The Dummy Handbook on False Positive Management for Fraud Detection

Become an expert in fraud detection that does not damage the customer experience

Contents

Introduction

It could have happened to anyone. She brought the extended family together, took everyone out to a fancy dinner in the big city, and insisted on picking up the tab. She'd just landed her dream job, and this was her chance to show everyone she'd made a success of herself. She didn't expect to have to ask someone else to pay because her credit card was declined. It wasn't her fault; her account was in good standing. But the out-of-the-ordinary transaction triggered an alert, a rule kicked in, and the charge was blocked. The experience was horrible.

Incidents like this are far too common. When a false positive happens, it's a sign something's broken in an organization's fraud prevention system. Good customers should never be punished for good behavior, and good companies should never lose trust—or business—because of a poor-performing fraud solution.

The Problem is Real

False positives are the antithesis of great customer experiences, and they're a costly problem. Per a recent report, Aite Group estimated that false positives grew to US $443 billion by 2021. Their research also found that younger consumers have the most negative attitudes about false positives, stating that, "of the millennial cohort, 59% say that they would be very or somewhat likely to leave their financial institution due to a credit card false decline."

> *False positives are the antithesis of great customer experiences. They represent unwanted and undeserved friction for good users—many of whom will abandon a given online experience before completing any or all of their intended actions.*

According to research from American Express, "more than half of Americans have scrapped a planned purchase or transaction because of bad service, and 33 percent say they'll consider switching companies after just a single instance of poor service." As noted by Merchant Fraud Journal, "Legitimate customers view declines as a personal insult, and will often retaliate by actively speaking badly about a brand. This kind of negative word-of-mouth is devastating for merchants seeking a foothold with their target audience."

A Change is Needed

These are heavy prices for organizations to pay, and the realization that something must be done is becoming progressively more widespread. According to PwC's Global Economic Crime and Fraud Survey, 34% of global C-level and senior management executives believe that existing approaches to combating online fraud are generating too many false positives. Per data from PYMNTS, "more than 60 percent of digital platforms say too many false positives are a significant point of friction in the conversion process, and more than 30 percent say it's their number-one challenge."

Why must false positives exist at all, and more importantly, can we eliminate them? We can, but doing so requires a new approach. Commonplace approaches that rely on legacy rules-based systems to filter and flag anomalies are failing, and businesses continue to struggle with determining how tightly to set their "filters." Too loose, and too much fraud can slip through. Too tight, and too many good customers experience too many false positives.

Next Generation Solutions

New technologies and solutions offer a way to reduce false positives while actually increasing accuracy, which in turn can actively enhance customer experience.

To understand how these solutions work, and why they represent significant improvements over existing and legacy solutions, we'll start by exploring false positives—what they are, and why they're a problem—and then move into an investigation of some of the more common approaches to fraud detection. We'll discuss the challenges these approaches are meant to solve, and how they do or don't measure up against their objectives. Finally, we'll look at cutting-edge new solutions that draw on the power of AI and machine learning to deliver reduced false positives alongside increased accuracy.

> *New technologies and solutions offer a way to reduce false positives while actually increasing accuracy, which in turn actively enhances customer experience.*

FALSE POSITIVES AND FRAUD PREVENTION

As DataVisor CEO and Co-Founder Yinglian Xie recently noted in a special report from The Times focused on "The Future of Fintech," the true goal of any fraud management strategy isn't actually detection, but prevention. It is precisely for this reason that false positives remain an ongoing challenge. To prevent fraud—as opposed to reacting to it in the aftermath of an attack—you have to see it coming before it happens. Fraud prevention, by its nature, is an inherently predictive effort, and a speculative one. In effect, fraud and risk teams rely on their solutions to make educated guesses about what may or may not prove to be fraudulent and act accordingly.

Poorly functioning fraud prevention solutions trigger too many false positives. Any organization dealing with this challenge must ask itself the following questions:

- Are we tracking the entire customer lifecycle to ensure holistic understanding and enable comprehensive protection?
- Are we looking at every possible attack vector?
- Are we identifying and dealing with new and emerging threats?
- Are we breaking down silos to centralize intelligence?
- Are we accurately distinguishing between legitimate and fraudulent actions and accounts in real time and at scale?

> *The true goal of any fraud management strategy isn't actually detection, but prevention.*

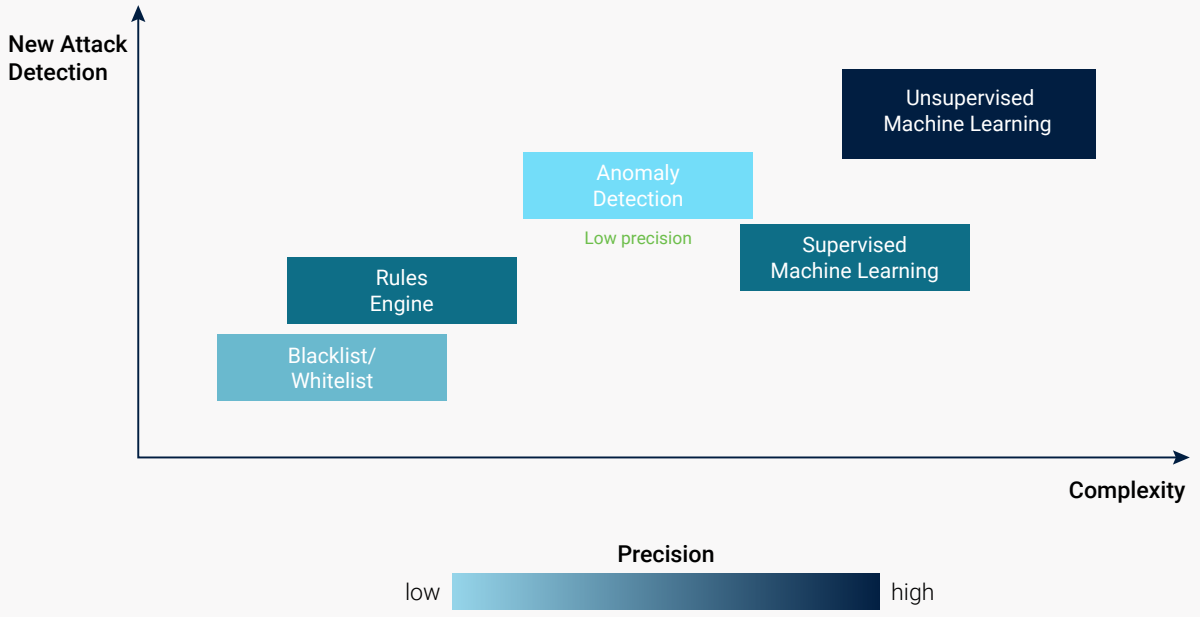
Let's look at some existing and legacy approaches and see how they stack up against these critical questions.

1 Fraud Detection Solutions: An Overview

As fraud has evolved over the years, solutions for preventing it have, as well. This cycle predates our digital era by decades, if not centuries. However, the scale of the problem today is massive, given our global digital economy. Today, there are a number of different approaches in active use across industries and use cases. These vary in efficacy with regard to speed, precision, and their respective abilities to address complex attack types and detect new and emerging fraud.

We'll begin by looking at one of the earliest forms of online fraud prevention: the "list" approach.

> *Reputation lists and reputation services no longer represent a viable defense against fast-evolving fraud. Reputation lists often lead to a surplus of false positives.*



REPUTATION LISTS

Reputation lists are lists of things such as email addresses, IP addresses, and device IDs that have been identified as either "safe" (white) or "unsafe" (black). If a particular IP address, for example, is flagged as being associated with fraudulent or malicious activity, it will be "blacklisted."

Reputation services work with businesses to try to help monitor and address blacklisted entities.

EXAMPLES

Email | IP Address | Phone # | Devices | Credit Card #

LIMITATIONS

- Quickly become out-of-date; need to be constantly refreshed
- Unreliable quality; lists often polluted, imprecise, or expired

Email Reputation Services

Email reputation services provide risk scores for email addresses. These scores are typically based on attributes such as the age of the email address, email frequency, and domain. Some email reputation services also validate identities using email-based information such as name, email address, IP address, and geolocation. Email reputation services use blacklists, along with email metadata, to assign scores.

IP Reputation Services

IP reputation services provide reputation scores for individual IP addresses. IP reputation services can signal that an IP address hosts malicious content, and also that it exhibits automated bot behavior. To assess risk, IP reputation services rely on the history of the IP—whether the IP address has exhibited malicious activity in the past, and if there are any changes since the last time the IP address was seen by the service.

Reputation lists and reputation services no longer represent a viable defense against fast-evolving fraud. Lists become outdated too quickly and require constant refreshing. It's also far too common for lists to be, at best, imprecise, and at worst, corrupted or expired. Modern fraudsters have also developed a wide array of techniques to successfully circumnavigate these lists. Finally, reputation lists often lead to a surplus of false positives.

Here are two examples of false positives:

1. Consider a scenario in which an IP address of a compromised machine is sending traffic to one service provider. There is a legitimate user on the machine (who is unaware that the machine has been compromised) and they are using it to access another service, so blacklisting would result in a false positive. Additionally, because consumer IP addresses get rotated quickly due to dynamic IPs, blacklisting the IP could also potentially result in blacklisting another family in the neighborhood—another false positive.

2. As a second example, consider a situation in which you have a large wave of attacks from a device emulator of an old Android version. In this scenario, an analyst might write a rule to block that particular version. At the attack's peak, that rule might have a 99% accuracy rate. However, when the attacker changes strategy and ceases to emulate that particular Android version, the rule will decay and ultimately become 0% accurate, as there will continue to be some legitimate users using that version.

RULES ENGINES

Rules engines represent a kind of next step forward from simple reputation lists. A rules engine is essentially backend software that can take predetermined actions based on specific criteria.

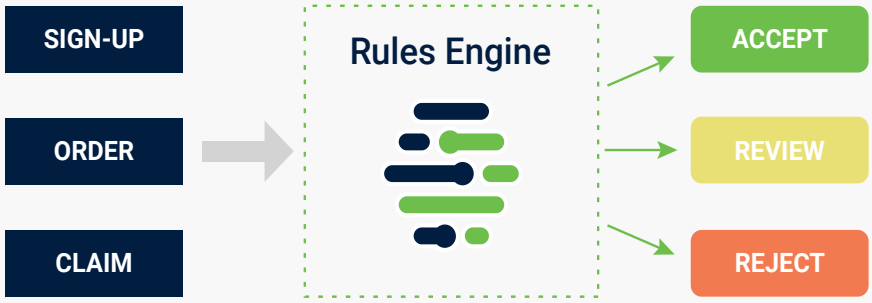
For example, if a business knows or believes that mismatches between billing country and IP country likely indicate malicious accounts or actions, a rule can be written to flag those instances, and that rule can be assigned a "weight" that represents the significance accorded that particular instance. Similarly, if it is known that accounts are more than 180 days old, a rule can be written to indicate a higher probability of safety. Based on existing rules, the system can then respond accordingly based on what it observes.

> *Rules can be hard to manage, particularly at large scale, and analysts or systems are required to regularly monitor, purge, and replace old rules.*

WHAT IS A RULES ENGINE?

Backend software that executes one or more business policies based off of a set of criteria and parameters.

- Operational business logic separated from application code
- Rules managed by fraud analysts



There are three main ways rules engines work:

- The system can go through its rules, one by one, and if it determines that any rule is "triggered," it will take the appropriate action and skip the other rules.
- Rules engines can also employ weighted scoring mechanisms. Every rule will have a score value, positive or negative, which can be assigned by an analyst. The points for all of the rules triggered will be added together to compute an aggregate score. An order can then be rejected, approved, or flagged for further review, based on that score.
- Rules can be combined, and logic can be applied. For example, if a user's email is from a free service and they are posting comments very fast, you might flag them as a spammer and block their ability to post content.

As standalone fraud defenses, rules engines have limitations. More critically, they're inherently reactive—they depend on previous experience. In other words, damage has to already have happened for a rule to be written in response. Additionally, experienced analysts are needed to effectively write useful rules, and the process can be slow and time-consuming. Given the rapid pace of modern fraud, rules engines invariably fall behind. There are logistical challenges, as well. Rules can be hard to manage, particularly at large scale, and analysts or systems are required to regularly monitor, purge, and replace old rules.

If this rules maintenance process does not happen, false positive rates go up accordingly as old rules expire, become less efficient, and lose relevance. When you combine poorly performing old rules with an inability to write new rules fast enough to address new and emerging fraud threats, you end up with an untenable situation, and no good choices—it's either more fraud, or more friction.

HOW RULES ENGINES WORK

- Check against rule lists
- Employ weighted scoring
- Combine rules with logic

LIMITATIONS

- Rely on previous loss experience
- Need experienced analysts to write rules
- Require constant maintenance

RULE	WEIGHT
IP address is anonymous proxy	+800
Account age> 180 days	-500
Email is private corporate domain name	-350
Mismatch billing country and IP country	+450
Phone number found on > 3 accounts	+250

```
IF (user email = free email service) AND
(comment character count > 150 per sec)
{
flag user account as spammer mute
commenting
}
```

SUPERVISED MACHINE LEARNING

Machine learning is a branch of artificial intelligence that enables algorithms to learn from existing data and then apply that knowledge to new data. Supervised learning is the most common type of machine learning. Supervised machine learning (SML) gets its name from the fact that the process of "learning" from a training dataset is a "supervised" process. Supervised learning requires that an algorithm's possible outputs are already known and that all of the data used to train the algorithm is already labeled with correct answers. Supervised machine learning is used to discover patterns and insights from a set of data to make predictions about future outcomes.

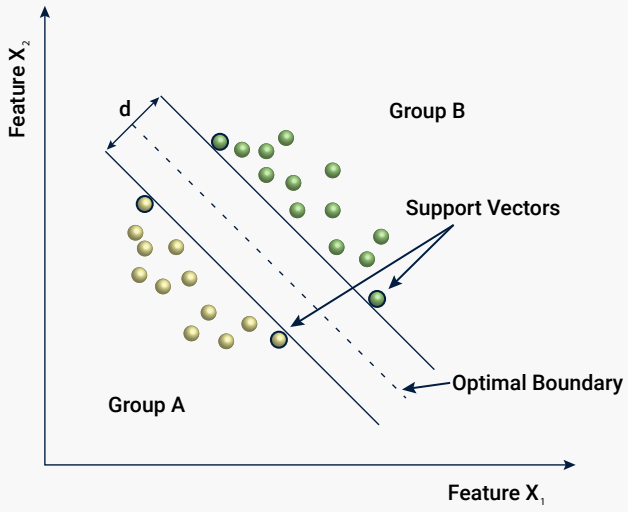
As a tool for fraud detection, SML is a powerful one. However, there are also significant limitations. For one thing, SML requires labels, and the process of getting labels can often be a matter of months. This alone makes SML a dangerously slow approach in the context of fast-evolving fraud. As with rules engines, SML depends on legacy knowledge—algorithms require known examples to learn to perform their tasks. Perhaps the biggest challenge with SML is its inability to detect new and unknown fraud attacks. Given the rapidity with which fraudsters can adapt their techniques, it is virtually impossible for SML-based solutions to keep pace.

HOW DOES IT WORK?

Algorithms learn to perform tasks from known examples ("training data"). Using supervised machine learning requires having data to train the model.

LIMITATIONS

- Labels are required; this process can take months
- Unable to detect new and unknown fraud patterns



ANOMALY DETECTION

As is self-evident from its name, anomaly detection is essentially a process of flagging outliers. Anomaly detection can function as part of an SML-based solution. For example, if an organization does not have existing labels to rely on, anomaly detection can be used to start identifying deviations from the norm within data that potentially signal suspicious accounts or activities. As these instances go on to be confirmed as either legitimate or fraudulent, labels can be created from these determinations and used to power SML models.

There are three main types of anomalies.

Point Anomaly

A point anomaly is a data object that deviates significantly from the rest of its data set. For example, if a user's online purchases are always under $100, the appearance of a $15,000 purchase would constitute a point anomaly.

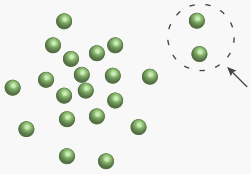
Contextual Anomaly

Contextual anomalies are data objects that deviate significantly with respect to the specific context for the objects. For example, a surge of higher-cost purchases in late spring might objectively seem anomalistic, but in the context of it being the period during which tax returns are mailed, it is, in fact, normal behavior.

Collective Anomaly

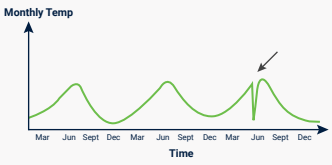
Collective anomalies occur when you have a subset of data objects that deviate significantly from an entire data set as a whole—in other words, when a group of related or linked data instances is anomalous with respect to an entire dataset. It is important to note that the individual data points may not be anomalies on their own—it is the collective that is anomalous. As an example, consider a shipping and delivery company. Individual shipping delays may be a normal and accepted part of the business. However, if a high number of delays all occur on one given day, that "collection" of delays could constitute an anomaly.

POINT ANOMALY



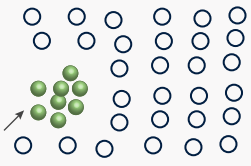
Data objects that deviate significantly from the rest of the data set

CONTEXTUAL ANOMALY



Data objects that deviate significantly with respect to a specific context for the object

COLLECTIVE ANOMALY



A subset of data objects that deviate significantly from the entire data set as a whole

In a fraud detection scenario, anomaly detection can be both a blessing and a curse. On the one hand, it's an effective approach for surfacing deviations from the norm that could suggest fraudulent or malicious activity. However, anomaly detection in and of itself can't determine whether something is actually fraud—only that something is different. On its own, anomaly detection is accordingly prone to producing high volumes of false positives, because deviations don't necessarily equal problems. The challenge of false positives associated with anomaly detection is growing as modern fraud continues to scale. In the scenario of a massive bot attack, for example, the perceived "norm" could actually be the fraud.

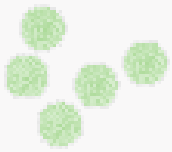
Other supervised machine approaches being applied to fraud detection include random forest and logistic regression. However, these approaches require datasets to have a sufficient number of fraud examples, and even then, threshold misadjustments can lead to increased false positives, as shown in an article titled, "Fraud Detection using Random Forest, Neural Autoencoder, and Isolation Forest techniques," which acknowledges that "privileging the decision toward fraudulent transactions produces additional legitimate transactions mistaken as fraudulent." There are concerns with other approaches, as well—as noted in a recent article from Altexsoft, overfitting is a common occurrence with random forests, support vector machine models require heavy engineering and powerful computing architecture, and neural networks suffer for a lack of human interpretability.

UNSUPERVISED MACHINE LEARNING

Unsupervised machine learning (UML) is a category of machine learning techniques that works without requiring labeled input data. Instead, UML infers a function to describe the hidden structures of "unlabeled" input data points. UML is used to discover patterns within large amounts of unlabeled data, and is especially effective for discovering new and unknown patterns.

Anomaly detection is one of three main UML approaches, with the other two being clustering analysis and graph analysis. While anomaly detection focuses primarily on identifying outliers within data, clustering and graph analysis focus on relationships and connectivity among input data.

CLUSTERING ANALYSIS



- K-means(++) efficient but difficult to pick K
- Hierarchial clustering expensive to compute
- Noisy results when dimensionality is high
  - Dimensionality reduction is the key
  - Efficiently search relevant dimensions

GRAPH ANALYSIS

- More about connectivity than compactness
- Accuracy vs. computation complexity
  - CC or SCC analysis
  - Graph cut

▶ Clustering Analysis

The challenge of detecting mass registrations offers an excellent example of how clustering analysis can be effective. Consider a scenario in which a number of users register accounts on an online platform. Individually, these registrations may appear normal and, accordingly, those wouldn't get flagged as being suspicious. However, using clustering analysis, it can be observed that these registrations share a number of attributes—they all used Google Chrome, they all signed up between 2 a.m. and 3 a.m., each of their respective GPS locations was within a mile of the others, and they all modified their account nicknames shortly after registering.

If one registration showed these attributes, there would be no cause for concern. However, if a large group of registrations all bear the same characteristics, it is likely the actions are coordinated and fraudulent. These registrations are "clustered" together accordingly, and the "cluster" is flagged as suspicious. Depending on the weight assigned the various attributes, and the score that results, the cluster might be blocked outright, or referred for subsequent manual review.

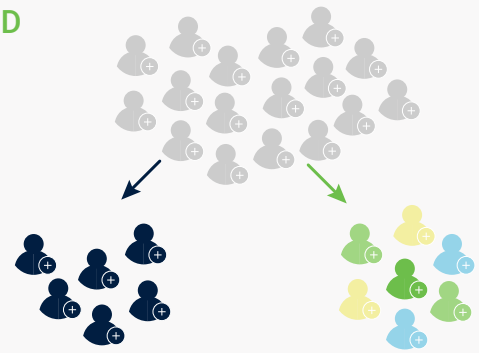
Graph Analysis

Graph analysis can be used to expose connections between user accounts that might otherwise go undetected. For example, consider a scenario in which a number of fraudsters are operating in coordinated fashion on a social platform for the purpose of spreading spam content. At face value, each social account appears normal. They like posts, they share posts, they comment on posts, and they friend different accounts. What cannot initially be determined is that all the different accounts are actually all operated by fraudsters, and they're actually only interacting with one another in order to present and build up a facade of normalcy. Slowly but surely, each of those accounts begins to post small amounts of spam content. No one account is particularly aggressive, so no flags are raised.

However, in aggregate, the group is combining to send a high volume of spam content. Graph analysis can be used to spot the connections between the accounts by detecting both explicit and implicit actions. Explicit actions could include one user sending a message to another user, and the two users friending each other. Implicit actions might include two users liking the same post. By spotting these connections, a coordinated fraud ring can be detected, despite the fact that no one user account is triggering any alerts.

APPLIED EXAMPLE OF FRAUD

- Characteristics
- Relationships
- Spam Behavior
- Grouping



New Account Registration		
• Sign-up time of day	• 3-4 a.m.	• 2-4 p.m.
• Income	• All $150,000	• $60,000 - $250,000
• Employer	• All Google or Facebook	• Google, FB, Citi, Walmart...
• Email pattern	• [Name].999@gmail.com	• Gmail, Yahoo, Hotmail...
• Application device	• All Android	• iOS, Android...
• Service provider	• All T-Mobile	• ATT, T-Mobile, Verizon...
DataVisor Score	**97**	**0**

2 Unsupervised Machine Learning: Fraud Use Cases

There are many fraud use cases to which UML can be applied, including:

▶ Application Fraud

Using UML, banks and financial institutions can analyze whole networks of applications to detect hidden connections that may appear legitimate when viewed in isolation.

▶ Bot Attacks

Using a UML-driven holistic data analysis approach, it is possible to analyze user histories, behavior changes, and suspicious patterns across millions of accounts. This enables the capture of significantly more bot-powered attacks.

▶ Money Laundering

UML algorithms can look at complex networks of transactions instead of individual ones, and can detect and eliminate launderers who deposit small denominations of funds to avoid CTR reporting.

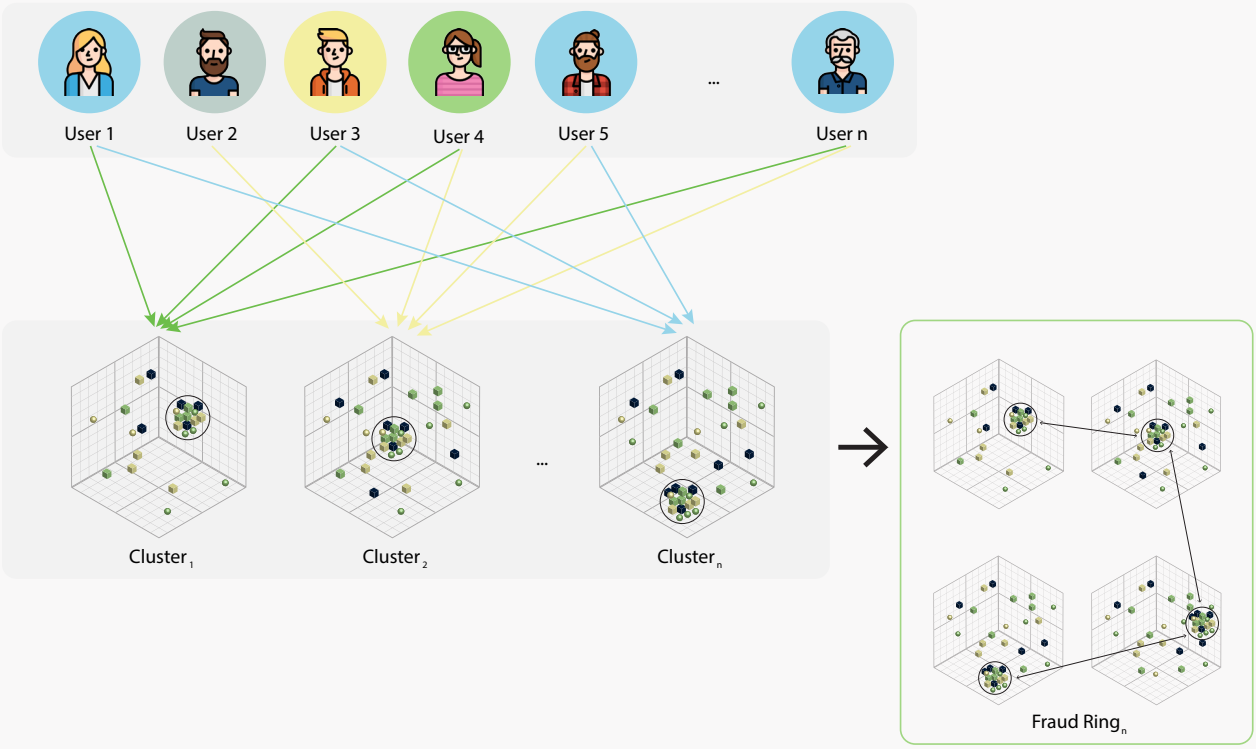
▶ Promotion Abuse

UML solutions can be used to capture all members of a fraud ring by identifying hidden linkages between fake account registrations and discovering unknown attacks before they're able to combine to fraudulently take advantage of promotions.

▶ Transaction Fraud

UML algorithms can be used to detect fraudulent accounts before those accounts can be used to conduct transactions that result in financial loss.

> *The signature advantage of unsupervised machine learning is its ability to operate without the need for labels. UML, by its nature, is proactive; it can analyze in real-time, with no prior legacy knowledge required.*



The signature advantage of unsupervised machine learning is its ability to operate without the need for labels. UML, by its nature, is proactive; it can analyze in real-time, with no prior legacy knowledge required. UML models are fundamentally self-tuning, and UML-powered solutions are free of the delays associated with SML and rules-based approaches. As such, UML is an ideal approach with which to challenge modern, fast-evolving fraud.

The use of UML-based fraud management tactics can have specific and lasting financial impact, especially when a given organization is wrestling with many different fraud types—and especially when the attacks are new or previously unknown. In the case of a recent DataVisor client—a food delivery unicorn—the transition from rules-based systems and SML models to unsupervised learning resulted in annual savings of $6 Million.

3 Why Unsupervised Machine Learning Leads to Low False Positives

The use of UML for fraud detection enables organizations to identify clusters of highly correlated users with suspicious commonalities and connections. This approach has a much higher precision rate than single-user based detection. A good user will typically evidence diverse behavior patterns, but it is highly unusual for an entire large group of users to all share those same behaviors. So while an individual instance of a certain characteristic or behavior may not be enough to confidently act upon, when we see the behavior repeated across an entire group, we can have a much higher degree of confidence about the accuracy of our results.

With unsupervised machine learning, the approach is entirely data-driven. We can mine for suspicious clusters, and we do clustering in multi-dimensional subspaces where the underlying algorithm can automatically pick features it thinks are most useful for clustering. With this approach, we don't need labels for detection. Feature engineering and selection are automatic. The ability to investigate correlated accounts all together at the group level enables teams to make accurate bulk decisions that can be applied to entire fraud rings. This significantly improves review efficiency, with more cases reviewed in less time.

> *Missed malicious activity is too costly to allow, but false positives are the enemy of customer experience. The answer to this dilemma is to adopt an approach based on unsupervised machine learning.*

The value of a UML-based approach comes from its accuracy and its precision, and the benefits are two-fold: 1) Organizations can proceed with confidence in their detection results, knowing that the actions and accounts they blocked are truly fraudulent, and 2) They can rest assured that their good customers can continue to enjoy friction-free experiences.

Missed malicious activity is too costly to allow, but false positives are the enemy of customer experience. The answer to this dilemma is to adopt an approach based on unsupervised machine learning.

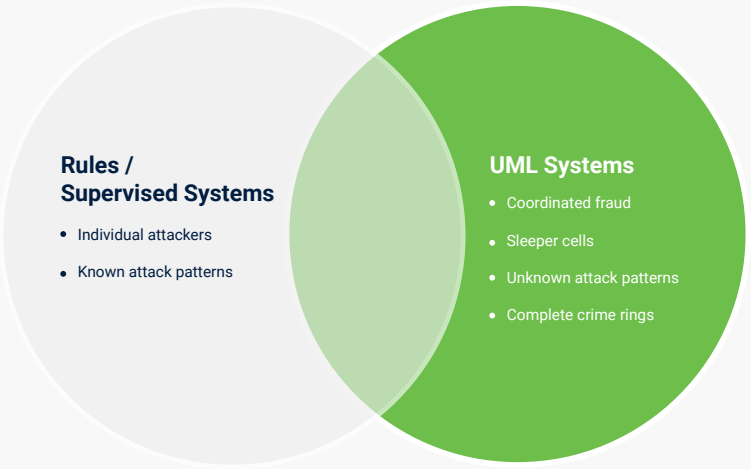
4 Comprehensive Fraud Management Platform for the Next Generation of Fraud

While unsupervised machine learning stands alone among fraud detection approaches for its unique ability to respond in real time to new and emerging fraud attacks, all the techniques we've discussed have something to offer, and a genuinely comprehensive fraud management solution will incorporate the best of everything on offer to deliver unrivaled power. Nothing less will suffice when it comes to combating the speed, scale, and sophistication of modern fraud.

THE DATAVISOR APPROACH

At DataVisor, we embrace the concept of comprehensive fraud management. Our platform integrates an extensive array of tools, techniques, and technologies that include UML, SML, rules, feature engineering, deep learning, and more. By leveraging the combined strengths of all of these different approaches, we can consistently deliver the high accuracy and low false positives that are the ideal of any fraud prevention strategy.

We embrace the concept of comprehensive fraud management. Our platform integrates an extensive array of tools, techniques, and technologies that include UML, SML, rules, feature engineering, deep learning, and more.



Using our integrated products and solutions, and relying on our comprehensive approach, organizations uncover correlated patterns and reveal hidden connections between accounts before sophisticated fraud attacks can launch.

The ability to accurately identify new and unknown attack types in real time with no need for historical data or lengthy training and retuning cycles gives businesses a defining advantage against adaptive and agile modern fraudsters. With the combined benefits our range of products offers, every business can implement and deploy the right solution for its unique scenario.

Conclusion

A truly successful fraud prevention solution achieves many goals. It works to defend a business from malicious attacks, and ensures comprehensive, proactive coverage. This results in reduced fraud losses, and increased operational efficiency. Additionally, a successful fraud solution serves to enhance customer experience, by ensuring that good customers do not experience undue delays or unwarranted friction. Finally, a successful fraud solution drives business growth and profitability, by safely opening the doors to more good customers, while keeping out bad actors.

As discussed in our introduction, false positives loom large when it comes to achieving these goals, because lowering the rate of false positives benefits all stakeholders and, ultimately, the elimination of false positives promotes a business-customer symbiosis in which everybody wins. Everybody except the fraudsters. And that's as it should be.

DataVisor is committed to providing comprehensive AI and machine learning-driven fraud solutions that empower organizations to defeat existing, emerging, and unknown fraud attacks with exceptional accuracy and unprecedentedly low rates of false positives. Our solutions organize detection results into fraud rings that enable efficient and effective analysis and action on confirmed malicious accounts.

Simultaneously, our solutions serve to enhance customer engagement by ensuring that legitimate users consistently enjoy friction-free experiences. The end results are security and efficiency, combined with growth and profitability. With DataVisor, organizations can move beyond simply protecting against bad actors, to focusing on acquiring and rewarding good customers.

About DataVisor

DataVisor is the world's leading fraud and risk management platform that enables organizations to respond to fast-evolving fraud attacks and mitigate risks as they happen in real time. Its comprehensive solution suite combines patented machine learning technology with native device intelligence and a powerful decision engine to provide protection for the entire customer lifecycle across industries and use cases. DataVisor is recognized as an industry leader and has been adopted by many Fortune 500 companies across the globe.

For more information:


info@datavisor.com


www.datavisor.com


967 N. Shoreline Blvd. | Mountain View | CA 94043