secuvyai

# Secuvy Contextual Data Discovery & Classification

Traditional approaches to data management are ill-equipped to handle the scale and complexity of today's new data paradigm. The inability to detect and identify sensitive information about data subjects across an organization's data real estate presents legal and regulatory risks. Protecting unstructured, structured, and semi-structured data across cloud data stores, on-premises databases, and hybrid environments requires a solution that can not only find sensitive data, regardless of where it is stored and what kind of data it is but also classify that data and enable data protection by identifying and remediating privacy risks before they become more significant security issues.

## Why Secuvy?

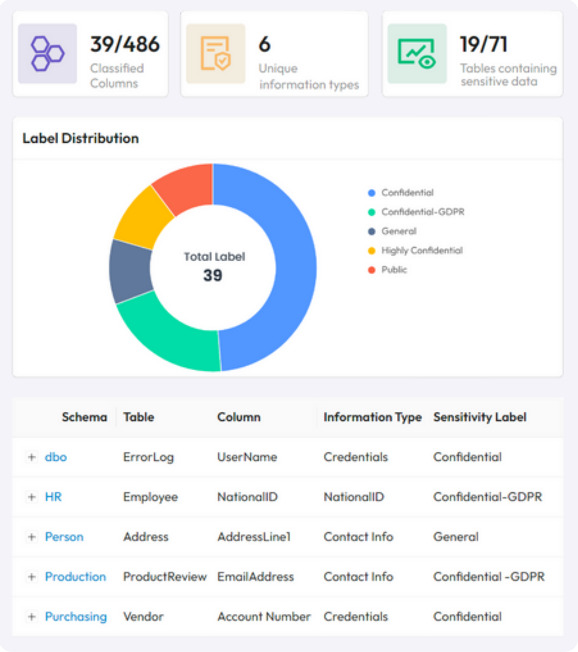
The Secuvy Platform is the first unified data protection platform for both data privacy and security built around next-generation self-learning AI to identify and classify data and determine contextual relationships, enabling robust data security and privacy compliance. Its advanced unsupervised machine learning algorithms autonomously analyze immense volumes of data and track them through their lifecycle, swiftly detecting atypical data flows and storage patterns and reducing the false negatives associated with traditional data discovery and classification. By automatically training these algorithms on historical data and providing feedback, Secuvy enables you to continuously improve accuracy and effectiveness in identifying sensitive information without needing additional headcount.

> "We were up and running with Secuvy in a matter of a few hours. Their radical approach to automate Data Privacy and Security workflows has reduced our project from months to hours."
>
> — **Privacy Program Manager**
> Fortune 1000 Company

# Secuvy Data Discovery

Legacy approaches rely heavily on manual or semi-automated processes and pattern matching to identify and protect sensitive data. However, sensitive data may not always adhere to conventional patterns, making anomaly detection a crucial aspect of classification. As a result, these methods often prove inadequate and unable to keep up with the ever-evolving threats of the digital landscape.

Supervised machine learning technologies, which rely on human training, constant monitoring, and retraining, slow the pattern recognition process as data scenarios evolve. Unless the supervised AI is explicitly based on your industry or data, the AI is useless to you.



39/486 Classified Columns

6 Unique information types

19/71 Tables containing sensitive data

Label Distribution

| Schema | Table | Column | Information Type | Sensitivity Label |
|---|---|---|---|---|
| + dbo | ErrorLog | UserName | Credentials | Confidential |
| + HR | Employee | NationalID | NationalID | Confidential-GDPR |
| + Person | Address | AddressLine1 | Contact Info | General |
| + Production | ProductReview | EmailAddress | Contact Info | Confidential -GDPR |
| + Purchasing | Vendor | Account Number | Credentials | Confidential |

By contrast, Secuvy's unsupervised algorithms excel at identifying anomalies within datasets, pinpointing irregularities that may indicate the presence of sensitive information. This ability is particularly valuable in scenarios where new types of sensitive data may emerge over time. Unsupervised machine learning algorithms also quickly self-learn as new types of threats and data breaches emerge, automatically adjusting the models. This proactive approach enables companies to stay one step ahead of cybercriminals and protect sensitive data effectively.

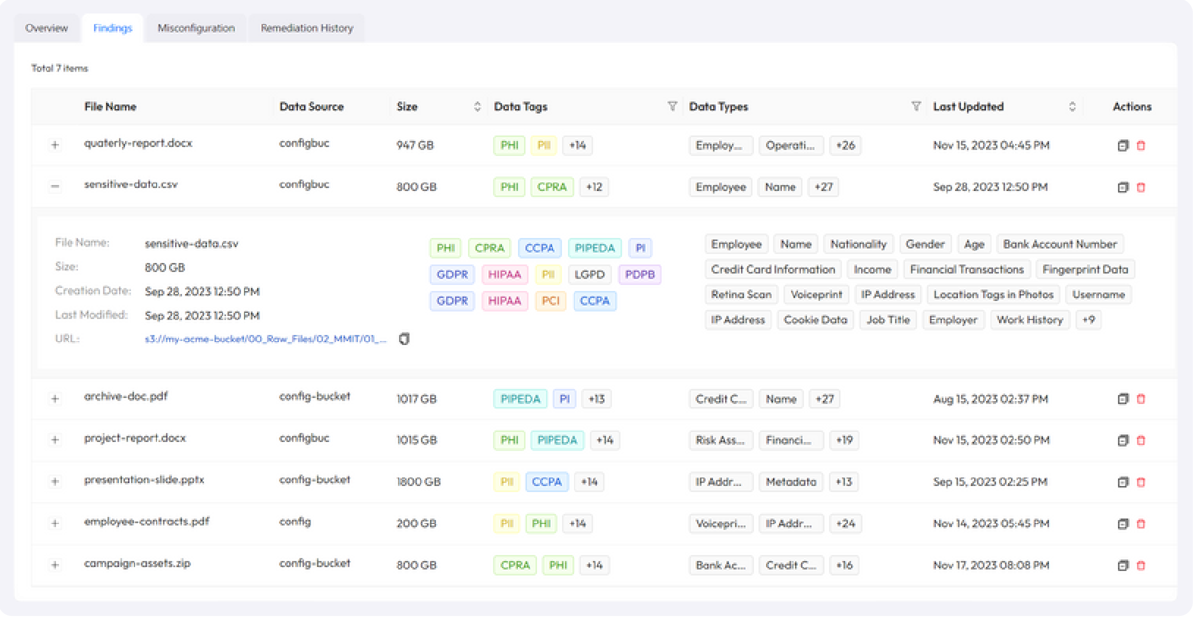
Secuvy's unsupervised machine learning stands out in scenarios where labeled training data is scarce or unavailable. Unlike supervised learning, where algorithms rely on pre-labeled datasets, unsupervised learning operates without the need for explicit guidance. This intrinsic capability is pivotal for sensitive data discovery, especially when dealing with vast datasets with diverse types of information.

Sensitive data often resides in datasets with high dimensionality, making it challenging to analyze and categorize effectively. To streamline complex datasets, unsupervised learning employs dimensionality reduction techniques, such as principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE). This not only enhances the efficiency of algorithms but also aids in identifying and classifying sensitive information hidden within the data's intricate structure.

# Secuvy Data Classification

It’s not enough to completely scan and accurately find sensitive data—context is important. This is where Secuvy shines. Secuvy’s self-learning AI enables organizations to automate the data classification process. Instead of maintaining a team of data scientists and privacy engineers to manually train the system, Secuvy’s AI understands the combination of data types and attributes within a file or data store that signals the potential for higher privacy and compliance violation risks.

For example, a data store where 30% of individual files contain a combination of names, SSNs, birthdates, and insurance claim numbers presents a higher privacy risk than a single file containing only a list of insurance claim numbers. The latter file does not contain important contextual information. This context helps the user recommend the sensitivity levels. This context is also helpful in assessing risks for privacy impact assessments, data protection impacts assessments, Records of processing activities (RoPAs) and streamlining data subject request (DSR) workflows.



Total 7 items

| | File Name | Data Source | Size | Data Tags | Data Types | Last Updated | Actions |
|---|---|---|---|---|---|---|---|
| + | quaterly-report.docx | configbuc | 947 GB | PHI PII +14 | Employ... Operati... +26 | Nov 15, 2023 04:45 PM | |
| − | sensitive-data.csv | configbuc | 800 GB | PHI CPRA +12 | Employee Name +27 | Sep 28, 2023 12:50 PM | |
| + | archive-doc.pdf | config-bucket | 1017 GB | PIPEDA PI +13 | Credit C... Name +27 | Aug 15, 2023 02:37 PM | |
| + | project-report.docx | configbuc | 1015 GB | PHI PIPEDA +14 | Risk Ass... Financi... +19 | Nov 15, 2023 02:50 PM | |
| + | presentation-slide.pptx | config-bucket | 1800 GB | PII CCPA +14 | IP Addr... Metadata +13 | Sep 15, 2023 02:25 PM | |
| + | employee-contracts.pdf | config | 200 GB | PII PHI +14 | Voicepri... IP Addr... +24 | Nov 14, 2023 05:45 PM | |
| + | campaign-assets.zip | config-bucket | 800 GB | CPRA PHI +14 | Bank Ac... Credit C... +16 | Nov 17, 2023 08:08 PM | |

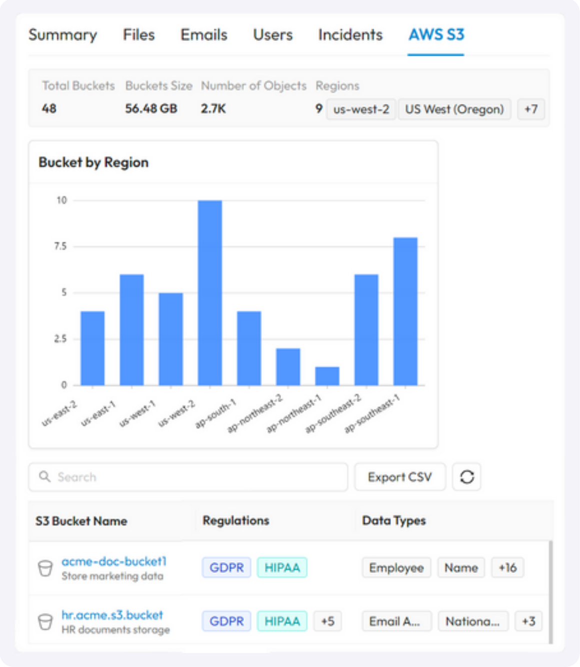
| | |
|---|---|
| File Name: | sensitive-data.csv |
| Size: | 800 GB |
| Creation Date: | Sep 28, 2023 12:50 PM |
| Last Modified: | Sep 28, 2023 12:50 PM |
| URL: | s3://my-acme-bucket/00_Raw_Files/02_MMIT/01_... |

PHI CPRA CCPA PIPEDA PI
GDPR HIPAA PII LGPD PDPB
GDPR HIPAA PCI CCPA

Employee Name Nationality Gender Age Bank Account Number
Credit Card Information Income Financial Transactions Fingerprint Data
Retina Scan Voiceprint IP Address Location Tags in Photos Username
IP Address Cookie Data Job Title Employer Work History +9

By accurately categorizing data into different sensitivity levels, companies can implement appropriate security measures based on the risk associated with each category, such as notifying the data owners and system owners, encrypting the data, moving the data to a secure location, applying retention policies, deleting the data, or applying data masking. This categorization can help companies prioritize their security efforts and more effectively allocate resources.

The ability to contextualize data, such as differentiating between DoorDash drivers, who could be both employees and consumers within the same data set, also better informs companies on how to protect that data.



| Total Buckets | Buckets Size | Number of Objects | Regions |
|---|---|---|---|
| 48 | 56.48 GB | 2.7K | 9 us-west-2 US West (Oregon) +7 |

| S3 Bucket Name | Regulations | Data Types |
|---|---|---|
| acme-doc-bucket1<br>Store marketing data | GDPR HIPAA | Employee Name +16 |
| hr.acme.s3.bucket<br>HR documents storage | GDPR HIPAA +5 | Email A... Nationa... +3 |

# Find Your Data.<br>Protect Your Data.

Identifying and remediating data risk using machine learning involves assessing potential vulnerabilities across systems, processes, and environments. Secuvy's revolutionary approach can adapt to unique data signatures in unstructured, semi-structured, and structured environments.

## With the Secuvy Platform, you can:

Discover any data type from any source, whether on-premises, in the cloud, or in a hybrid environment.

Identify and correlate data relationships based on context-based risk.

Classify data and enforce data privacy policies anywhere.

# About Secuvy

Secuvy is a cybersecurity company with a mission to automate and simplify privacy compliance through our low touch self-learning AI. Secuvy provides customers with 360° continuous visibility into all their personal and sensitive data with the greatest accuracy, unparalleled speed and the lowest cost.

## Embrace the AI Revolution in Data Privacy with Secuvy

→ **See Secuvy's AI in action—request a demo to elevate data protection and compliance.**

WWW.SECUVY.AI/DEMO

secuvyai

**Data Privacy Compliance & Automation Using the Power of AI.**

WWW.SECUVY.AI

PLEASANTON, CA (USA)

INFO@SECUVY.AI