

80% Faster Data Categorization of Disparate Datasets

CASE STUDY

AT A GLANCE

DATA LANDSCAPE

- Disparate
- Unstructured
- Heterogeneous
- Immovable
- Sensitive
- Various Formats
- Petabytes

BEFORE

- Inefficient data categorization
- 5-6 weeks to categorize 1GB of data
- Delayed insight

AFTER

- Automated categorization
- < 4 hours to categorize 2TB of data
- Accelerated time to insight

CHALLENGE

The US Intelligence community has mountains of data they need to filter and categorize to identify insights about national intelligence targets. The data is located in different private cloud environments across the globe. It's unstructured, consisting of text, PDF, images, and videos, and cannot be moved due to its sensitive nature.

Historically, local analysts would manually sort through the data to organize it into different national security topics. This process was very laborious and time-consuming. On average, it would take one analyst six weeks to categorize 1GB of data. In total, the intelligence community has petabytes of data. Even if a thousand analysts were tasked with categorization, it would take them over a hundred years.

SOLUTION

Enter Devron—a federated machine learning platform designed specifically to unlock disparate, heterogeneous datasets such as this for advanced analytics. Because Devron offers access to data in situ, their data science team was able to automate the categorization process, deploying a supervised machine learning algorithm to each private cloud that categorized the data in a fraction of the time.

By design, Devron natively keeps the source information private, never sharing the raw data—only model learnings are sent back to the global model. This inherent privacy enabled more analysts within the intelligence community to harness the data for insights instead of the small few with high enough security clearances to see the raw data.

RESULT

With Devron, the US intelligence community categorized 2TB of data in less than 4 hours, improving their speed by nearly 80%. As a result, they accelerated their time to insight, uncovering new learnings about national intelligence targets faster.