Money laundering is currently the third largest “business” in the world after currency exchange and the auto industry. The Financial Action Task Force (FATF) estimates money laundering to be 2 to 5 percent of the world’s gross domestic product (GDP), which amounts to $1.38 trillion to $3.45 trillion. However, FATF further clarifies that this is merely an estimate. Because money laundering is a global issue, there is really no way of knowing the true picture of legal and illegal funds that flow through the world’s financial systems.

Financial criminals seem to be one step ahead and at the ready to launder illicit proceeds from legitimate financial products as they are developed and commercialized. The introduction of technology-based digital products and services such as mobile banking, virtual currencies, remote deposits and prepaid
cards pose a higher risk to financial institutions. These offerings are particularly vulnerable to criminal enterprise. However, innovation and technology are an absolute necessity for financial institutions and law enforcement in order to combat financial crime in this dynamic digital world.

Driving the Need for Better Detection

Technology has become a double-edged sword as bad actors navigate new conduits to facilitate their money laundering activities and develop new skills to escape detection. Survival of the fittest dictates who wins the war on financial crime. That is why financial institutions are embracing analytics and more advanced technologies to close gaps in their anti-money laundering (AML) programs and improve their ability to identify risk.

A recent report by Kennedy Consulting Research and Advisory indicates that a variety of internal and external industry drivers are fueling this trend. Among the external factors cited in the report are increased regulatory pressure, competitive pressure and technology advancement. Internal drivers include data proliferation and growth, the increasing sophistication of users and the maturation of enterprise resource planning (ERP) systems.

*Artificial intelligence plays a significant role in AML and fraud technology*

Other research agrees that while regulatory pressure is a key external driver, the threat of fines and concern about potential loss of reputation and market share are other notable considerations. At the same time, sophisticated systems that employ artificial intelligence and machine learning are more affordable and therefore available to a wider audience. Until recently, the use of these advanced technologies was limited to top-tier financial and government institutions with
deep pockets and a high level of expertise. Innovative cloud-based analytics, data visualization and other tools designed for the business user are changing the technology landscape.

The Evolution of AML Technology

The vendor landscape for fraud, AML and financial crime solutions is a broad mix of point solutions and more integrated enterprise platforms. These solutions are predominantly rules-based systems, which are designed to interact with data silos. The efficacy and accuracy of any rules-based system, simple string/pattern matching, or basic statistical approach is highly dependent on the scale and context of the problem.

Entity resolution, for example, relies on mining the enormous universe of unstructured data to identify hidden risk and find the bad guys. When trying to solve the problem of entity resolution, rules-based systems fall short for several reasons:

- Rules are static and cannot handle scale, scope or uncertainty
- Rules are built for lists that are small, curated, structured and discrete, making it difficult, if not impossible to handle Internet-scale continuous know your customer (KYC)
- Rules need constant updating, which is typically a manual effort
- Rules do not solve the false positive problem
- Rules do not provide an observation space from which one can reason

Struggling to handle the deluge of false positives, a financial institution’s first response is usually to throttle the scope—increasing thresholds, parsing lists, eschewing data points—at the expense of creating false negatives and missing
the very things that are the purpose for their effort. As a result, rules-driven solutions leave huge gaps. Unlike principles-based solutions, rules-driven solutions are unable to handle the massive, messy, open-ended Internet ecosystem with its diverse population of data content, structure and reliability.

Technology trends and globalization have increased the reach and risk posed by transnational organized crime (TOC), powerful organized non-state malefactors and a global criminal/political nexus. For these reasons, Internet-scale (i.e., global, multilingual, heterogeneous, diverse, semi-unstructured KYC relevant data sources) solutions that are qualitatively different in scale, scope and reliability are necessary for monitoring entire customer databases in near real time.

**Converting Data to Actionable AML Intelligence**

A new paradigm is emerging where principles-based AML systems grounded in scientific disciplines are replacing inflexible rules-based solutions. Advanced data-driven techniques that rely on a principles-based approach—like machine learning—are changing how institutions identify fraud and financial crime.

To understand machine learning, one must understand the other three interrelated scientific disciplines found in advanced technologies: data mining, statistics and artificial intelligence. At the center of these disciplines is data mining, which is the process of extracting patterns from data. It relies on the use of real-world data and applies mostly machine-learning algorithms to solve domain-related problems. Data mining pattern recognition aims to classify data (patterns) based either on a priori knowledge or on statistical information extracted from the patterns.
Intelligence data is often relationship data. Data mining is becoming an increasingly important tool to transform data into information. It is commonly used in a wide range of profiling practices, such as surveillance and fraud detection. Data mining was cited as the method used by the U.S. Army unit Able Danger that reportedly identified Mohamed Atta and three other 9/11 hijackers as possible members of an al-Qaeda cell operating in the U.S. for more than a year before the attack.

Statistics, the oldest of the four disciplines, is believed to have started around 1749 when the first census was established in Sweden. Statistics is the practice or science of collecting, organizing, analyzing and interpreting numerical information from data. Business statistics use regression analysis to determine the validity of relationships, an effective technique in AML advanced technologies. On the cutting edge of AML technology are those solutions that also employ other statistical data analysis techniques such as clustering and classification to find patterns and associations among groups of data, matching algorithms to detect anomalies and models, and probability distributions to prioritize the likelihood of true matches.

In 1956, the computer scientist John McCarthy coined the term “artificial intelligence” (AI) to describe the theory and development of computer systems able to perform tasks that normally require human intelligence, such as visual perception, speech recognition, decision-making and translation between languages. AI plays a significant role in AML and fraud technology. Based on the assumption that human thought can be mechanized, many AI engines employ fuzzy logic matching algorithms to facilitate watch-list filtering. The most effective solutions for detecting and managing fraud incorporate several AI techniques, including:
- Data mining to classify, cluster and segment data and automatically find associations in the data that may signify interesting patterns
- Pattern recognition to detect approximate classes, clusters or patterns of suspicious behavior either automatically (unsupervised) or to match given inputs
- Machine learning techniques to automatically identify characteristics of fraud
- Neural networks to learn suspicious patterns from samples and be used later to detect them
- Link analysis to evaluate relationships and connections between organizations, people and transactions

### Machine Learning

The origins of machine learning date back to the late forties through 1959 when it emerged as a branch of AI and was defined by Arthur Lee Samuel, an American pioneer in the field of computer gaming, AI and machine learning. Machine learning is a scientific discipline that is concerned with the design and development of algorithms that allow computers to solve problems by automatically learning to recognize complex patterns and make decisions from labeled data. These algorithms build a model based on inputs from which predictions or decisions are made, rather than follow only explicitly programmed instructions.

Machine learning is employed in a range of computing tasks where designing and programming explicit, rules-based algorithms is infeasible. By making inferences from masses of evidence, machine learning actually benefits from the dynamics of scale and combinatorial nature of evidence. As data grows, the
possible mixtures of evidence in which to make distinctions grow exponentially. Millions of data points and billions of data pairs offer trillions of different combinations that can be weighted by total evidentiary value per case and consolidated and mapped to a manageable range of total risk-based bands. This method involves machine learned/AI techniques that seek to emulate human judgment rather than execute a finite rule set.

The underlying premise of machine learning is that the difficulty lies in the fact that the set of all possible behaviors or results giving all possible inputs is too complex to describe generally in programming languages. This explains in part why rules cannot solve the problem of too many false positives when screening customers and transactions for sanctions, politically exposed persons (PEPs) and reputationally exposed persons (REPs) found in adverse media. As data grows linearly, pairs of data points grow exponentially, causing false positives to explode and the diagnostic power of any static rule to degrade to zero. This is the problem that plagues the AML market today.

The Benefits of Machine Learning in AML Technology

In 2014, a significant number of regulatory enforcement cases against banks cited the lack of a systemized approach to AML supervision and the lack of surveillance on a daily basis. Increasing regulatory expectations are forcing banks and other financial institutions to integrate external data into their operational risk framework. This is driving the need to use machine learning to pull external data and set filters to get relevant and scaled data on an automated basis.
Machine learning and neural analysis can cross-check banking data with public record databases to determine bad actors or suspicious transactions. However, the advanced solutions best suited for solving the problem of entity resolution related to financial crime not only incorporate machine learning techniques and processes but do so with a principles vs. rules approach.

When tackling the problem of entity resolution, AI and machine learning techniques can be utilized to compare two records and measure the probability and risk associated with a match. This is representative of classification, a desired output of a machine learned system. Inputs are divided into two or more classes and a model is produced that assigns unseen inputs to one or more of these classes. Ordering these results in a two-dimensional observation space enables the AML analyst to determine if two records represent the same entity. It also allows the analyst to establish thresholds of what to review and what to exclude from review. Other advanced analytic tools such as social network analysis and scatter plot visualizations can facilitate the detection of relationships and anomalies.

An AML solution designed to support machine-learning tasks and advanced analytics is an effective model to identify and monitor risk across the enterprise. It helps institutions to:

- Identify hidden risk
- Prioritize and order alerts
- Improve operational efficiency
- Optimize resources for more targeted investigation efforts
- Eliminate false negatives and significantly reduce false positives
Machine learning applied to AML technology provides a principled approach and framework that handles entity resolution in a formal but flexible and extensible way. It explicitly factors in the scale, inherent uncertainty and rapid change involved in large real-world tasks. Whereas fragile systems crumble at Internet scale, robust systems can withstand increasing stress and anti-fragile systems actually improve with stress. Robust, anti-fragile solutions make better inferences by bottom-up learning from data, top-down knowledge engineering or horizontal discovery of new evidence artifacts.

Conclusion

While it is difficult to accurately measure the global cost of fraud, money laundering and other financial crime, the AML community is well aware of the cost of compliance. However, the costs go beyond compliance to the business itself and when viewed in that context, they are staggering. PWC’s 2014 Global Economic Crime Survey indicates that economic crime remains a fact of life for every segment of the global business community, exploiting the tension between two fundamental business goals: profit and compliance. As financial institutions continue to evaluate their AML programs and supporting technology under intense regulatory scrutiny, they must not be afraid to consider a complete change in approach to entity resolution and the rules that currently drive most of the solutions they have embraced. Machine learning has proved to be an effective technique for identifying risk because it embraces principles-based methodologies, which better handle the enormous, unstructured and diverse Internet ecosystem.

Carol Stabile, CAMS, senior business manager, Safe Banking Systems LLC, Mineola, NY, USA, carol.stabile@safe-banking.com