance, Feature Impact and Feature Effects. You can identify which features have the strongest correlation to the target, the strongest predictive power for a particular model, and explore via partial dependence plots the relationship between a feature and the target across a feature's range of values.



HotSpots identify clusters of features that contribute to similar predictive outcomes, providing insight into possible segments of your population.

**AI Operators** require Trusted AI that is linked directly to value creation, promises operational efficiency and integrity, and stands up to auditing and regulatory inspection:

- **Ease and Flexibility of Deployment** - Deploy to a rest API endpoint, to Hadoop, or download scoring code with ease, integrating the model into the business process to begin generating value quickly.
- **Model Monitoring** - In one centralized deployment dashboard, DataRobot enables the monitoring of both DataRobot and custom models in production, including accuracy tracking and alerts on data deviation. Operators can also keep an eye on key model metrics to quickly and proactively adapt to changing conditions.
- **Approval Workflows** - Role-based permission management limits access to productionized models and facilitates the establishment of an approval architecture based on prediction consumption in order to manage organizational risk and satisfy regulatory requirements.
- **Compliance Documentation** - Automatically generate detailed, well-formatted compliance documentation, leveraging data, modeling, and interpretability visualizations and features to smooth the way through regulating entities.



The Service Health dashboard in model monitoring shows real-time data and system metrics, to ensure the system is performing operationally to your needs.

**AI Consumers (Internal & External)** want to make sure that they can rationalize the decisions made about them by an AI system, that their privacy is protected, and that they have been treated fairly:

- **Prediction Explanations** - With every prediction made, DataRobot can output the set of features with the most predictive impact on the particular outcome, including their values and how they influenced the prediction. These prediction explanations can be easily formatted into a report, lending transparency to the consumer as to the factors behind a decision.
- **Security and Privacy** - PII detection has been built into the platform to alert users when personally identifying information may have been uploaded and guard against privacy and security risks for your organization and your customers.
- **Bias and Fairness** - Interpretability tools within DataRobot help users identify potentially problematic and biased features. Model inputs are automatically examined so that a data scientist can learn how the model treats a specific factor level that may indicate bias. DataRobot University provides a course on Ethical AI to further educate and enable users to build models with awareness of bias and fairness.