ESCAPE THE ANALYTICS SOFTWARE ‘BLACK BOX’

New developments in forensic data analytics are breaking down traditional barriers to access, transparency and usability. You don’t need an advanced degree to use “data blending” software.

You’re the regional compliance or internal audit manager of a global manufacturing company responsible for monitoring high-risk transactions and investigating suspicious or noncompliant activities. For years, you’ve relied on spreadsheets, database queries and random sample selections to scrutinize data and select areas for testing and due diligence. But you’re now so overwhelmed reacting to whistleblower hotline calls coming down from headquarters that you rarely have time to think about proactive anti-fraud data analytics procedures. Sometimes, you feel like you’re drinking from a firehose and wish you could be more proactive than reactive.

As a CFE, you know how to analyze data almost as well as you know how to interview people. This is particularly true with financial data such as payments, sales and commissions, and travel and entertainment, among other sources. You’ve even had your fair share of third-party due diligence and compliance reviews and the occasional email investigation where you found that smoking-gun document.

You also hear about digital transformation within your industry, but you wonder how it practically applies to you when you’re measuring risk management, compliance and business integrity. You’ve heard war stories of how difficult it is to extract, normalize and integrate data within your organization so you can properly analyze it. “There’s simply too much data out there,” you tell yourself. “And even if I could get it centralized, there’s no way to cost effectively synthesize that data into something meaningful where I could take action.”

You fear that any commercial data analytics (or business intelligence) solution is just a confining, rigid “black box” — where you don’t fathom the algorithms, the math is too complex and
the results show too many false positives, such as potentially bogus invoices or vendors.

However, this model has changed.

**Digital transformation will impact your role**

Ten years ago, or even five years ago, most fraud examiners would’ve agreed with your frustrations. The traditional anti-fraud analytics mindset was to spend lots of time and money integrating all the data sources across the enterprise and hope that expensive black-box commercial software would point you to anomalies and suspicious activities.

Highly specialized anti-fraud software platforms, which require equally specialized human expertise, have given way to much more simplified analytical tools that can do the same, if not more, sophisticated analysis, and you don’t need an advanced statistics degree.

“Companies of all sizes recognize the tremendous potential for data, but many struggle turning that data into actionable insights,” said Dean Stoecker, CEO and founder of Alteryx, a software company. “Understanding data to make imperative business decisions can no longer be the responsibility of one role, one team or one department. Organizations need to empower every data worker — regardless of technical acumen — to advance their data science and analytics skills quickly in either a code-free or code-friendly environment.”

**The black box is dead**

Expect more and demand transparency in your digital analysis process. There’s no excuse for blindly trusting the model that purports to identify high-risk transactions, employees, customers or third parties. “When organizations look at transactions, it’s critically important that they identify the specific triggers that are driving the risk scores and what’s influencing the predictive models,” says Todd Marlin, CISSP, CCSP, CISM, principal with Ernst & Young LLP’s Forensic & Integrity Services and leader of the EY Global Forensic Data Analytics team.

“Transparency is key,” Marlin says. “CFEs should evaluate transparency and analytics considerations, including access rights to the data you’re using, other data sources that have been combined with it and the assumptions you’ve made.

“Do the technical decisions you’ve made match the business intention?” Marlin asks. “Depending on the approach and objectives of the technology (e.g., transaction monitoring, email analysis, etc.), your responses to these areas will result in greater transparency — particularly for those anti-fraud professionals who aren’t technology focused.”

**From solutions to insights**

A traditional “solutions approach” to anti-fraud monitoring is based on the premise that a single, all-encompassing software package will be the end-all-be-all to whatever asset misappropriation, corruption or financial misstatement fraud scheme might arise. If you install it, you’ll find fraud.

However, with recent improvements in analytical capabilities that enable analysis of increased data volumes and complexities, organizations now should follow an “insight approach.” They should seek agile data analytics capabilities to model known risks and develop the insights to discover unknown risks. “Today’s technology platforms are far more open, object-oriented and visual, as compared to command-line and code-driven platforms,” Marlin says.

Your objective to increasing business transparency and improving your integrity culture (see “What’s your integrity agenda?” Fraud Magazine, September/October 2018 tinyurl.com/yc37gmtz) is to not just provide business intelligence solutions to known, historical fraud risks — aka, looking backwards — but to season your program to predict future fraud risk areas and prescribe insightful, timely mitigating activities to the user.

**High-impact use cases**

Here are some examples of how organizations used leading forensic analytical techniques to provide greater business transparency into their corporate data to help prevent and detect fraud or non-compliance activities.

**Data blending and predictive modeling to spot fake customers.** A large consumer products company saw a significant increase in whistleblower complaints related to bogus or fake customer schemes in a Latin American country. Sales personnel allegedly were creating fake customer accounts to meet...
sales targets and taking advantage of sales and marketing discounts to cover up missing inventory.

A team comprised of internal audit, compliance and IT — plus an external forensic analytics team — used data-blending software that helped integrate multiple data sources into a single repository for more efficient analysis.

In this case, the company integrated multiple sales and customer-related data sources to paint a more robust picture of sales activity. Next, they deployed machine-learning techniques to develop predictive models that successfully identified the profile — key attributes — of a potentially bogus customer, based on known characteristics of historical bogus customers from previous investigations.

In approximately eight weeks, the teams developed a predictive model that achieved a 96 percent accuracy rate and identified a bogus customer, based on key high-risk variables (or attributes) identified in previously known bogus customers. Internal audit and compliance deployed this model on site at the company to quickly identify a potentially bogus customer in its system and followed up or investigated the sales representative associated with the potentially bogus customer. The company plans to expand the model to other countries that could save millions of dollars annually in unauthorized incentive and marketing expense payments.

**Text mining and social media.**

Consumer banking-account scandals and regulatory inquiries have led to a substantial increase in banks monitoring front-line banker activities. A banking company recently mined Twitter, Facebook, Glassdoor and other leading sites to identify high-risk, employee-related activities that violated the company’s code of conduct. This text-mining technique identified risk areas by region, district and banker, and provided transparency into high-risk behaviors, such as abusive language or negative remarks that could put the bank’s reputation at risk.

The bank now uses these techniques to better supervise and train bankers on their code of conduct, which is helping to mitigate reputational risks, among other factors.

**Robotic process automation for timely anti-fraud monitoring.** Data collection is the systematic approach to gathering and measuring information from a variety of sources. “Data munging” is the process of cleaning, parsing and proofing data to accelerate the process of gaining business insights. A recent study by CrowdFlower found that data scientists spend 80 percent of their time on data collection and munging.

(See “Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says,” by Gil Press, Forbes, March 23, 2016, tinyurl.com/ya72kfq.) Wouldn’t that time doing data munging be better spent on actual analysis and providing insights?

A pharmaceutical company recently used robotic process automation (RPA) to automate many of the data munging tasks such as:

- Continuously requesting data feeds from systems and consistently compiling them into a single “data lake” that could be used for a wider variety of internal audit, compliance and legal-use cases.
- Evaluating the quality of data and performing automated cleansing steps through statistical modeling to rapidly source data discrepancies, such as incomplete date or dollar-value fields.
- Notifying the users in the organization of source-data discrepancies via automated emails.
- Staging data and creating automated alerts to users when certain key tasks were complete, such as a desktop notification when the latest batch of data was available for review.

The pharmaceutical company eliminated hundreds of hours of manual work each month by automating the data-refresh process via RPA and increasing the frequency of updates to the analytics and reports for more timely and relevant compliance monitoring.

**What can anti-fraud professionals do now?**

When you objectively evaluate your anti-fraud and compliance monitoring program, management encourages you to ask bigger, more insightful questions of your organization’s data. As with any
business process or control activity, here are some questions to ask:

• What are our most significant ethics, integrity and compliance risks?
• Who’s accountable for managing them?
• What are they doing?
• Are they working?
• How do we know? (aka, show me the data)

What are some new questions you could ask if you combined multiple data sources or deployed more advanced analytics? Could you be more proactive in your investigative activities and use analytics to help prevent and detect fraud from happening in the first place? Where could you use predictive modeling to spot fake customers, vendors or employees? Could you analyze past litigations to build a predictive model of what risk and cost factors are likely to be incurred with new cases? What’s the dollar impact to the business and what legal and compliance risks would you mitigate by increasing your business transparency?

Asking bigger, more insightful questions of your data is now quite attainable with easier-to-use tools that’ll help you escape out of that software black box. Now go and find that hidden money in your organization.

Vincent M. Walden, CFE, CPA, is a partner with EY’s Forensic & Integrity Services practice. Contact him at vincent.walden@ey.com.

Todd J. Marlin, CISSP, CCSP, CISM, is a principal with EY’s Forensic & Integrity Services practice. Contact him at todd.marlin@ey.com.

© 2018 Association of Certified Fraud Examiners, Inc.