WHITE PAPER

How AI and Machine Learning Are Redefining Anti-Money Laundering

The bad actors are making strong plays. Take every advantage to fight back.

sas

Contents

The evolving landscape of money laundering 1
Here's where AI comes in 1
Flavors of machine learning 2
- Supervised machine learning 2
- Unsupervised machine learning 2

Machine learning at work – six use cases for AML 3
- Machine learning as a supplement to transaction monitoring 3
- Machine learning for anomaly detection 4
- Machine learning for customer segmentation 5
- Machine learning for customer risk ranking 5
- Machine learning for social network analysis 5
- Machine learning for threshold setting and tuning 5

The three top challenges to adopting machine learning 6
- The necessary knowledge, people and systems are not in place 6
- Training data is difficult 6
- Machine learning models tend to be black-box systems 7

How SAS can help 7
- Key features 7

The evolving landscape of money laundering

Financial criminals are shrewd in disguising the origins of their illegal profits and getting that money into the financial system, either for personal gain or to support criminal enterprise. While most laundered money stems from drug trafficking and organized crime, the events of 9/11 also put the spotlight on covert funding for terrorist activities, which has traditionally been even more difficult to detect.

It's a daunting challenge. Consider the immense volume of data that financial institutions are expected to comb through to meet regulatory requirements to detect and report suspicious activity. The data is usually diverse and subpar. It's common for systems to use only a subset of available data when generating alerts. Traditional transaction monitoring systems are unwieldy to maintain and rely on rules and thresholds that are easy for criminals to test and circumvent. Investigation processes tend to be highly manual, from gathering the supporting data for a case to submitting a complete SAR (suspicious activity report).

Meanwhile, the money launderers are working night and day to remain hidden, constantly engineering new ways to conceal the flow of funds.

Traditional anti-money laundering (AML) and combating the financing of terrorism (CFT) tools and tactics take longer and cost more than they should. To fortify the defense, financial institutions need ways to:

- **Automate tasks** that formerly required human intervention, such as disposition of alerts.
- **Detect more risk** and effectively prioritize it with sophisticated analytics techniques.
- **Provide richer context** for investigations with access to more comprehensive insights.

Here's where AI comes in

The concept of artificial intelligence (AI) conjures up visions of robots who learn too much, grant themselves too much power and vanquish their creators. The reality of AI is far less dramatic. Broadly speaking, it's about allowing a machine to make a decision a human could have made.

When Amazon and Netflix recommend things you might like, AI is behind the scenes. When Siri and Alexa intelligent personal assistants help you organize your life and make recommendations, or when facial recognition authenticates your online or mobile payments, artificial intelligence is at work. From driverless cars to personalized product offers to detecting credit card fraud, AI and machine learning technologies are benefiting a host of industries and creating new ones not dreamed of.

A subset of AI, *machine learning* enables a computer program to learn from data rather than through explicit programming. These programs work by taking example data, finding patterns in it that might be too complex for a human to intuitively see, then applying the findings to new data. When this learning capability is coupled with modern computing power, you have a recipe for a system that can make complex decisions in an automated way.

Some machine learning techniques can distill new data elements not previously known or available for AML/CFT models. These models can interpret previously unknown patterns from vast data sources, which in turn can feed into other modeling efforts. Other machine learning techniques can be used to directly make predictions based on the patterns they find. The more training a model gets with feedback data, the more accurate it becomes and the less tuning it requires.

Flavors of machine learning

The right machine learning approach for AML/CFT depends on the input you have for training the model and what you hope to achieve.

Supervised machine learning

With supervised machine learning, the model is presented with sample inputs and their associated outputs, and the goal is to devise a general rule that maps those inputs to outputs. For example, what attributes were associated with cases that turned into SARS? What findings were associated with false positives or false negatives? The model learns how to better predict the outcome when it is applied to new data.

Typical supervised learning techniques include:

- **Bayesian statistics**, which describe the conditional probability of an event based on incoming data as well as prior information or beliefs about the event.
- **Decision trees**, a tree-like "if/then" model of decisions and their possible consequences.
- **Neural networks**, where "neurons" are connected by weight, a widely used technique for speech recognition, image analysis and adaptive control.
- **Regression analysis** to model and analyze the relationships between a dependent variable and one or more independent variables.
- **Random forests**, an ensemble learning method for classification, regression and other tasks, that operates by constructing a multitude of decision trees.

Unsupervised machine learning

With unsupervised machine learning, the algorithm learns from sample data that has not been labeled, classified or categorized. The algorithm is on its own to find structure or hidden patterns in the data. Since you don't know which data represents suspicious activities, you want the model to create a function that describes the structure of the data, flags anomalies and then applies this knowledge to new data.

Common unsupervised learning techniques include:

- **Affinity analysis**, a data analysis and data mining technique that discovers relationships among activities performed by (or recorded about) specific entities.
- **Clustering**, a form of exploratory data mining that groups "objects" in a group (cluster) where objects are more similar to each other in some sense than to those in other clusters.
- **Nearest-neighbor mapping**, an optimization exercise of finding the point in a given set that is closest (or most similar) to any given point.

By looking at all accounts, customer attributes and events together, unsupervised machine learning can uncover meaningful patterns that would not be seen with traditional rules-based AML systems.

In short, machine learning enables a computer program to learn from the data rather than through explicit programming. The program takes sample data, finds patterns in it that might be too complex for a human to see, and then applies the findings to new data.

With advances in modern computing power, highly complex decisions can be automated with greater speed and accuracy. It's easy to see the value of machine learning for keeping pace with evolving AML/CFT schemes.

Machine learning at work – six use cases for AML

Thanks to advancements in handling big data, machine learning can now change the architecture for AML. Financial institutions can either:

- Replace their rules-based engines with machine learning models.
- Use machine learning as a support system to develop models that feed into and out of the rules-based engine, bringing new intelligence to such activities as risk ranking, rule tuning and alert prioritization.

Let's look at some ways machine learning is already being put to use for AML.

Machine learning as a supplement to transaction monitoring

Historical alerts associated with suspicious behavior, productive dispositions or SAR filings – known "good" alerts – can be labeled, and a supervised machine learning model can be trained on this data to become smarter in scoring and prioritizing new alerts. Investigators can then:

- Focus on the alerts that are more likely to result in a SAR filing.
- Uncover hidden links between today's bad actors and their historical counterparts.

One approach is to build an *alert prioritization* machine learning model on top of an existing transaction monitoring system to score alerts by their level of suspiciousness. Or a machine learning model can be designed to augment or entirely replace rule-based scenarios. Either way, a machine learning model can help investigators prioritize workflow queues for review, or auto-disposition alerts – escalate high-risk alerts and "hibernate" low-risk alerts.

With machine learning models, you can:

- **Reduce false positives with tighter segmentation.** Rule-based scenarios and segmentation models are usually broad and based on conventional data fields such as account type and customer demographics. A broad-brush approach leads to more false positives, more workload for investigators and higher cost to the organization. In contrast, machine learning techniques such as clustering can define more granular segments for a more focused assessment of risk.
- **Increase true positives with powerful analytics.** Machine learning models can identify suspicious activities and behaviors that traditional rules-based transaction monitoring systems would miss. These models can uncover hidden links and relations, unearthing new patterns of suspicious activities that were previously undetectable.

Models can process massive amounts of data about demographic, geographic and activity information of the bad actors and their networks. Armed with a holistic, global view, investigators can carry out more effective research.

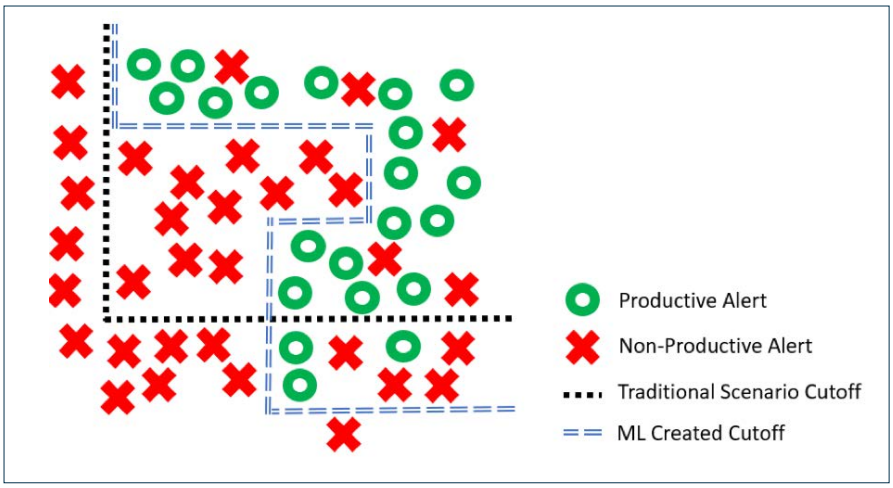

Figure 1. A machine learning model can provide more productive alerts on suspicious activity by considering more complex relationships.

An Asia Pacific region bank reduced false positives by 33% by developing an alert scoring and "hibernation" approach for predicting worthy investigations.

Machine learning for anomaly detection

AML programs are based on the concept of finding suspicious activity, but there's no objective definition of "suspicious." That's a problem for supervised machine learning, which needs to learn from labeled examples, such as "suspicious" or "not suspicious."

Anomaly detection techniques address this issue by identifying observations that appear mathematically "distant" from the expected – different either from their own historical activity or the current activity of peers. The outliers – behavior outside the norm – might warrant a closer look. With little direction and no labeled data, anomaly detection methods can spot potentially suspicious activity not defined in a rule.

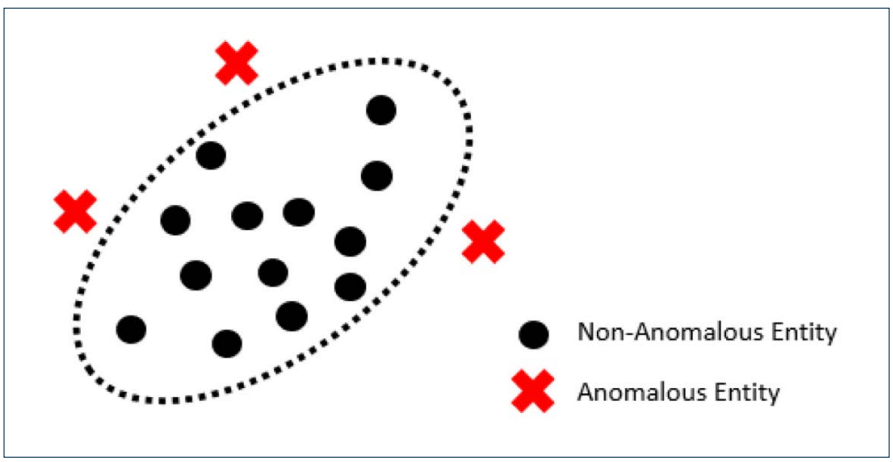

Figure 2. A simple, conceptual example of outlier detection based on two variables.

A Tier 2 regional US bank modernized its rules-based AML transaction monitoring system by deploying a SAS® neural network model, which reduced work items by 50% and increased SAR conversion rates from 5% to 15%.

Machine learning for customer segmentation

Different customers transact differently and therefore should be monitored differently. So, naturally, AML programs segment customers into peer groups. There are many logical ways to segment peer groups – by demographics, business structure or transaction volume, for example. But unsupervised machine learning algorithms offer more sophisticated possibilities.

For example, a clustering algorithm such as K-means clustering classifies entities into similar groups by identifying observations that are mathematically "close" to each other by one or multiple factors. Just as two cities might be similar based on a combination of factors such as physical distance, density plus population size, different entities may be seen as similar across a combination of attributes and be monitored accordingly.

A Tier 1 global bank tested the validity of applying AI to detect hidden risks or false negatives within its client base and identified 416 customers suspected of operating as money service businesses, 89 of which were previously undetected.

Machine learning for customer risk ranking

Machine learning algorithms can use known data about a customer's historical behavior and suspicious activities that result in SAR filings to risk-rank (or score) other customers with similar attributes. This approach can be combined with data outside the organization to give a more holistic and global view of the customer as part of the know your customer process. Machine learning models can be trained for customer screening and onboarding by assigning scores and using these scores as labels to train transaction monitoring system models.

Machine learning for social network analysis

Machine learning can identify interwoven networks of suspicious activities and actors. Unsupervised machine learning models can look at accounts and customers involved in transactions with a bad actor or a person on the politically exposed persons (PEP)/sanctions list. Machine learning models can quickly uncover hidden patterns, links and networks among millions of accounts and variables — and connect bad actors with others in their network.

A Tier 1 global bank improved the customer experience for trade finance business by accelerating due diligence times and improving accuracy. The bank deployed automated, deep learning algorithms to classify the types of documents under review, thereby reducing the review process effort from two person-weeks to about a minute.

Machine learning for threshold setting and tuning

To be effective, rules/models must adapt to what's happening in the world. Imagine the power of a system that can automatically examine masses of data to help establish rules and keep them up to date. Machine learning can be used to establish data-derived thresholds that calibrate themselves based on what they learn. These models can process large volumes of data on productive alerts to establish more accurate and precise thresholds. Based on the output of these models, organizations can set up ongoing, automated tuning — a far better way to manage an otherwise time-intensive process.

The three top challenges to adopting machine learning

In spite of the proven benefits of machine learning in multiple industries, financial institutions have lagged in adopting it for AML. From SAS' experience with clients in this industry, three issues are the biggest blockers:

- The necessary knowledge, people and systems are not in place.
- Training data is difficult.
- Machine learning models tend to be "black box" systems.

Let's take a closer look at these challenges.

The necessary knowledge, people and systems are not in place

It's no secret that there's a dearth of data science professional talent. Despite the emergence of many new college degrees and certification programs, companies still struggle to find and retain qualified professionals to carry out their analytics ambitions. A good model development process is iterative, ongoing and time-intensive. It entails testing various model types, combinations of input variables and model parameters – and then continuously updating, testing and redeploying.

For organizations that are short on data science talent, part of the solution is to have a software toolkit that automates much of this analytics work and manages the model life cycle. Nonetheless, it is still useful to have human expertise at hand to design and tweak these systems in an intelligent and efficient way.

SAS has a deep bench of highly experienced analytical consultants who implement solutions in a wide variety of industries. Our Security Intelligence Solutions team consists of industry experts with AML analytics experience at Tier 1 banking institutions.

Training data is difficult

For a machine learning algorithm to work, the data to build it has to be identified, gathered, merged into a centralized data mart and managed for quality. Even supposedly "clean" data can have unknown issues such as missing fields or outliers.

Supervised machine learning models also require sample output data – for example, data on productive and non-productive alerts. Some organizations may be able to use data from an existing AML system, but any new features will need freshly labeled data, which requires work from investigators.

These necessary data preparation steps are often overlooked.

SAS has established data architectures and processes in place to tackle these data management problems as part of our standard AML solution.

Machine learning models tend to be black-box systems

While machine learning models are fantastic at finding patterns, they don't do it in a way that is easy to understand and explain to a regulator. Many of the machine learning models receiving the most buzz right now – neural networks, random forests and gradient boosting – are opaque approaches. The output of these models is complex to the point of being impractical to decipher. Techniques exist to reverse-engineer the output in a way to understand them better, but these models are still challenging to explain.

Alternatively, "white box" statistical modeling techniques, such as regression models and decision trees, trade some predictive power for transparency, which is useful for explaining to a regulator. And rules-based scenarios are naturally the easiest to understand and explain.

So what type of analytics should your organization use? It depends on the nature of your problem and the level of model opaqueness your organization is comfortable with. More analytically mature financial institutions have seen positive results with black-box models, but not every organization is ready to take that step. Depending on the data and use case, the incremental gain from a more complex approach might not be justified. You may decide to experiment with both to get a full understanding of the trade-offs.

SAS supports both approaches and can help guide you through this decision.

How SAS can help

SAS has been developing analytics software for more than 40 years. Working closely with financial services customers, we've developed a solution that covers AML processes across all key areas, such as suspicious activity monitoring, customer due diligence, watch-list filtering and investigations case management.

Key features

- **Data management.** Address the data challenges of AML, from processing big data to accessing and integrating legacy sources – all in a single platform.
- **High-performance analytics and visualization.** Get rapid insights from big data with an infrastructure that enables you to test hypotheses, ask questions and simulate scenarios.
- **Suspicious activity monitoring and reporting.** Rely on a robust, flexible scenario engine that more accurately detects suspicious activity and generates alerts for events that meet defined parameters.
- **Watchlist matching.** Use fuzzy-matching algorithms, intelligent scoring and alert consolidation to identify persons, organizations or jurisdictions that represent regulatory risk.
- **Investigation and alert management.** Use a web-based interface to gain a big-picture view of investigation work items, with easy access to a knowledge center database.

- **Peer-group anomaly detection.** In-memory analytics rapidly identifies potentially suspicious activity by comparing an entity's behavior to historical and peer behavior.
- **Search.** Search large data repositories with distributed indexing, replication and load-balanced querying, automated failover and recovery and more.
- **Multitenant architecture.** Securely serve multiple groups with a single installation by segregating the data.

The newest generation of SAS® machine learning software is offered on the SAS® Viya® platform, which works in an in-memory, distributed environment. SAS automates the end-to-end processes associated with AML, from data management to model management to case management and governance.

SAS Viya serves a broad range of users, from business analysts to data scientists. There are intuitive point-and-click interfaces or the option to program with the Base SAS programming language or supported APIs from other popular programming languages such as R and Python. And of course, there are intuitive data visualizations that enable business users to explore the data in multiple dimensions. Your developers, analysts and leaders can choose their preferred mode of working.

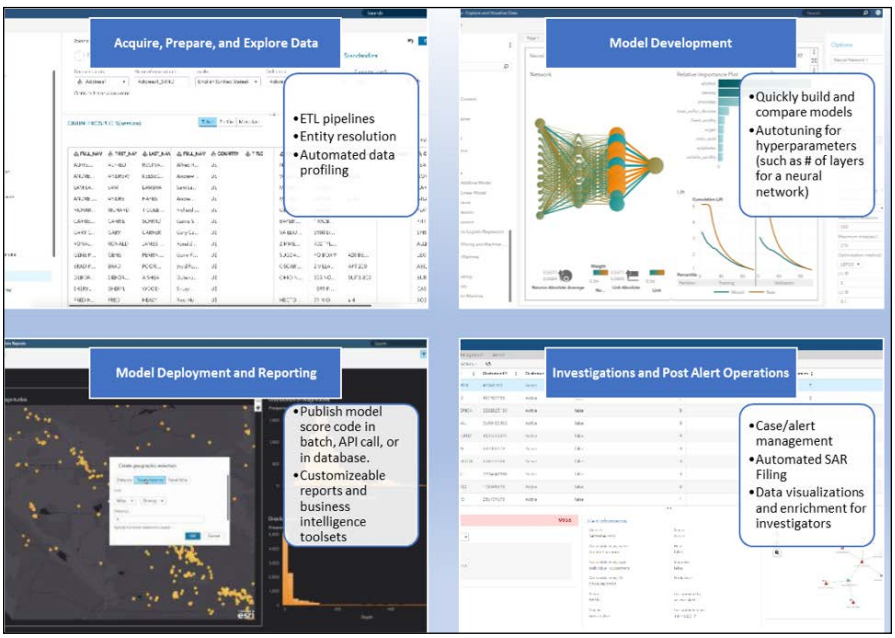

Figure 3. SAS supports the AML continuum, from managing data and models to managing alerts, investigations and reporting.

Learn more at sas.com/en_us/software/anti-money-laundering.html.

Contact us at GHUSPSDAMLPresales@sas.com.

To contact your local SAS office, please visit: sas.com/offices