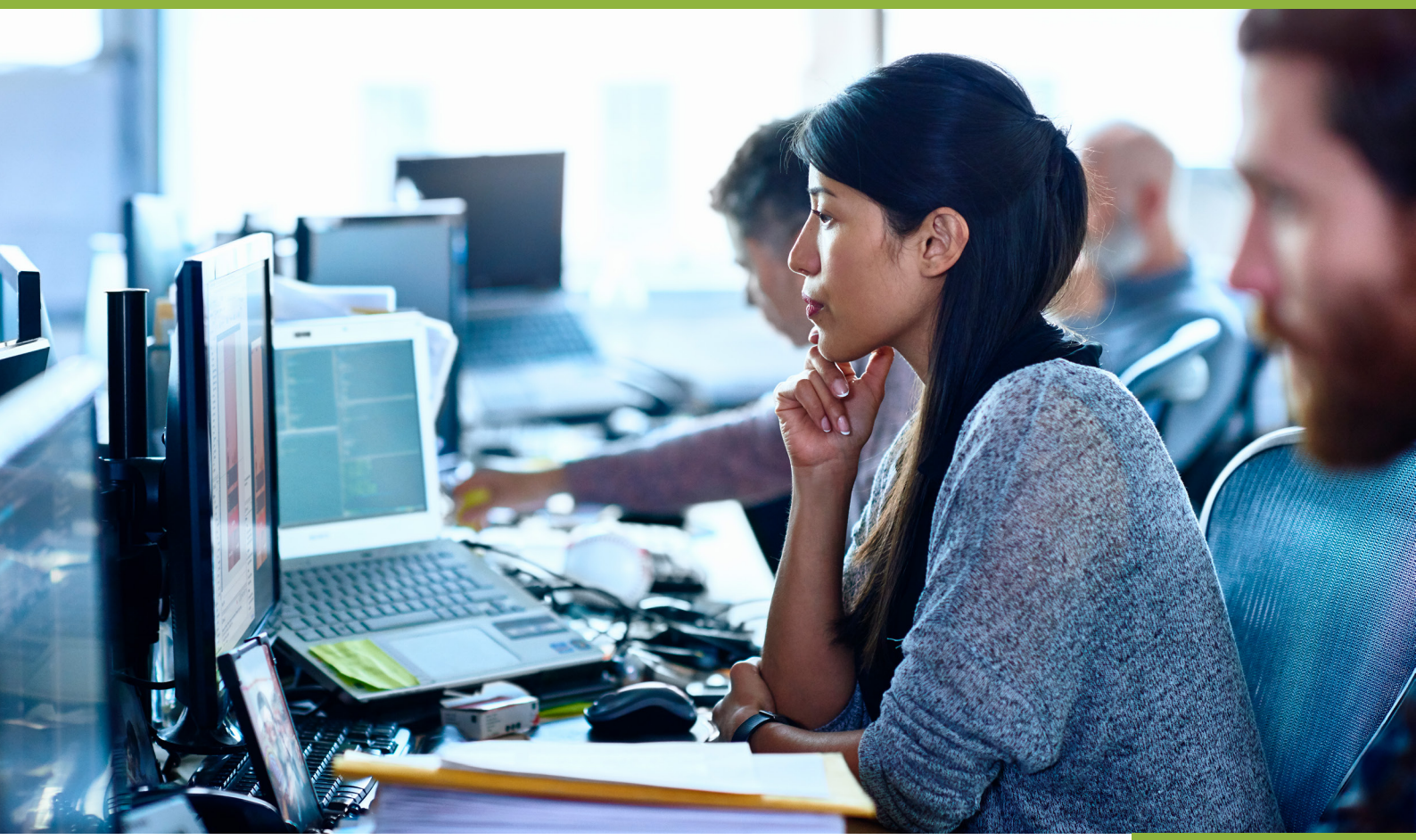
WHITE PAPER

How to Succeed With Fraud Analytics

What you need to know about data management, discovery analytics and deployment



sas
THE POWER TO KNOW.

Contents

Featuring:
James Ruotolo, CFE, FCLS, Director of Fraud and Security Intelligence Solutions, SAS
Kim Kuster, CIFA, CIFI, FCLS, Senior Fraud Solutions Architect, SAS

Is it real or is it cubic zirconia?

In 2017, a virtual consortium of 28 fake ad agencies generated at least a billion online ad impressions that pushed malicious software, tech support scams and other fraudulent schemes.

The fraudsters prepared well. They created well-rounded synthetic identities, with a website, Twitter and Facebook accounts, machine-generated content and LinkedIn executive bios for each agency.

They were clever. They cultivated relationships with legitimate ad platforms, which enabled their fake ads to reach 62 percent of ad-sponsored websites each week.

They planned ahead to foil detection. The malware analyzed users' browsers to identify machines that had elements used by security investigators. Clicks from those machines did not get redirected to malicious sites.

The scheme, dubbed Zirconium by the security firm Confiant, worked. Even one of the big three credit reporting bureaus was fooled into posting a link to a fake Flash installer.

Like the cubic zirconia gemstone – so optically flawless it can spoof a diamond – the Zirconium ads looked genuine, while creating costly havoc for users and website owners.

The perennial allure of something for nothing has fueled all sorts of innovation in online fraud. Of course, there's still the classic insurance fraud with staged accidents and fabricated medical bills. And health care fraud where medical providers exploit patients' personal information to get payments for nonexistent prescriptions and procedures. Or the lucrative bust-out scheme in financial services where a fraudster opens a line of credit for a fake identity, cultivates a good history for that account, then cashes out the big payoff and disappears.

Those tried-and-true strategies are still rampant as new threats emerge, such as cryptocurrency fraud and ransomware. Some of the largest and most cyber-savvy organizations have had their data encrypted, tied to ransom demands of $1,000 to $20,000 in bitcoin to decrypt it. Worse yet, many of the businesses who pay the ransom never recover their data.

What's the world coming to when you can't trust your adware imitators and cyber-criminals?

Fraud on the rise

In an Experian survey of consumers and businesses worldwide, almost three-quarters of businesses (72 percent) cited fraud as a growing concern over the previous 12 months, and nearly two-thirds (63 percent) report seeing the same or higher levels of fraud losses in that time. More than half of businesses (54 percent) are only "somewhat confident" in their ability to detect fraud.[1] These are sobering statistics in the growing war on fraud.

Even the rollout of EMV chip credit cards three years ago hasn't slowed the swelling tide of fraud. According to Gemini Advisory, in a 12-month period, 60 million credit and debit card numbers were stolen in the US and posted to Dark Web sites where they are routinely sold. Most of those were chip cards.

The study found that 75 percent of the numbers posted to those sites (nearly 46 million of them) were stolen from a physical point-of-sale terminal in a store, while the rest (15.3 million) were stolen from online breaches, which EMV can't protect against. Granted, part of the issue is that not all merchants have deployed the expensive chip readers, but certainly a big component of card fraud, even with EMV cards, relates to vulnerabilities in data security and online fraud detection.

The good news is that advances in data management, fraud analytics and case management are redefining the tools and tactics for fighting back — and the results.

Four flavors of fraud analytics

Retailers, banks, health care organizations, insurance companies – any large organization that could be a target for fraud – has vast resources of data that can power the defense against fraud. They just need to be able to separate signal from noise, understand those signals across disparate systems, and connect the dots to uncover activity worth investigating, now and likely into the future. That's where analytics comes in.

"This is a classic opportunity for analytics to play a significant role in this big data problem," says James Ruotolo, Director of Fraud and Security Intelligence Solutions at SAS. "The use case is ripe for transformation following improvements in hardware, real-time analytic engines and new analytic methods that blur the lines between cybersecurity, fraud management, audit and risk functions."

[1] Experian, *The 2018 Global Fraud and Identity Report: Exploring the links between customer recognition, convenience, trust and fraud risk*, Jan. 1, 2018. https://www.experian.com/assets/decision-analytics/reports/global-fraud-report-2018.pdf, accessed Dec. 12, 2018.

Ruotolo classifies fraud analytics into four flavors that represent progressively greater analytical maturity and insight:

- **Descriptive analytics** tell us what is happening now or what has happened in the past. How many alerts were generated? How many hours were spent on each investigation? What were the outcomes? How many led to Suspicious Activity Reports? How many were false positives? What was the dollar amount protected or recovered?

  These key performance indicators are essential for running the business, but once you get a handle on the current state of affairs, you want to understand the drivers behind that state.

- **Diagnostic analytics** helps explain why something happened. Find statistical outliers, correlations and clusters that point to connections and root causes. What elements of a campaign were associated with higher fraud losses? Why is property and casualty fraud trending up in this geography? Is the new digital account opening process helping or hurting overall profitability when we factor in elevated fraud risk?

  Such diagnostic insights improve planning, but a higher goal of fraud management is to predict what's likely to happen in the future, such as the odds of finding fraud in an application or claim.

- **Predictive analytics** reveals patterns among data elements that point to a high propensity for fraud. With predictive insight, fraud teams move more into prevention mode versus pay-and-chase. A data-informed view into the future helps you understand future trends and outcomes if present conditions persist.

  But what if you don't want present conditions to persist; you'd prefer a brighter future?

- **Prescriptive analytics**, or optimization, helps identify the best outcome or scenario that could happen, given a set of business constraints. Beyond forecasting, the purpose of optimization is to maximize (or minimize, as appropriate) some goal by modifying decision variables that satisfy the constraints.

  "This is the most mature state of an analytics program, because you're not just predicting an outcome; you're trying to predict an optimal outcome, given your business situation, market conditions, regulatory environment, transaction volumes, etc.," said Ruotolo. "This is about understanding what factors need to change to reach that optimal outcome, how to be the most successful with the resources you have available."

Machine learning and the trade-off between accuracy and transparency

Unlike rules-based systems, which are fairly easy for criminals to test and circumvent, machine learning adapts to changing behaviors in a population through automated model building. The system automatically creates analytical models that adapt to what they find in the data. Over time, the model "learns" how to deliver more accurate results, whether the goal is to make better credit decisions, retail offers, medical diagnoses or fraud detection. It's easy to see the value of adaptive modeling to keep pace with emerging fraud tactics.

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

- With **supervised 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. To train a supervised model for fraud detection, you present it with records associated with both fraudulent and legitimate activities, and the model learns how to predict the presence of fraud when applied to new data.
- With **unsupervised learning**, the learning algorithm isn't given any labels (dependent variables, in statistics terms). The algorithm is on its own to find structure or hidden patterns in the data. Since you don't know which data represents financial crime, you want the model to create a function that describes the structure of the data, flags as an anomaly anything that doesn't fit the norm, and then applies this knowledge to new and unseen data.

There are just two catches to the marvels of machine learning:

- **You need a lot of data.** Machine learning strategies require a significant amount of data with good classifications – for example, lots of data on previous known fraud. The computer iteratively goes through masses of data to find patterns and create highly accurate models. More data means better pattern recognition and more accurate fraud detection.
- **The resulting model is somewhat of a black box.** Because the computer has free rein to define the model with minimal human intervention, the methods used to arrive at the final model are often not that transparent to the end user.

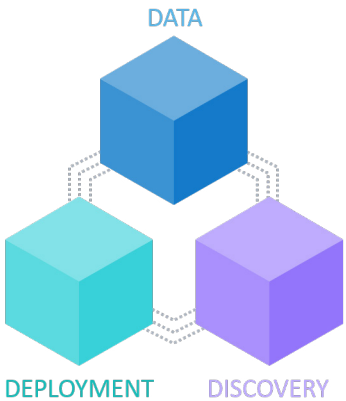
  "You may get a very high-quality score and good results, but may not understand why a file scored that high," says Ruotolo. "There's a trade-off here between being able to have the most accurate results and having a result you can articulate to an investigator, analyst, regulator or jury."

With advances in machine learning, detection systems can learn, adapt and uncover emerging patterns of financial crime – all the better to catch it before the losses are significant, or to prevent it altogether.

The necessary foundation for analytics success

While there tends to be much focus on the shine of sophisticated algorithms and modeling techniques, the real promise of fraud analytics rests on the supporting context:

- The **data** you harness for fraud detection, prevention and investigation.
- The **discovery** process that uses the right analytic techniques for the situation, alone or in combination.
- The **deployment** stage that puts discoveries into production and maintains high-quality results.



Data: Integration, quality and enrichment

The first step is the most fundamental, but one many organizations still struggle with or neglect to some degree. The most sophisticated analytics techniques cannot override the reality of "garbage in, garbage out."

A data audit is a good place to start. What data do you have? Who in the organization understands it, where it came from, the business context and the value? What trends might be found in it?

Many organizations have grown over the years through acquisitions or mergers, so they may have multiple systems even for the same function – such as several different policy, application and claims systems within one insurance company – as well as multiple external data sources. "Simply bringing the data together from those disparate systems and mapping them correctly can be a significant effort," says Ruotolo.

Then there's the issue of data quality. "If you've ever come across a claimant or customer with a Social Security number of 999-999-9999, you know what I'm talking about," says Ruotolo. "It's important to spend a fair amount of time addressing data quality, entity resolution and data enrichment up front before entering the next phase of the analytics life cycle."

Once you have mapped diverse data sources, cleansed the data and enriched it with third-party resources, you're ready to move to defining the discovery phase.

> "In talking with companies that have launched fraud analytics projects in the past, we sometimes find they moved very quickly into building a model and didn't spend as much time as they should have on data management. If you don't invest the time to ensure the necessary data quality, you're going to get poor results out the back end."
>
> James Ruotolo, Director of Fraud and Security Intelligence Solutions, SAS

Discovery: Advanced analytics and hybrid approaches

The discovery process naturally begins with basic descriptive statistics, then works its way up the analytical maturity levels described earlier – from descriptive to diagnostic to predictive to prescriptive analytics.

"We often find value in using a hybrid approach that combines a number of approaches for fraud detection," says Ruotolo. "You can start with something as straightforward as heuristic business rules, which are basically a set of if-then statements. Business rules can carry a lot of value, but by themselves they often produce a high false-positive rate. So we like to combine those techniques with other approaches, such as supervised predictive modeling – if you have enough historical data to draw on. If you don't have rich data on past fraud outcomes, you can use unsupervised modeling to look for patterns in the data and find anomalies."

Fraud teams are also capitalizing on network analysis and visualization. Network analysis plays a big role in understanding the connections (or lack thereof) among applicants, devices, open accounts and application data. For instance, does an account have authorized users who are not family members? Are payments from the same source (bank, account or device) being used to pay otherwise unrelated accounts? Visualizations of these links can be useful both in assessing applications and conducting investigations.

"You pick the right types of analytical methods, the best fit for the data you have and the problem you're trying to solve," says Ruotolo. "If you have a strong SIU [special investigations unit] with a very robust history, with lots of data from prior years' investigations, you have a good starting place from which to build powerful, supervised analytical models. If you have new systems and not a lot of historical data to train the model, this is a case where we pick a different set of methods or a different combination of methods to achieve an optimal result."

Once you have been through the process of creating a strong data foundation and building the models, the next step is deployment – putting the analytics to work.

Capitalize on the power of artificial intelligence and machine learning to automate your surveillance and detection activities.

Hybrid analytics approaches

By combining multiple analytics methods, you can find more fraud, faster, and spot emerging fraud tactics that don't resemble historical patterns. For example:

- Anomaly detection and predictive analytics can uncover new types of fraud by examining what's happening right now, not just comparing it to the past.
- Social network analytics can establish links among entities in broad context.
- Self-learning techniques, such as machine learning, take identity fraud detection to the next level, keeping pace with changing behaviors in a population.

Deployment: Deliver, prioritize and act on alerts

"Building a good model is really only half the battle here," says Ruotolo. "The key is getting that into production and putting it in front of users in a way that's useful and adds value to their work. So make sure you have a process in place for delivering those alerts, whether in a real-time or batch process.

"As the volume of alerts grows, it can become cumbersome to use the traditional spreadsheets and emails to disseminate alerts. You'll want to move more toward a dedicated user interface that supplies all the information needed to make a decision about triaging an alert."

Then capture the dispositions and outcomes. The ultimate findings of any investigation, whether fraud was found or not, should be fed back into a learning cycle of continuous improvement.

Eight realities to consider when planning a fraud management program

1. Big data is a boon, if you can wrangle it.

Many organizations are just now catching up to the challenges of big data. The rapidly growing volume, variety and velocity of data streams often strains the systems designed to ingest and digest it. At the same time, executives and consumers are putting conflicting pressures on organizations. They expect:

- Greater use of data to make more accurate decisions to reduce fraud.
- Greater speed in approving transactions and reducing friction for the user.
- Greater protections for the privacy and security of their data.

Even the most mature organizations struggle to strike the right balance between expediency and protection.

2. There's an explosion of new data sources.

Increases in data volume and variety are driving better fraud detection and prevention. Most organizations already analyze data from traditional sources – such as claims systems, retail transaction systems, internal billing systems, third-party medical billing, policy data and more – from all the lines of business. Now organizations can also bring in data from new, nontraditional sources such as industry and government databases and from Internet of Things (IoT) devices in homes, businesses and vehicles

"A lot of companies I talk to are looking to bring in data from novel sources, such as remote device sensors, weather data, digital maps, online transactions, mobile apps, social media, aerial imagery from drones and more," says Ruotolo. "With the push for telematics and usage-based insurance, for instance, a fraud program has to address new data sources that can have a real impact on scoring, prediction and investigation."

Adding value to investigators' time

"People ask, 'If I do a really good job with this technology, am I going to put myself or my investigators out of a job?'" says Ruotolo. "We typically find the opposite.

"More accurate detection makes investigators more valuable, because they can focus on the alerts that really matter and deliver stronger results. Even with the efficiencies they gain with analytics, a lot of companies end up adding headcount because they've found so much suspicious activity that had previously been flying under the radar.

"For example, one company looking at its medical provider network knew there was big risk and a lot of organized fraud activity, but they couldn't identify it. They implemented fraud analytics with the goal of finding one case of fraud a month. In the first three months they found about 60."

3. There is gold in all that unstructured data.

"If you're making decisions on only the structured data fields in your systems, then you're really making decisions with less than 20 percent of the available data," says Ruotolo. "To reach the level of accuracy we're after, it's critical to consider unstructured data – text, audio and images."

The technology is improving at such a rapid rate that companies can extract significant value from such data sources as high-resolution drone images, National Insurance Crime Bureau (NICB) aerial imagery after catastrophic events, text comments in claims forms and customer service records, and audio recordings of phone calls.

This trend is driven by improvements in graphic processing unit hardware, analytical methods such as cognitive analytics and artificial intelligence, and the maturity of optical character recognition, speech recognition and image analytics. High-resolution satellite, drone and vehicle imagery will continue to accelerate this trend.

4. It's time to speed the pace of fraud detection.

"The dwell time – the time between a hacker's entry into a system to being identified – currently averages about 196 days," said Ruotolo. "That's months and months that a hacker has to troll through your network and find interesting information before getting caught. I think we can all agree that 196 days is far too long."

The goal isn't to stop people from getting into the network, Ruotolo asserts, especially since some of the threats come from insiders. "That perimeter game is over. We've all lost, and the bad guys will eventually get in. The goal is to identify them as quickly as possible when they do make it in, before they can do a lot of damage or steal a lot of data." This calls for close collaboration across fraud, cybersecurity and risk teams in the organization.

5. The cloud is clear.

"Historically, privacy-conscious enterprises such as financial services institutions were apprehensive about moving services to the cloud, primarily due to security concerns," says Ruotolo. "Companies are now realizing that cloud providers such as Amazon Web Services or Azure are probably spending a lot more on data center security than their company ever would. So having a company hosting them is not necessarily any greater risk than keeping it inside their own four walls.

"In fact, a number of customers that in the past absolutely refused to allow their data outside their internal networks are now coming back to us saying their go-forward strategy is 100 percent cloud. They don't want any new projects on premise."

Several prominent organizations have announced goals to have a zero data center footprint in the near future, with all their data hosted in the cloud by third parties. "They see the cloud as the path forward and the way for them to be successful and profitable in the future," says Ruotolo, "and a lot of organizations are following."

If the cloud is in your organization's plan (and it probably is), consider the options it opens up for the big data and high-performance processing for fraud analytics.

> "That perimeter game is over. We've all lost, and the bad guys will eventually get in. The goal is to identify them as quickly as possible when they do make it in, before they can do a lot of damage or steal a lot of data."
>
> James Ruotolo, Director of Fraud and Security Intelligence Solutions, SAS

6. Be aware (or beware) of the Lemonade effect.

Founded in 2016, Lemonade Insurance Company created a new business model for insurance based on behavioral economics and technology. The company uses artificial intelligence and chatbots to deliver insurance policies and handle claims for its users on desktops and on mobile devices through its iOS and Android apps. No brokers required.

Lemonade customers can take out a policy on a smartphone, submit a claim by snapping and sending a photo from their phones, and receive payment within minutes. "They're setting a new bar for the way business is done, and the industry is reacting to that," says Ruotolo. "For consumers, Lemonade has created a different mindset around the art of the possible. Some insurers are trying to implement their own analytic methods to compete, while others are pushing back against it. It will be very interesting to keep an eye on this as things progress over the next few months and years, as companies figure out the right way to use technology in their organizations."

This trend has critical implications for fraud programs in all industries. For one, the anonymity of digital accounts opens the door for new forms of fraud. The expectation for frictionless, speedy processing restricts authentication options and tightens the window for due diligence and fraud detection.

7. It's not a 'set it and forget it' thing.

"I worked at an insurance company that (until it implemented analytics) had been using the same manual red flag list, kindly provided by NICB, for probably 20 years," Ruotolo recalls. "There were some tweaks and updates every now and again, but by and large, we'd been using the same list of things to look for to identify suspicious claims. We'd give the same training to our adjusters every year. The mindset was, 'This is the way it's always been done.'"

That mindset is yielding to a posture of continuous improvement, Ruotolo says. "Fraud analytics is not about buying something, putting it in place, and then walking away. It is something that needs to be constantly updated and nurtured. Fraud patterns change over time, the bad guys modify their methods to continue to slip in under the wire, and we need to be doing the same thing, updating our detection techniques."

8. The investigation unit of tomorrow will look very different from today's.

"When I started in SIU, a team would be primarily staffed with field investigators with a lot of law enforcement background," Ruotolo says. "They were great at being out in the field, taking statements and doing the legwork of investigations. As time went by and we started to add analytics, SIU teams brought on different types of analysts, skilled in desktop research and investigation."

The trend has only accelerated. SIUs will rely more on analytics and data science skills. Effective, evidence-based storytelling will be a highly sought-after skill. Instead of hiring former law enforcement officers who are well versed in the SIU experience, you're likely to be hiring more quantitative PhDs out of universities and training them in how SIU investigations work.

As companies evolve the digital experience, they see online behavior patterns change, and traditional markers that might signal fraud are no longer reliable.

Six ways to make the business case for fraud analytics

Here's your shorthand script to make a business case for funding a fraud analytics project:

Volume. "There's more fraud out there, and we're not getting enough referrals from our adjusters. We need to implement the technology to do more and find more of these cases with the same resources."

Accuracy. "We're getting a lot of referrals, but many of them are false positives. They're not good leads; we're not conducting investigations on them or getting quality outcomes, so we need to improve the accuracy rate."

Efficiency. "Our analysts and investigators shouldn't have to go out and pull data from 16 different systems, when they could have it all presented to them on one screen. The time to triage a case and generate an alert can be reduced from two days down to 20 minutes, and we can do a much greater volume of casework with the same staff."

Expediency. "If we can get involved earlier on and more quickly identify suspicious activity, we have a better chance for a higher-quality outcome. I can show you a case where that dwell was reduced from 68 days to two days."

Prevention. "By acting faster, we can prevent many fraudulent payments from being made instead of taking a 'pay and chase' approach."

Defensibility. "When explaining our decisions to regulators or legal teams, we will have a great answer – our decisions were made by a consistent, algorithm-driven process that is applied universally across the board, removing human bias."

Fraud analytics in action

Fraud analytics have delivered notable results in financial services, retail and other industries. Here's an example from a major property and casualty insurance company.

Scenario and challenges

- Concern that certain medical providers were operating in collusion.
- Lack of awareness of potentially problematic individual claims.
- Lack of transparency into the destination of payments made through attorneys.
- Data inconsistencies across billing, claims and policy systems.
- Reliance on rudimentary rules that yielded excessive false positives.

Fraud analytics solution

- Integrated the internal and external data scattered across multiple source systems.
- Performed entity resolution to resolve multiple instances of the same identity.
- Enriched the data with additional sources from NICB and cross-industry databases.
- Created scoring models for both claim-level and medical provider-level alerts.
- Used known fraud history to train a highly accurate predictive model.
- Presented insights to analysts and investigators in a comprehensive, interactive user interface.
- Provided intelligence to better triage alerts and investigate claims and entities that needed attention.

Results

"The results were pretty staggering," says Kim Kuster, Senior Fraud Solutions Architect at SAS. "Using some of the unique skills of its SIU team, the customer made great impacts throughout the organization while gaining and strengthening its analytic maturity." Some tangible results:

- Reduced false positives and improved detection, resulting in a 74 percent alert acceptance rate.
- Flagged more than $7.5 million in previously unidentified questionable personal injury protection claims and medical provider exposure.
- Improved major case identification, identifying 162 percent more suspicious providers.
- Also flagged 16 percent of accepted alerts for medical management issues, demonstrating value beyond fraud detection.
- Nearly doubled SIU capacity without abandoning manual referrals.
- Increased referrals to underwriting for policy concerns, such as one driver with 16 listed vehicles.

> "There's not an SIU team out there sitting on their hands waiting for work. This company's SIU was busy before fraud analytics came into their shop, but with analytics they were able to trim down the white noise they may have been working previously and focus on activities that are a value-add back to the business."
>
> Kim Kuster, Senior Fraud Solutions Architect, SAS

Closing thoughts

Leading-edge analytics and hybrid modeling techniques find threats faster and more accurately. Embedded machine learning enhances every step of the process – continuously fine-tuning detection algorithms, streamlining and automating case management, and boosting overall performance.

By applying advanced analytics and powerful machine learning on a unified platform, your organization can detect more financial offenses, reduce false positives and run more efficient investigations.

Learn more at sas.com/fraud.

The insights in this paper were drawn from a presentation to the International Association of Special Investigation Units (IASIU), "Analyzing Fraud Analytics," in October 2018.