

Securing the Unseen in the Big Data Era:

Balancing Data
Privacy with
Comprehensive
Discovery and
Classification



TABLE OF CONTENTS



Big Data and Privacy: An Evolving Paradox	03
Current Challenges in the Data Landscape	04
Big Data, Big Responsibility: Why Data Discovery and Classification Are Key to Privacy	06
Enhancing Precision with Data Discovery	06
Streamlining Operations with Data Classification	10
Leading Public Sector Bank in India Enhances Data Security – A Case Study	12
The Evolving Data Deluge: Why Smarter Classification Mechanisms Are Crucial for Payment Data Security	13

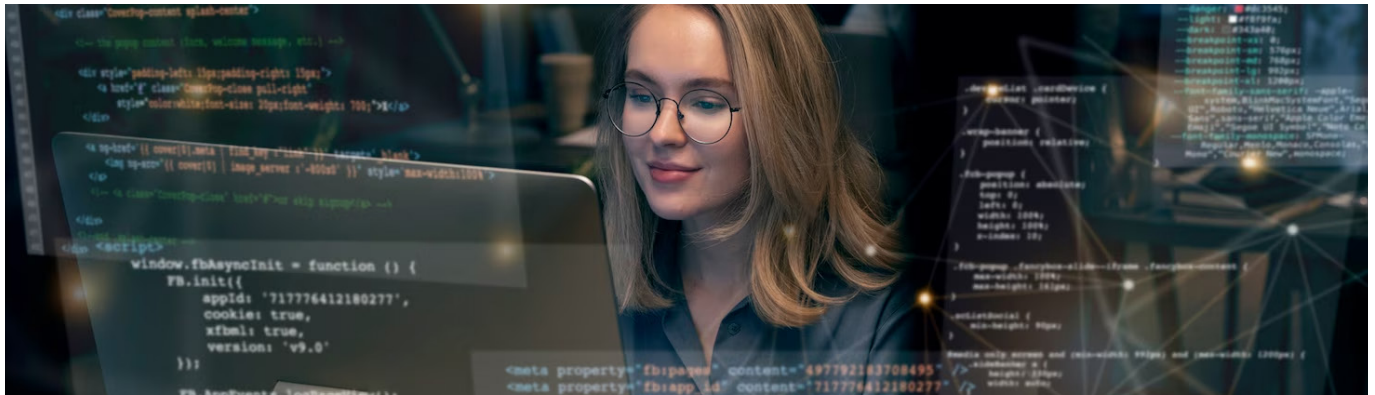
Big Data and Privacy: An Evolving Paradox

In 2023, 34.5% of executives saw cyberattacks that targeted the accounting and financial data of their companies¹. As per Forbes, the number of malware programs circulating in 2023 exceeded one billion². These statistics highlight a critical issue: each data transaction carries the potential for a privacy breach, impacting not just the organization's bottom line but also the lives of the individuals whose data it represents.

With estimates indicating that the total transaction value in the global digital payments market reached \$9 trillion in 2023, with an expected annual growth rate of 11.8 percent, reaching \$15 trillion in the next five years, the result is a vast and complex ecosystem with disparate data stores that often do not have clear organizational owners³.

Under these circumstances, it is easy to see how blind spots and nearly unchecked data proliferation can occur. Unfortunately, organizations cannot protect or manage data they do not know exists. Alarming, almost 50% of companies say they do not know where their sensitive and confidential data is stored within their systems⁴. This indicates a clarion call to weave cybersecurity more prominently into digital transformation efforts.

Current Challenges in the Data Landscape



01 Data Visibility and Complex Technology Stacks

Research shows that nearly every business (98%) reports struggles with data visibility due to complex technology stacks⁵. These environments are complex, comprising multi-cloud, SaaS, and on-premises elements. According to Statista, 81% of companies keep their sensitive data in cloud-based files such as Google Drive, 80% in email, 70% on versatile devices (PCs, laptops), 61% within chat services (e.g., MS Teams), and 50% in GitHub repositories⁶. The problem is that data is scattered within complex multi-cloud infrastructures, on-premises, or hybrid environments. Companies store data in various forms—unstructured, semi-structured, and structured and are subject to new and changing regulations, creating gaps that adversaries can exploit.

02 Unsecured Data Sources

Among companies that reported data breaches in the last two years, the most common type of breach was a database or other data source left unsecured (48%), while only 38% reported that a hacker or other outsider had maliciously accessed data⁷. Further researchers from Stanford University found that approximately 88 percent of all data breaches are caused due to human error⁸.

03 Regulatory Compliance and Penalties

Regulators are increasingly serious about cracking down on organizations that do not properly protect consumer data. As regulatory frameworks around data privacy continue to evolve and expand—such as the General Data Protection Regulation (GDPR) in Europe, the California Consumer Privacy Act (CCPA), and the Digital Personal Data Protection (DPDP) Act in India—organizations face growing complexity in compliance. The challenges of compliance with big data are further amplified by the sheer volume, velocity, and variety of data being processed. Organizations must not only track where data is stored and how it is used but also ensure that all handling practices are compliant with varying and often conflicting regulations across different regions. Each regulation may impose its own set of rules, and failure to comply can result in hefty fines and legal challenges. For instance, GDPR violations can lead to penalties of up to 4% of annual global turnover or €20 million, whichever is higher, making compliance a significant concern for global enterprises⁹.

04 Overprivileged Third-Party Integrations

Research revealed that 22% of external data shares utilize open links, so “anyone with the link” can access the data. Alarming, 94% of these open link shares are inactive, meaning that people have access to these files, folders, recordings, records, etc., even when they no longer need access. The research also shows that 100% of organizations grant full access to sensitive data (emails, files, calendars, source code) to at least one third-party integration, and one-third (33%) of integrations are granted access to sensitive permissions and data¹⁰. These SaaS-to-SaaS integrations are increasingly targeted by attackers.

The digital payments market is experiencing explosive growth, valued at a staggering **USD 92.67 billion in 2023** and projected to climb even higher with a **CAGR of 14.89% by 2030¹¹**. This rapid expansion is driven by the declining use of cash globally and the increasing adoption of digital payment solutions across various industries. Traditional card payments still hold a significant market share, especially in regions where they have a longstanding presence. However, digital payment methods such as payment apps, digital wallets, buy now pay later (BNPL) options, and account-to-account (A2A) payments are rapidly gaining traction. This shift is largely fueled by evolving customer behaviors and the demand for seamless, frictionless payment experiences embedded within financial journeys.

The transition to a largely cashless society, accelerated by the surge in contactless payments and e-commerce transactions during the COVID-19 pandemic, has placed immense pressure on merchants across all sectors to offer secure and effortless payment solutions. Industries ranging from retail and hospitality to healthcare and education are witnessing exponential data growth, each with unique dynamics and requirements. As consumers increasingly embrace digital payments, merchants and payment providers must continually innovate to meet the rising expectations for speed, security, and convenience. The ability to offer diverse payment options, minimize transaction steps, and ensure robust security will be crucial for retaining customer loyalty and staying competitive in the digital economy.

The escalating privacy concerns in the era of big data underscore the pressing need for comprehensive data discovery and classification strategies. To effectively navigate this complex landscape, organizations must implement robust data governance frameworks and leverage advanced technologies for data management. By doing so, they can safeguard privacy and maintain trust with their stakeholders.



Big Data, Big Responsibility: Why Data Discovery and Classification Are Key to Privacy

As companies dive deeper into the data-driven landscape, the responsibility to manage and safeguard this data intensifies. The cornerstone of this responsibility lies in effective data discovery and classification. These processes are crucial not only for regulatory compliance but also for maintaining customer trust and ensuring ethical data use. By implementing robust data discovery and classification practices, organizations can navigate the complexities of big data, balancing the scales between harnessing its power and upholding privacy standards.

Enhancing Precision with Data Discovery

Automated data discovery revolutionizes how organizations handle the vast amount of data generated daily. By employing intelligent tools, companies can scan a wide array of data sources—from endpoints and network servers to databases, web applications, and even storage files. These tools are adept at analyzing data stored in various formats, whether text, spreadsheets, images, audio, or PDFs, thus expanding the scope of discovery.

Moreover, these advanced tools can navigate through structured, semi-structured, and unstructured data across an organization's network and cloud environments. Integration with artificial intelligence (AI) and machine learning (ML) technologies further enhances these tools, enabling them to adapt and respond to the dynamic challenges that security teams face in monitoring ever-evolving data landscapes.

Key Components of Data Discovery

Data Exploration

- Examining dataset characteristics and properties.
- Analyzing size, format, variables, and data distribution.

Visualization

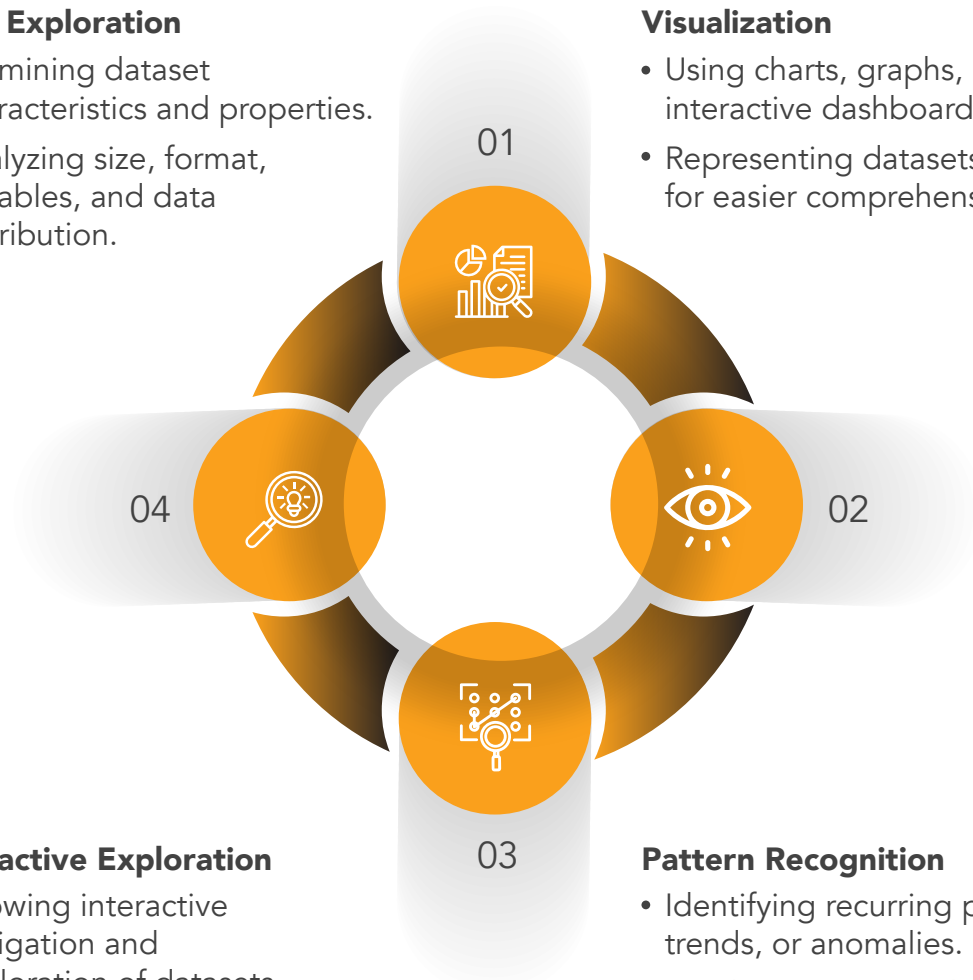
- Using charts, graphs, and interactive dashboards.
- Representing datasets visually for easier comprehension.

Interactive Exploration

- Allowing interactive navigation and exploration of datasets.
- Adjusting parameters for deeper insights.

Pattern Recognition

- Identifying recurring patterns, trends, or anomalies.
- Employing statistical methods, machine learning algorithms, and advanced analytics.
- Making predictions based on historical data.



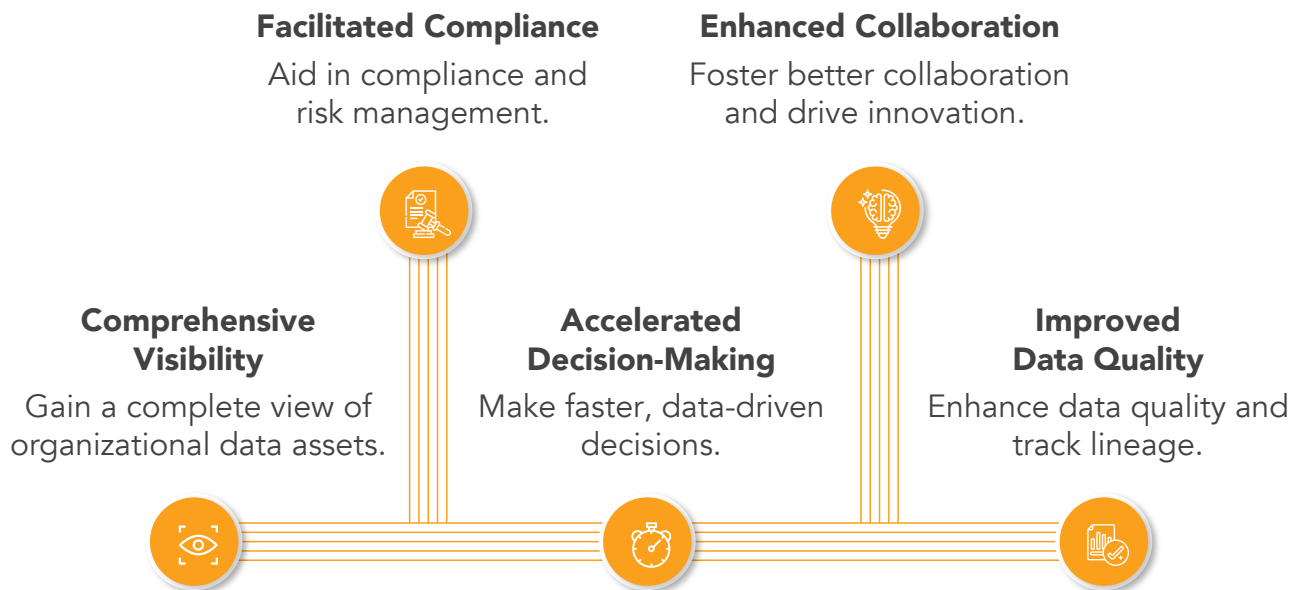
Importance of Data Discovery

Data discovery enables organizations to accurately identify payment data flows within their systems. By mapping the exact pathways of payment data—from entry points like online transactions and mobile wallets, through internal movements across storage locations and processing platforms, to final destinations such as third-party vendors—organizations can craft effective data protection strategies. Additionally, data discovery can uncover hidden repositories of payment data in legacy systems, log files, or employee desktops. Locating these often-overlooked data points ensures they are subject to the same security protocols as actively used data, closing potential security gaps.

Beyond enhancing security, data discovery provides actionable insights by analyzing data patterns, trends, and anomalies. These insights drive innovation and inform decision-making processes, enabling businesses to respond proactively to market changes and customer needs. Moreover, understanding data relationships and content through discovery tools reveals how different data sets interact, supporting more effective data management and operational efficiency. This process also aids in compliance with increasing data privacy and security regulations, ensuring organizations have the necessary visibility and control over their data. In summary, data discovery empowers organizations to leverage their data more effectively, driving informed decision-making and innovation while ensuring robust protection and management of sensitive information.



Benefits of Data Discovery



Challenges of Data Discovery



Streamlining Operations with Data Classification

Following the discovery phase, data classification comes into play. This process involves analyzing the types of data unearthed during discovery and organizing them into categories based on predefined keywords, patterns, and labels. Data classification tools are crucial for linking disparate pieces of data, thereby facilitating the organization's ability to process different datasets for various purposes.

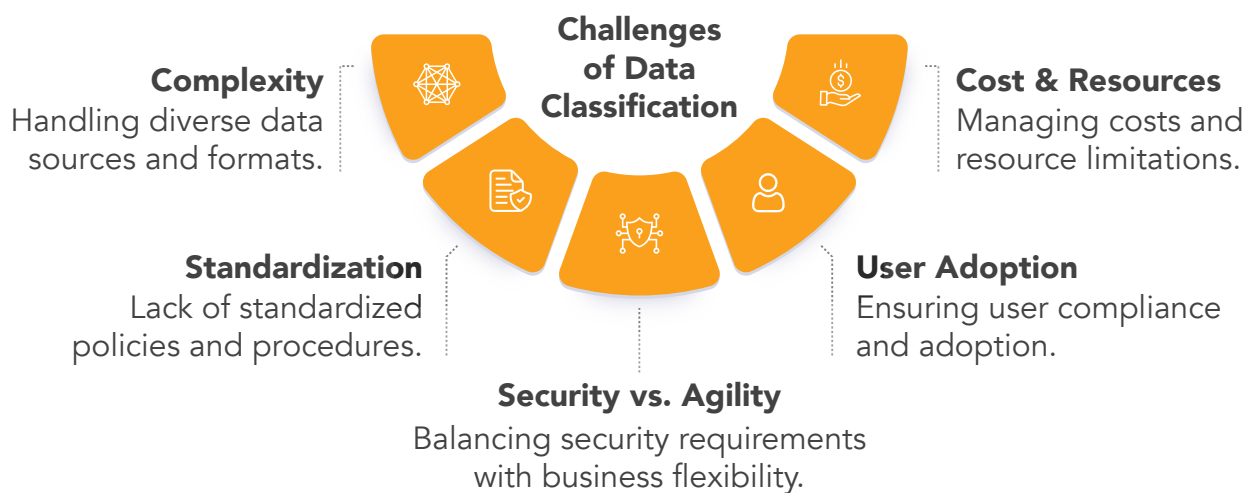
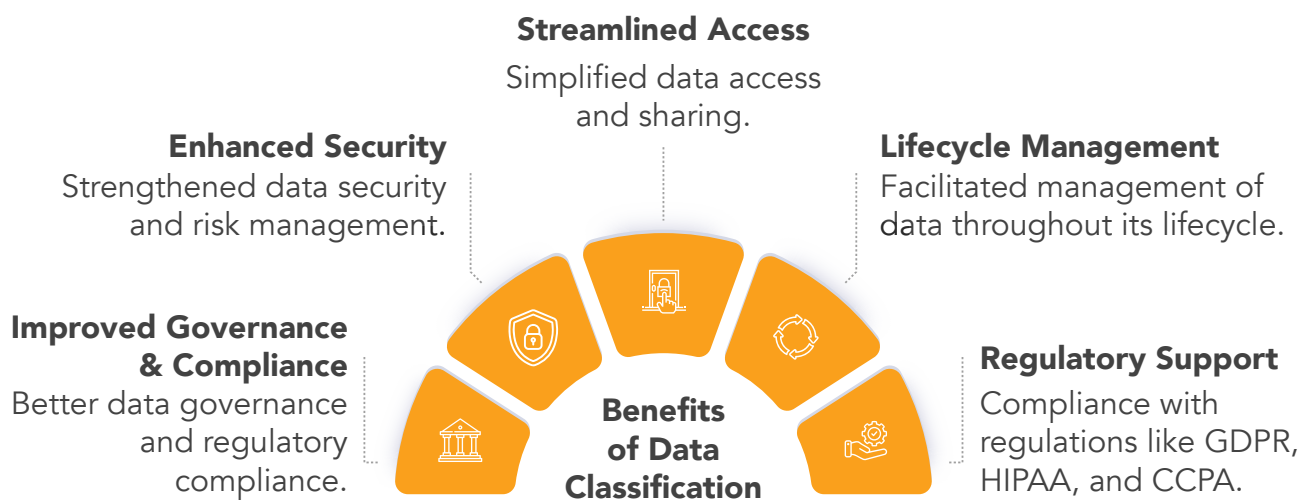
Automated data classification tools address the limitations inherent in manual classification, such as the inconsistency and inaccuracies that can arise when multiple employees handle the data. These tools provide a consistent, error-free approach to data classification, saving the time and effort that security teams would otherwise spend on manual processes.

Importance of Data Classification

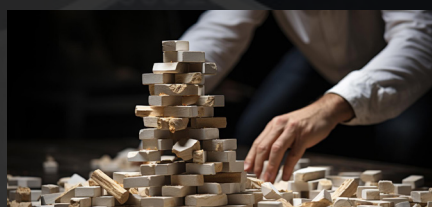
Data classification enables organizations to categorize payment data based on its sensitivity. This classification process includes identifying critical information such as credit card numbers, bank account details, and other sensitive data that could be exploited for fraudulent purposes. By assigning appropriate classifications, organizations can prioritize specific security measures for their most sensitive payment data, ensuring that the highest levels of protection are applied where needed most. This targeted security approach allows for efficient allocation of resources, focusing stringent controls like encryption and access restrictions on high-value data, while applying more streamlined protocols to lower-risk information. As a result, organizations can maintain robust security without unnecessarily hindering business operations.

Furthermore, data classification supports compliance with regulatory standards such as the Payment Card Industry Data Security Standard (PCI DSS). By clearly identifying and classifying payment data, organizations can demonstrate compliance by ensuring PCI-compliant security protocols are applied precisely to the required data. This not only simplifies audits but also reduces the risk of non-compliance penalties, providing a clear framework for maintaining data security. Additionally, having a structured classification system in place allows organizations to respond more effectively to security incidents and audits, as they can quickly identify and address vulnerabilities in their payment data handling processes. In essence, data classification optimizes security resource allocation and enhances regulatory compliance, safeguarding sensitive payment information while supporting efficient business operations.

Benefits and Challenges of Data Classification



Leading Public Sector Bank in India Enhances Data Security with SISA Radar – A Case Study



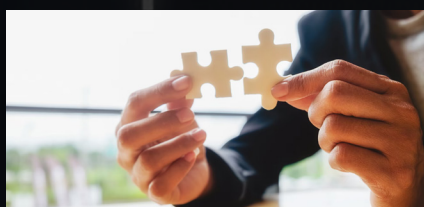
Challenge

A major public sector bank in India embarked on a digital transformation, creating a complex digital infrastructure with over 100,000 endpoints and 90,000 email users.

To address the growing concerns around data security, the bank sought an advanced data discovery and classification solution.

The bank faced several key challenges in securing its extensive digital ecosystem:

- Prevent accidental data sharing and internal security threats.
- Secure sensitive information across diverse data sources (Linux Endpoints, Databases, One Drive, SharePoint, Cloud).
- Adapt and scale effectively with a projected 25% data volume increase.



Solution

Implementing a comprehensive data discovery and classification solution, Radar, helped address these challenges effectively. Key features include:

- **Automated Data Discovery and Classification:** Identifying and classifying sensitive data across all sources using AI/ML technologies.
- **Email Security Integration:** Securing email communications by integrating with various email environments and enforcing strict DLP policies.
- **DRM Compatibility:** Ensuring compatibility with existing Digital Rights Management systems to protect copyrighted content.
- **Scalability and Adaptability:** Seamlessly scaling to manage data growth and adapting to changing security requirements.



Derived Benefits

- **Streamlined Data Protection and Governance:** Automated data discovery and classification streamlined the bank's data protection and governance processes.
- **Enhanced DLP Policies:** Automated data classification enabled the enforcement of stronger Data Loss Prevention (DLP) policies.
- **Improved Threat Monitoring and Incident Management:** Enhanced monitoring capabilities led to more effective threat detection and incident management.
- **Tighter Access Control and Content Management:** Improved access control and content management practices reduced the risk of unauthorized data access.
- **Protected Sensitive Content:** Secure handling of copyrighted and sensitive content prevented data leakage through email and internal channels.
- **Scalability:** The solution scaled effectively to manage the projected 25% increase in data volume.

The Evolving Data Deluge: Why Smarter Classification Mechanisms Are Crucial for Payment Data Security

The once-predictable realm of structured data, conveniently arranged into databases of rows and columns, has given way to a vibrant and varied ecosystem of data sources. This shift from structured to largely unstructured data formats—ranging from sensor outputs and social media content to complex image files—demands a revolutionary change in how we classify data. In this ever-growing tide of unstructured data, particularly payment data, smarter classification mechanisms are no longer a luxury but a necessity. Here is how these advanced mechanisms address the challenges of managing and securing sensitive payment information.



01 From Uncontrolled Chaos to Organized Security

Unstructured payment data, such as transaction logs and receipts, often resides scattered across various systems. Traditional methods struggle to manage this sprawl, leaving it vulnerable to breaches and compliance issues. Effective classification transforms this chaos into organized security by employing:



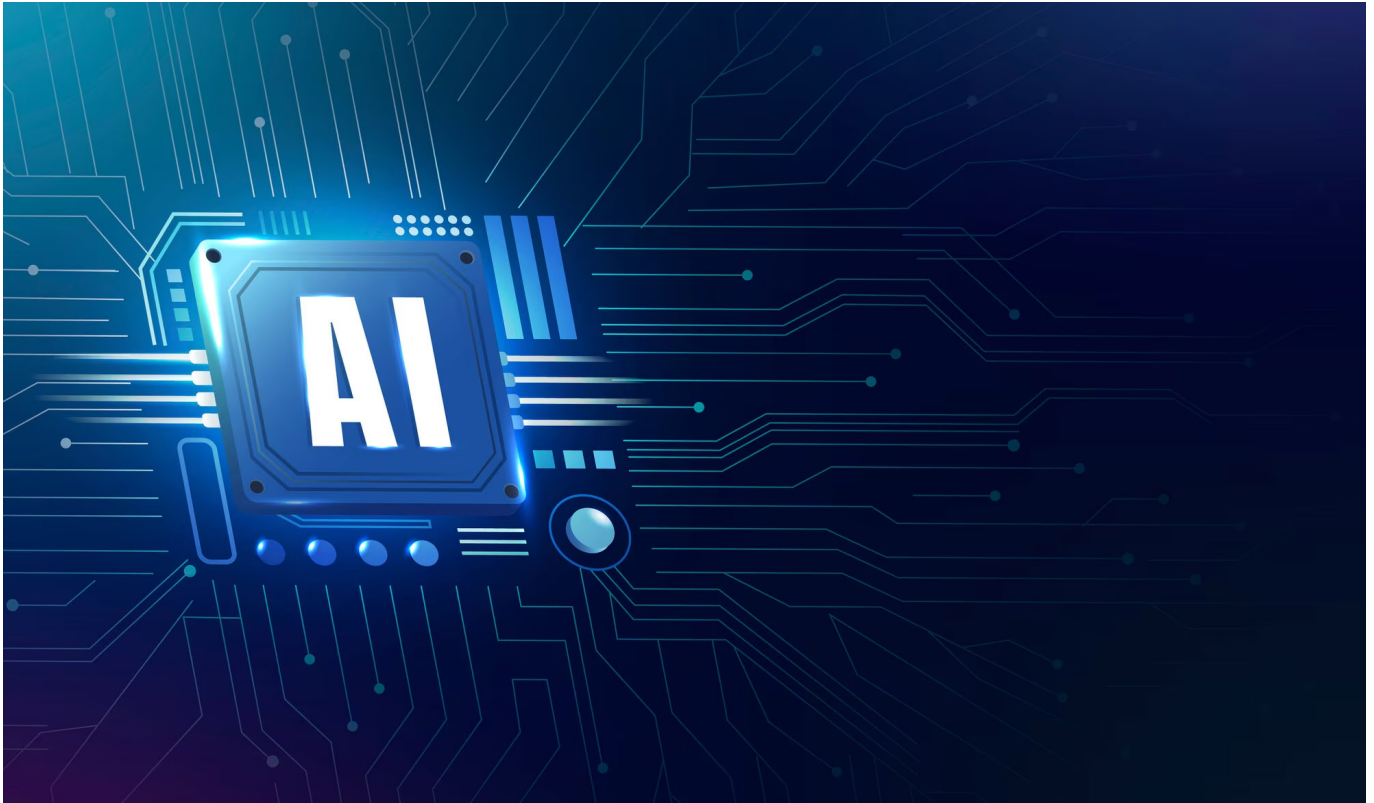
- **Content-based Analysis:** This involves identifying keywords and patterns within the data itself, such as credit card numbers, bank account details, and transaction amounts. This helps in pinpointing sensitive information quickly and accurately.



- **Contextual Understanding:** Considering the source, purpose, and context of the data is crucial. For instance, classifying transactions based on merchant categories (groceries, utilities) can provide valuable insights for budgeting and fraud detection, helping organizations to not only secure the data but also derive actionable intelligence from it.



- **Utilization of Existing Metadata:** Leveraging system-generated information like creation date, last access time, and user access logs helps in identifying inactive or obsolete data that can be archived or deleted. This not only frees up valuable storage space but also reduces the amount of data at risk.



02 AI and Machine Learning: Powering Smarter Classification

The sheer volume of payment data makes manual classification impractical. Here's where AI and Machine Learning come in, transforming how data is managed and secured:



- **Automated Scanning:** AI-powered tools can rapidly scan vast amounts of data, searching for specific keywords and patterns indicative of payment information. This automation significantly reduces the time and effort required for data classification, making it feasible to handle large datasets effectively.



- **Enhanced Metadata Enrichment:** These tools delve deeper than traditional methods, analyzing content beyond basic file types. They can identify specific items within documents or images, such as account numbers masked within a receipt or transaction details embedded in emails, providing a comprehensive view of the data landscape.



- **Continuous Learning:** Machine learning algorithms constantly refine their classification abilities based on past experiences. This ensures they stay ahead of evolving threats and identify new types of payment data, adapting to changes in data patterns and emerging security risks.

Implementing these advanced mechanisms ensures not only the protection of sensitive payment data but also compliance with regulatory standards, operational efficiency, and the ability to derive meaningful business insights. By leveraging AI and ML based smarter data classification solutions, organizations can develop dynamic systems that not only secure payment data more effectively but also unlock its full potential, paving the way for deeper insights and informed decision-making. This ensures that data privacy is maintained while maximizing the utility and value derived from payment data in today's data-driven world.

Best Practices for Data Classification

Clear Criteria & Policies

Define and enforce classification criteria and policies.



User Training

Conduct training and awareness programs for users.

Automated Tools

Use automated tools and technologies for efficient classification.



Regular Audits

Regularly audit and refine classification processes.



Workflow Integration

Integrate classification into existing data management workflows.



References

1. <https://www.forbes.com/sites/chuckbrooks/2023/03/05/cybersecurity-trends--statistics-for-2023-more-treachery-and-risk-ahead-as-attack-surface-and-hacker-capabilities-grow/?sh=580d664a19db>
2. <https://www.forbes.com/sites/chuckbrooks/2023/03/05/cybersecurity-trends--statistics-for-2023-more-treachery-and-risk-ahead-as-attack-surface-and-hacker-capabilities-grow/?sh=580d664a19db>
3. <https://www.biometricupdate.com/202405/who-is-looking-out-for-your-data-security-in-an-era-of-wide-spread-breaches>
4. <https://circleid.com/posts/20240110-sensitive-data-discovery-the-first-step-in-data-breach-protection>
5. <https://ciosea.economictimes.indiatimes.com/news/security/crowdstrike-and-rubrik-partner-to-transform-data-security/108958675>
6. <https://www.statista.com/statistics/1233673/sensitive-data-location/>
7. <https://www.capterra.com/resources/data-classification-breach-factors/>
8. <https://blog.knowbe4.com/88-percent-of-data-breaches-are-caused-by-human-error>
9. <https://gdpr.eu/fines/>
10. <https://www.helpnetsecurity.com/2024/05/03/saas-security-priority-for-organizations/#:~:text=The%20tenant%20research%20revealed%20that,don't%20need%20access%20anymore.>
11. <https://www.maximizemarketresearch.com/market-report/digital-payment-market/16835/>