
DataRobot
AI for Homeland Security
Delivering Scalable Mission-Critical Capabilities

Table of **Contents**

INTRODUCTION

Nations today face threats from complex landscapes amidst shifting priorities and changing technologies. Security at the border, while traveling, at large events, and security of our personal information has all changed in the last decade. The role of the government to enable that security mission has changed as well.

Larger and more evasive threats have called for a broader, more comprehensive approach to security. In the United States, part of that approach was the formation of the Department of Homeland Security (DHS) in 2002 whose missions and organizations are as diverse as the Coast Guard, the Cybersecurity and Infrastructure Security Agency (CISA), the Federal Emergency Management Agency (FEMA), Customs and Border Protection (CBP), and more.

Artificial intelligence (AI) can serve as a key enabler for these dedicated public servants in this ever-changing threat landscape. The Department of Justice (DOJ) also published its AI Strategy plan to realize the full potential of AI within the Department and guide its implementation in a way that fosters public trust. In 2020, DHS issued its AI Strategy to help guide AI adoption while managing the risks. The plan makes investments in emerging technologies and uses AI to assist with both policy infrastructure and mission-specific events while creating a workforce that's ready to use these tools. As DHS points out in their AI strategy, "AI presents opportunities for the Department of Homeland Security (DHS) to more effectively or efficiently accomplish our mission to secure the homeland." DataRobot is committed to helping organizations like DHS succeed in these opportunities.

Artificial intelligence (AI) is transforming the world around us. With the ability to gather data from many sources, sift through it, and provide both predictive and prescriptive recommendations in a trusted and governable manner, AI is an invaluable mission-enabling tool for security-focused organizations.

DataRobot brings years of experience supporting similar functional capabilities in the commercial and public sectors, many of which have been shared in other publications:

- Cultivating an AI-Ready Workforce
- AI for Cybersecurity
- Enhancing AI Governance
- Deploying Trustworthy AI

DataRobot webinars have hosted Homeland Security speaker events for financial management and federal executives. And DataRobot's More Intelligent Tomorrow podcast has featured discussions on important topics like human capital management, AI and the courtroom, AI ethics, and the changing role of technology on citizenship.

In this eBook, DataRobot presents additional mission-specific use cases for homeland security organizations to help turn data into value, augment humans for better decision making, and help keep citizens safe.

ARTIFICIAL **INTELLIGENCE** in Action

AI can enable organizations with homeland security missions in unexpected ways. From accounting and compliance to information technology and operations, DataRobot brings a decade of experience helping organizations fulfill their objectives faster and more effectively. Four different AI use case studies are presented below that are uniquely applicable for homeland security missions:

- Visual inspections
- Resource planning and staffing requirements
- Extreme weather events
- Employee retention

Each use case demonstrates how agencies focused on security and other government organizations can take advantage of AI. DataRobot helps organizations move beyond theoretical AI to value-added AI with a dedicated focus on the end-user: more than 1,000 customers and over 3 billion models built represent this dedication. DataRobot's Trusted AI platform and services ensure that law enforcement and homeland security professionals can continue their growth in AI adoption, while establishing guardrails for overstepping ethical limits and biases that could jeopardize the public's trust.

USING AI for

Automated Visual Inspection

Security inspections are part of modern life. From sporting events to travel and concert venues, these inspections are an important part of keeping everyone safe. For example, approximately 2.9 million passengers fly in and out of U.S. airports every day according to the Federal Aviation Administration (FAA). More than 150 million Americans attended professional sporting events in 2018. At events like these, effective and efficient screening of attendees for weapons and contraband are key to keeping individuals safe while providing a high level of service.

AI can accelerate inspections by automating and prioritizing reviews. Unlike humans at the end of a long shift, an AI's performance does not degrade over time.

In this example, the DataRobot team applied DataRobot's Visual AI and AutoML capabilities to build models capable of detecting firearms in bags using open-source databases of X-ray security scans.

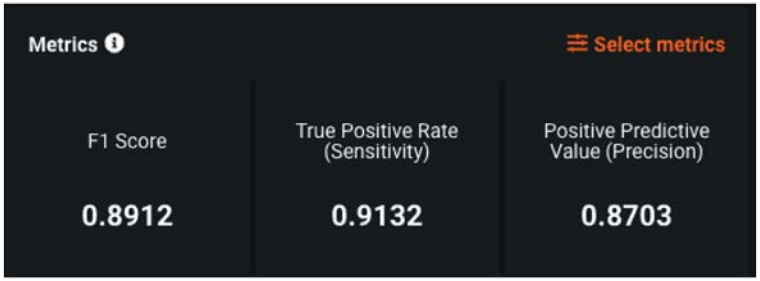
The process starts with gathering datasets and creating models. The dataset used to train the AI model in this example contained approximately 5,000 X-ray security images. Of the total dataset, approximately 30% of the images included a firearm. DataRobot can build both multilabel and multiclass predictions, meaning it can identify multiple objects in an X-ray. For this example, a binary classification was used: Does this bag contain a firearm or not?

After only a few hours, DataRobot trained and validated a model that is about 90% accurate at identifying images containing firearms.

Example of X-ray images with firearms.

There is variability in the images used to train the models because they were taken using three different types of security X-ray machines. This variability takes the form of different resolution levels and background noise in the images. Although this degrades performance, DataRobot is able to overcome this obstacle and still create high performing models by automatically applying industry best practices through modeling blueprints.

Another obstacle to creating high-performing computer vision models is that training datasets may not contain sufficient images of the target object with different backgrounds and from different directions. This data deficiency can cause the model to fail to recognize the target object (e.g., firearms) when scoring new images. DataRobot's Visual AI provides an easy way to overcome this object with automated image augmentation. This process flips, rotates, and scales images to increase the number of observations for each object in the training dataset and increases the probability that the model correctly identifies objects when scoring new records.



Training and validating a model that is about 90% accurate

Auto-generated activation maps improve explainability by illustrating which areas of an image are most important for a model's predictions—similar to feature impact on other models. DataRobot's AutoML automatically builds and compares hundreds of model blueprints to find the best performing model for identifying firearms. In this example, the winning blueprint was a neural network classifier that was built without a requirement for expensive processors like GPUs.

After only a few hours, DataRobot trained and validated a model that is about 90% accurate at identifying images containing firearms. With additional tuning, this model's performance could still be significantly increased. For example, an organization seeking to minimize false negatives (e.g., failing to identify firearms in X-rays) could change the prediction thresholds to optimize for this criterion.

DataRobot's combination of capabilities allow users to build and deploy a high-performing visual AI objection detection model in only a few hours and without the need for writing any code. This model can be quickly improved with additional advanced tuning and deployed to cloud-connected or edge environments. Applying DataRobot to this problem does not require new non-intrusive inspection (NII) machines and demonstrates how organizations can apply advanced Visual AI capabilities to existing infrastructure for rapid security improvements.

USING AI to

Predict Staffing Needs at Border Checkpoints

On any given day, nearly 500,000 passengers and pedestrians, more than 150,000 privately owned vehicles, and approximately $7.6 billion worth of imported goods cross U.S. borders. Delays at the crossing points along the border are a recurring problem. They exacerbate the supply chain problems and create undo stress for both the people looking to cross the border and the agents tasked with keeping everything secure.

With a limited number of agents, officers, and government professionals conducting operations across more than 300 ports of entry every day, wait times to enter the U.S. from Mexico can exceed 10 hours and cost upwards of $7 billion in economic activity annually.

DataRobot's AI Cloud Platform can enable effective and secure border transportation by predicting activity at crossing points to support better staffing level decisions. This use case can reduce wait times to spur economic trade, as well as ensure enough personnel are on hand to screen for illegal goods and criminal activity. For instance, every day Customs and Border Protection (CBP) arrests an average of 25 wanted criminals at ports of entry and seizes over 4,700 pounds of drugs. Having more agents in the right spot for more effective inspections can increase those seizures and help keep

America safer. AI-enabled staffing can also improve efficiency by predicting periods where activity will be low and allow CBP to reduce staffing to minimal levels without impacting safety.

The U.S. Department of Transportation (USDOT) Bureau of Transportation Statistics (BTS) provides publicly available statistics for both the U.S.-Canada and U.S.-Mexico borders at the port-of-entry level. The database contains entry data from Mexico to the U.S. for 26 years dating back to 1996. It includes pedestrian, bus, personal vehicle, rail container, train, and truck data. For this example, DataRobot is only predicting truck crossings.

Port Name	Measure	2021 January
Brownsville	Trucks	26,142
Calexico East	Trucks	34,759
Columbus	Trucks	1,226
Del Rio	Trucks	6,178
Douglas	Trucks	2,194
Eagle Pass	Trucks	16,132
El Paso	Trucks	13,008
Hidalgo	Trucks	60,500
Laredo	Trucks	200,777
Lukeville	Trucks	39
Naco	Trucks	312
Nogales	Trucks	36,630
Otay Mesa	Trucks	73,855
Presidio	Trucks	880
Progreso	Trucks	4,624
Rio Grande City	Trucks	2,977
Roma	Trucks	2,497
San Luis	Trucks	3,931
Santa Teresa	Trucks	12,816
Tecate	Trucks	4,719
Ysleta	Trucks	53,513

Example of border transport data from USDOT.

This chart shows the total truck crossings at each port of entry along the U.S.-Mexico border in January 2021. In this example, DataRobot used all 26 years of data and Automated Time Series Modeling to predict unexpected changes in the volume of truck crossings at a specific port of entry for the next month. With this information, leaders could modify staffing levels, alter lane openings and closures, and plan major repairs around surges or shortfalls in expected volume, thereby decreasing wait times and increasing trade throughput.

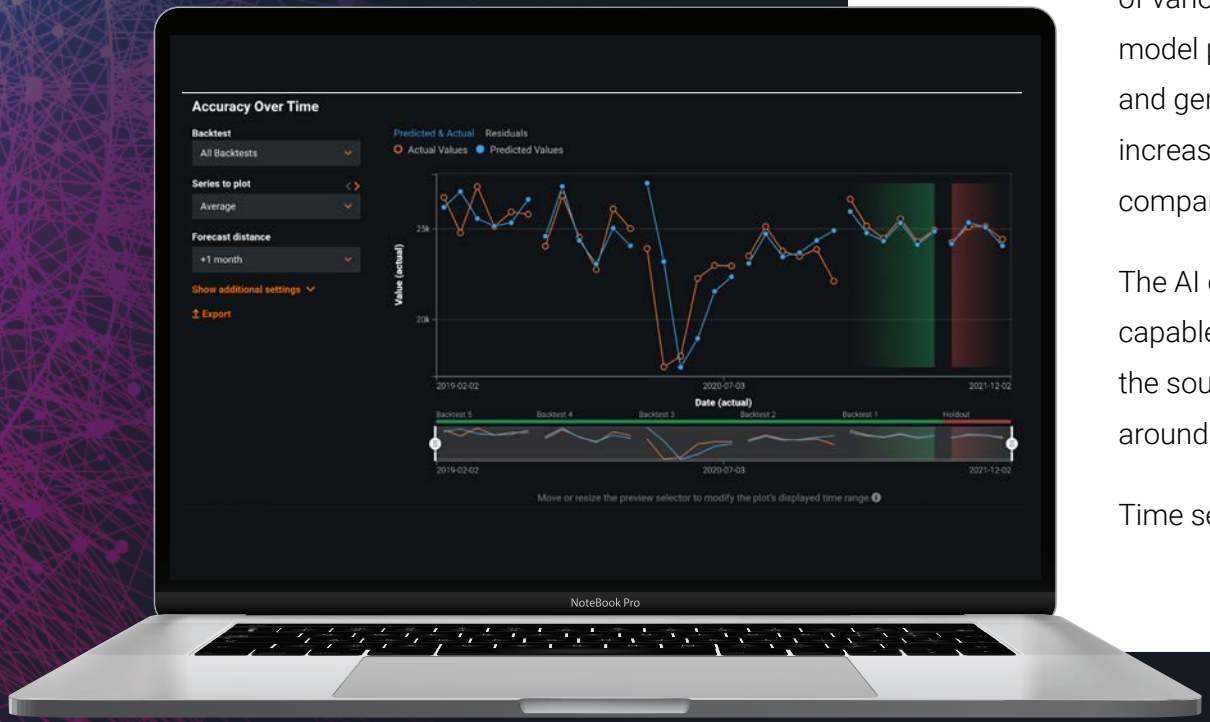
An indicator variable was created in the dataset to account for COVID-19 (known as a "regime change" in data science). For more accurate predictions, truck traffic could be aggregated at a more precise level such as hourly or daily. DataRobot model performance could also be improved by training on organizational-specific data such as border-specific events and historic staffing levels at ports of entry.

DataRobot automatically identified the ports of entries as different series in the dataset and treated them independently. DataRobot also automatically handled the complicated time series requirements like date and time partitioning while generating explainable predictions and visualizations, which increases a model's explainability and builds trust with users.

A six-month feature derivation window generated the best results for forecasting the truck volumes of the next month. DataRobot enables quick and easy iterations of various backtest configurations to rapidly find the best performing model parameters. DataRobot also took the nine original input features and generated 135 new features during automated Feature Discovery to increase the model performance. Using these new features, 63 models for comparison were automatically built.

The AI quickly produced a multiseries time series forecasting model capable of predicting surges of truck traffic at each port of entry across the southwest border. Performance of the model dropped immediately around the beginning of COVID-19, then rapidly regained accuracy.

Time series modeling can be applied to numerous use cases across



Example of model accuracy over time

DataRobot's AI Cloud Platform can enable effective and secure border transportation by predicting activity at crossing points to support better decisions about staffing levels.

Homeland Security's organizations, including staffing, demand forecasting, supply chain management, predictive maintenance, anomaly detection, and more.

This type of modeling builds forecasting models to scale across an organization's needs. Time series modeling is different from other types of machine learning and requires specialized data handling, preprocessing, and modeling capabilities. Using DataRobot's built-in automation and no-code user interface, users can easily access the full spectrum of time-based machine learning techniques.

With this information, leaders could modify staffing levels, alter lane openings and closures, and plan major repairs around surges or shortfalls in expected volume, thereby decreasing wait times and increasing trade throughput.

USING AI to

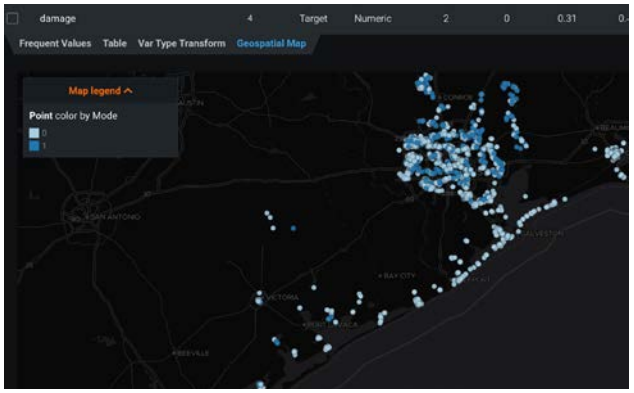
Predict Weather and Climate Events

Climate change and natural disasters are a concern for both the public sector and commercial organizations. The scale and cost of weather disasters in the U.S. is substantial and growing. From 2018 to 2020, the U.S. experienced 50 independent weather and climate disasters that cost over $1 billion each. In the past three decades, the National Oceanic and Atmospheric Administration (NOAA) estimates that climate and weather disasters have cost the U.S. over $1.875 trillion.

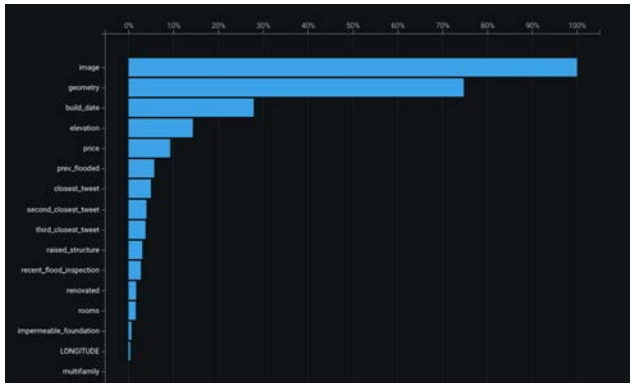
The DataRobot team has proven experience supporting weather and climate applications like identifying clean drinking water, fighting forest fires, and enabling renewable energy companies. The DataRobot AI Cloud Platform can also help identify infrastructure and buildings at risk of damage from natural disasters.

DataRobot enables the user to easily combine multiple datasets into a single training dataset for AI modeling. DataRobot also processes nearly every type of data, such as satellite imagery of buildings using DataRobot's Visual AI, the latitude and longitude of buildings using DataRobot's Location AI, tweets with geotagged locations using DataRobot's Text AI, and a variety of other details such as the home price, whether it was previously flooded, when it was built, and elevation. DataRobot can combine these datasets and data types into one training dataset used to build models for predicting whether a building will be damaged in a hurricane.

In 2017, Hurricane Harvey struck the U.S. Gulf Coast and caused approximately $125 billion in damage. Using only training data, which would have been available only before Hurricane Harvey hit the Gulf Coast, we can see an example of using DataRobot's AI to provide predictions about which structures were most vulnerable.



Example of geospatial distribution of damaged properties



Example of feature impact

DataRobot's AutoML rapidly built and compared hundreds of models using customized model blueprints. In less than a day, DataRobot produced a damage-prediction model that correctly predicted damaged properties 87% of the time and performed especially well at predicting which 30% of homes were most at risk of damage from Hurricane Harvey. DataRobot's Explainable AI features like Feature Impact inform the reader that the satellite imagery is the most important factor in determining damaged homes in the top-performing model.

With DataRobot, professionals and organizations can solve an array of difficult predictive analytics questions related to natural disasters and rapidly gain value from their data. Some additional DataRobot applications related to this include predicting:

- Fraudulent insurance claims
- Infrastructure resiliency
- Electrical grid demand
- Demand requirements for critical supplies
- Staffing requirements for emergency responders
- Outages in communications systems
- At-risk communities

USING AI for

Human Capital Management

Hiring and retention across most government organizations is a challenge. For instance, in 2019, the Government Acquisition Office (GAO) researched and identified 35 high-risk hiring areas. The Office of Personnel Management (OPM) began tracking time-to-hire (T2H) requirements in 2008 and found that the average government hire took around 100 days. External reports have found that public sector T2H averages three times as long as the private sector—119 days compared to 36 days.

Several factors contribute to this situation. Congressionally mandated reports on the government hiring process have highlighted the "unintelligible" job descriptions for many government positions. For example, a job description for a software engineer in the private sector was only one paragraph long, while the government job description was seven pages long. DataRobot can help identify qualified workers, accelerate the hiring process, and reduce turnover in a workforce with the DataRobot AI Cloud Platform already solving staffing issues at one of the largest staffing firms in the world.

Let's look at how the DataRobot AI Cloud Platform can build models to predict if an employee will depart an organization in the next six months. Accurately forecasting which employees are likely to leave, as well as what the indicators or causes for their departure might be, enables leadership to intervene early and plan proactively. The goal of attrition and retention models is usually to find a balance between retaining top talent and accepting a reasonable level of attrition. This use case predicts a binary classification target of "likely" or "unlikely" to leave in the next six months.

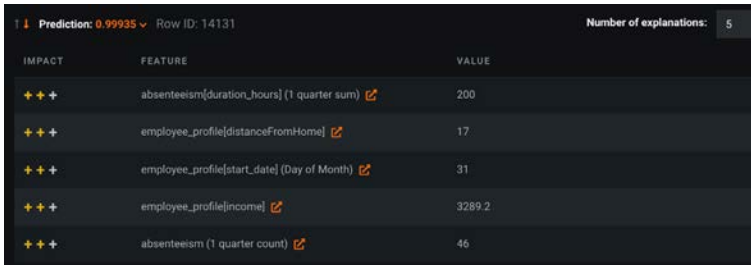
Example of Feature Lineage in DataRobot

DataRobot combined multiple datasets from different sources to create a single comprehensive training dataset for this model. We joined five secondary datasets to one primary dataset using Feature Discovery. Each dataset contained different information about an employee. For example, one dataset contained basic employee profile information while another included information about the employee's tardiness and absenteeism. The final training dataset consisted of 20,880 employee records with an approximate 70/30 split among retained and departed employees. Other than monetary bonuses, the dataset contained no information about employee performance or employee satisfaction.

DataRobot's Automated Feature Discovery used the 692 features from the original datasets to derive an additional 257 features which were used to increase predictive performance. Feature Lineage helps to visualize the steps that DataRobot took to create the new feature. The image above shows the model finding the minimum time an employee was absent during the last six months.

In only a few hours, the DataRobot platform produced a model capable of helping leaders identify at-risk employees and assisted with mitigation strategies so leaders and managers could focus attention on employees who may be at risk of leaving. A prioritized list of at-risk personnel allowed leaders and human resources teams to more accurately apply limited time and resources to prevent regrettable turnover.

DataRobot automatically provides explainable insights such as prediction explanations to help leaders understand "why" a specific prediction is made. In the image on the right, DataRobot shows that the employee in Row ID 14131 of the dataset is highly likely to depart in the next six months. The top reasons for this prediction are absenteeism, how far the workplace is from their home, when they started at the company, and their salary.



Example of DataRobot prediction explanations

Another example of DataRobot's explainable features is the word cloud, which can produce unique insights. For this example, this model's word cloud showed if an employee at this organization was absent towards the end of the month (e.g., day 28, 29, or 30), they were more likely to leave the company in the next six months.

DataRobot is an industry leader in bias and fairness testing for models. AI bias and fairness features are especially important for use cases involving humans, such as hiring and retention. DataRobot's unique bias and fairness tools test models for multiple forms of bias and enable users to perform root cause analysis to identify likely sources of bias in a dataset or in a model. This built-in capability allows

Example of word clouds in DataRobot

DataRobot is an industry leader in bias and fairness testing for models. AI bias and fairness features are especially important for use cases involving humans, such as hiring and retention.

organizations to fix issues before models are deployed into production.

The DataRobot AI Cloud platform can support human capital use cases across the government and specifically within Homeland Security organizations. Here are some other examples of using AI to address staffing needs at organizations like DHS:

- Reviewing resumes and applications for minimum experience requirements
- Predicting the best recruiting approaches to prioritize a recruiter's time
- Predicting the likelihood of staffing turnover across organizations
- Predicting unexpected requirements for additional staff support

The AI Cloud Platform for HOMELAND SECURITY

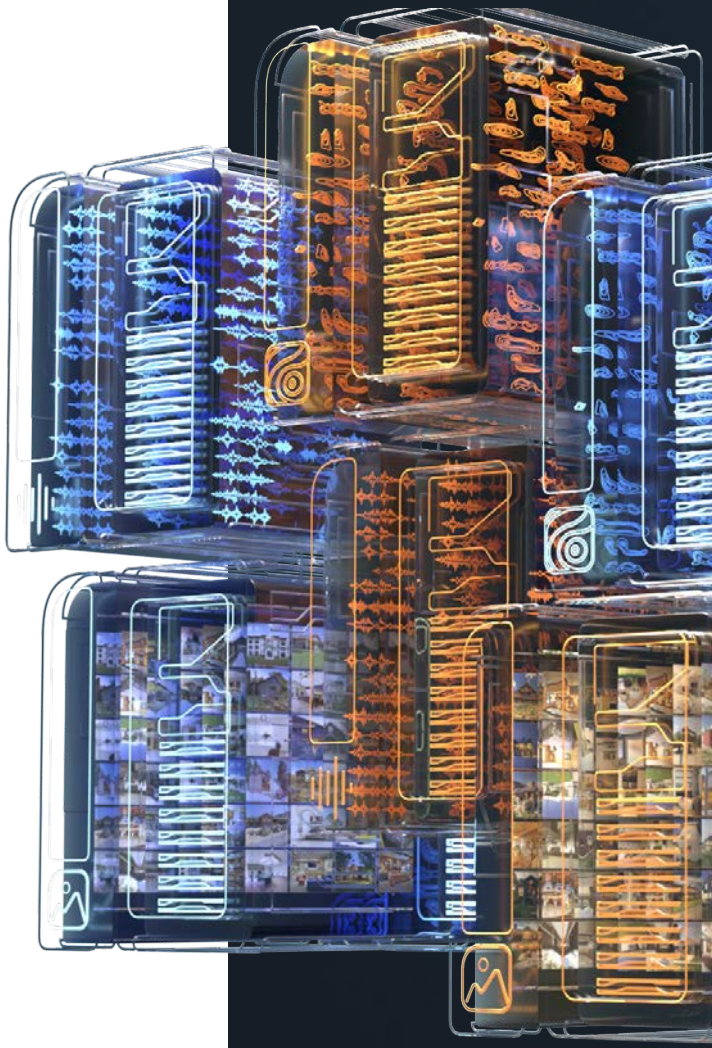
To accelerate an organization's AI maturity and gain the largest mission value, the right platform is needed. DataRobot is the global leader in trusted enterprise AI with a proven track record of delivering AI with measurable results for departments and agencies across the federal government. DataRobot maximizes value to the mission by delivering AI at scale while continuously optimizing performance over time.

- DataRobot enables both no-code data analysts and code-first data scientists to harness the predictive power of AI.
- MLOps capabilities ensure seamless model governance, data drift tracking, and model accuracy analysis and retraining to build trust with leaders and operators and adhere to responsible and ethical AI practices.
- DataRobot does not believe in black boxes; "Explainable AI" capabilities include dozens of features to ensure understanding at the model and prediction levels.
- DataRobot's team of AI success leaders and data scientists have trained and enabled more than 1,000 government professionals to turn their data into mission-oriented value.
- DataRobot seamlessly integrates with common technical tools already used in the government (Tableau, UiPath, Palantir, Snowflake, ServiceNow, and many more).
- DataRobot ingests and models from any data—text, geospatial, time series, numerical, categorical, etc.
- DataRobot helps future proof your organization with agnostic multi-cloud deployments (AWS, Azure, GPC) ensuring you always owns your data and models.

DataRobot proposes a few key considerations for homeland security professionals when selecting an AI platform. First, an AI's predictions must be trusted by users. Trust is essential in any relationship and must be earned. This is why DataRobot has invested to become the leader in delivering trusted and explainable AI. Part of that trust comes from understanding bias and fairness. DataRobot leverages five different industry standard fairness metrics, which organizations can automatically implement to check for model bias.

Second, organizations need a secure solution. Organizations processing sensitive data need vendors who support access, availability, processing integrity, data confidentiality and privacy, and physical security. The DataRobot team and platform is trusted by government officials with data at all levels of sensitivity and classification. DataRobot implements risk-based and standards-based security protocols to secure both our services and customer data. As a part of our comprehensive security program, our managed cloud service is SOC 2 Type II and ISO 27001 certified by independent third-party auditors to ensure compliance with industry standards and best practices for information security, corporate controls, and software development.

Third, organizations need to have tools that can cater to a wide audience of users, from security analysts to data scientists. With wider capabilities needed for an efficient end-to-end AI project, DataRobot is able to democratize AI by putting those AI capabilities into the hands of non-data scientists—making intelligence accessible to every area of your organization.



Finally, an AI platform for homeland security has to be scalable and able to overcome bottlenecks. From small tasks to large data mining projects, an AI platform needs the flexibility to handle mission-specific needs since the mission will not change to accommodate the software. DataRobot provides an end-to-end tool for all data and all model deployments. It provides a single place to deploy, monitor, and manage all production models, regardless of how they were created or where they are to be deployed, in a fully governed manner. DataRobot users today include more than a third of the Fortune 50 with thousands of deployed models supporting global operations.

DataRobot provides a single place to deploy, monitor, and manage all production models, regardless of how they were created or where they are to be deployed, in a fully governed manner.

CONCLUSION

In the public sector, Trusted AI built with DataRobot enables homeland security officials to make better decisions faster. DataRobot helps turn data into value and provide augmented intelligence capabilities to those who need it most. DataRobot proudly partners with governments and organizations to tackle their unique challenges and leverage AI for value they can trust.

The mission never waits, and the best time to make your organization AI enabled is now.

- What big data problems is your organization struggling with?
- What projects and processes are delayed from a lack of data understanding?
- What key indicators and goals have your organization's leaders made for information and predictive analysis?
- What value would your organization gain from AI?

Reach out to DataRobot today to discuss how AI can help enable your organization's homeland security missions and learn more about why DataRobot is the Trusted AI provider of choice for more than 1,000 organizations around the world.

DataRobot

DataRobot AI Cloud is the next generation of AI. DataRobot's AI Cloud vision is to bring together all data types, all users, and all environments to deliver critical business insights for every organization. DataRobot is trusted by global customers across industries, including a third of the Fortune 50, delivering over a trillion predictions for leading companies around the world.

Learn more at **datarobot.com**.