Data analytics: Data in motion tends to stay in motion

By Matt Reeder and John Kim

Matt Reeder (mtreeder@orrick.com) is an Associate with Orrick Herrington & Sutcliffe LLP in Washington, DC. **John Kim** (john.kim@controlrisks.com) is a Director with Control Risks in Washington, DC.

This is the second article in a two-part series.

In the first part of our series, "Developing a data analytics–enabled compliance program for the real world,"[1] we offered a three-step method for building data analytics–enabled compliance systems incrementally upon existing data sets and capabilities. Now, we turn our attention to how data analytics can enhance compliance programs that have inventoried their current capabilities, identified useful data sets, and mobilized their resources to execute an established plan.

The hallmarks of an effective data analytics–enabled compliance program are adaptability and specificity. The more adaptable a data analytics program is, the more readily it can integrate new data sources, respond to new regulatory or legal requirements, and be applied across changing business practices. Taking a multitiered approach to data analytics and implementing a continuous feedback loop will ensure an appropriate level of adaptability. Data analytics outputs must be sufficiently specific. This specificity ensures that effectiveness is measurable. Vague evaluation criteria or outputs based on loose correlations do not yield actionable information.

With these hallmarks in mind, we describe three data analytics techniques that will empower a mature, data-enabled compliance program to apply data analytics more effectively. They are rules-based tests, statistical and trend analyses, and machine learning. Understanding these techniques will allow compliance professionals to work toward incrementally adopting a multitiered approach to data analytics.

Each of these techniques has unique costs and benefits, but they can work together. Adopting all three maximizes detection rates, minimizes false positives, and marshals more useable data in service of the compliance function. Furthermore, applying all three techniques creates a virtuous feedback loop that fosters adaptability. This full-fledged application of data analytics creates momentum for the compliance function that augments, amplifies, and multiplies the effects of the more traditional components of a compliance program.

Rules-based tests

Rules-based tests are often small pieces of computer code (scripts) meant to identify behaviors, characteristics, or actions. They are typically designed to signal a specific behavior, the presence (or absence) of which may require the attention of—or correction by—a compliance professional. These behaviors can then be coded as procedural rules that are structured in an "if/then" format that targets a specific transaction set or behavior type.

Rules-based tests find information such as:

- Expense amounts below—but within a certain range of—an approval threshold,

- Duplicate names or addresses between vendors and employees, and

- Duplicate or sequential invoice numbers.

Satisfying one of these rules-based tests does not by itself indicate misconduct. But misconduct is often

accompanied by behaviors that do satisfy such tests. Thus, when aggregated and included in compliance monitoring or testing workflows during red-flag testing, rules-based tests can reveal patterns and trends that merit further inquiry or investigation. Each positive can be flagged, and the totality of these red flags can offer insights into risks across an employee group, within a business unit, relating to a specific transaction type, in a limited geographical area, etc.

Benefits

- Rules-based tests are proven and effective tools for harvesting the low-hanging fruit in existing data streams.

- The tests are often easy to implement.

- Learning to understand and use the results of rules-based tests is fast, simple, and cheap.

- Compliance professionals can collaborate directly with their IT departments and internal stakeholders to develop, adopt, and deploy these rules-based tests on an ongoing basis.

Drawbacks

- The tests are often limited to known and observed behaviors, characteristics, and actions.

- The known behavior that a subject matter expert recommends as a test criterion is likely a behavior that a would-be bad actor would know about and could use to game the test or avoid altogether.

- Rules-based tests are rigid and therefore cannot "learn" to identify rule-avoidance behavior. Thus, relying too heavily on rules-based tests can set the stage for a game of compliance whack-a-mole that involves an endless series of time-consuming and labor-intensive refinements.

- Since rule-based tests run on broad swaths of enterprise data, false positives can create significant noise in the compliance monitoring signal that diminishes the testing data's usefulness.

Statistical and trend analyses

Statistical and trend analyses consider fixed data points across a separate variable data set, like transaction amounts across time, or transaction type across amounts, or rules-based red flags across departments or job functions. Creating such an analysis requires the combined efforts of a compliance professional and a statistician to observe behaviors and form and test various hypotheses. Viable hypotheses can be translated into a data-based testing procedure that automatically identifies defined patterns and highlights anomalous data points.

Examples of statistical and trend analyses:

- **Benford's law (Newcomb–Benford law)**. The statistical law of anomalous numbers that states that the frequency distribution of leading digits in many real-life data sets follows a commonly observed, logarithmic pattern.

- **Time series analysis**. Analyzing patterns of behavior through time by normalizing data sets to account for seasonality.

- **Distribution and standard deviation models**. Understanding an underlying population of data and determining outliers based on the dispersion or variation in a set of values.

These tests inform a compliance professional's understanding of a data population's underlying characteristics. They can also detect patterns and behaviors that are imperceptible by manual human review or rules-based testing. These analyses do, however, require a degree of mathematical and scientific expertise not always

resident within a traditional compliance department.

Benefits

- Statistical and trend analyses improve detection rates and diminish false positives compared to rules-based testing. Consequently, they can identify behavioral patterns that do not follow rigid, linear patterns.

- Because these analyses rely on multiple and variable data ranges, bad actors cannot easily "game" them, because there are no easily identifiable rule sets.

Drawbacks

- Setting up effective statistical and trend analyses takes more manual effort from and close collaboration between compliance professionals and a statistician. Furthermore, because setup requires a redundant process of observing, hypothesizing, and testing, this collaboration can be time consuming. Such lead time is not without cost to the organization.

- While these analyses offer additional insights into sometimes hidden data patterns and employee behavior, they still share a degree of the rigidity that can limit the utility of rules-based testing. Like rules-based tests, statistical and trend analyses cannot learn behaviors and adapt. Thus, refinement and fine-tuning are essential to successful statistical and trend analyses and to creating an ongoing resource demand.

Machine learning

Machine learning combines sophisticated analytical tools and human input to more fully automate the process of targeted issue identification that learns to identify the data points underlying behavioral patterns indicative of misconduct. The inherent complexity of applying machine learning techniques in a compliance workflow often requires collaboration with a data scientist. The best resources for developing machine learning models often are the data points and investigative findings from rules-based test results and statistical and trend analyses. These machine learning algorithms can assimilate these data and reduce false positives to better correlate positive results with prohibited conduct.

Examples of machine learning techniques:

- **Random forests.** Builds multiple iterations of decision trees (probability distributions based on course of action) and uses consensus across the iterations to classify and categorize the data.

- **Segmentation and clustering analysis**. Groups "like" transactions and classifies transactions based on shared attributes and similarities.

Machine learning algorithms can help compliance professionals detect high-risk behaviors that are preparatory to misconduct. A successful machine learning model can outmaneuver bad actors who might succeed at defeating a data analytics–enabled compliance program relying solely on rules-based tests and/or statistical and trend analyses. Certain machine learning techniques, like segmentation and clustering, also can prioritize employees, vendors, transactions, or other data sources by magnitude of risk in order to help compliance professionals focus attention where it is needed.

Benefits

- Unlike the first two data analytics techniques we have discussed, machine learning tools thrive on complexity.

- Machine learning tools operating on rich data sets can become highly accurate with remarkably low false

positive rates.

- Highly flexible algorithms allow the tool to learn suspicious behavior and stop bad actors from gaming the compliance system.

- Unlike rules-based tests and statistical and trend analyses, machine learning tools improve over time without added human intervention.

- They require maintenance, but do not need constant refinement and fine-tuning.

Drawbacks

- Developing a machine learning capability demands substantial upfront investments of time and capital. Data scientists, compliance professionals, and statistical analysts are all essential to the process.

- The learning curve is steep, and any expertise gap or breakdown in collaboration can create significant difficulty.

- The output of a machine learning model is only as good as the data it ingests. As such, high-quality and sometimes hard-to-obtain data sets are necessary.

Conclusion

Regardless of which analytics technique or techniques a company adopts, compliance professionals must diligently monitor and improve their tool sets and models. Human input and refinement will ensure that detection mechanisms remain accurate, that scripts and algorithms are trained on the right compliance targets, and that the data analytics outputs serve the needs of the compliance mission. It is this human/data interaction that ensures that the data analytics–enabled compliance program is sufficiently adaptable to the ever-changing requirements of today's compliance environment. Data analytics are not a set-it-and-forget-it solution. Rather, they augment, amplify, and enhance trained compliance professionals' ability to identify, detect, prevent, and mitigate misconduct.

Taking the following steps at regular intervals will ensure an appropriate level of adaptability in a data analytics–enabled compliance program:

1. Gather analytics results;

2. Segregate testing population from results;

3. Review and investigate testing population for false and valid positives;

4. Analyze accuracy of testing population and identify commonalities in—and differences between—valid positives, valid negatives, and false positives;

5. Identify additional or different testing criteria to improve validity rate; and

6. Tune and refine test criteria to improve results.

Data analytics can help compliance professionals identify red flags, mitigate risks, prevent misconduct, quickly remediate breaches, and, if the enforcement authority comes calling, demonstrate a level of diligence that reduces punitive exposure and avoids costly reputational harm. An adaptable and carefully tailored data analytics program can create momentum for the compliance function that builds on itself. If properly resourced and maintained, such a program will pay dividends for years to come.

Takeaways

- Proceed incrementally to build upon existing data analytics capabilities.

- Ensure that data analytics are specifically tailored to your organization's needs; articulate the goal of your data analytics program simply and concisely.

- Remember that a simple and effective use of data analytics will generate better results than a grand but unwieldy one.

- Maintain adaptability in your data analytics program by taking a multitiered approach and establishing a continuous feedback loop.

- Monitor, reassess, and improve your data analytics program over time.

1 Matt Reeder and John Kim, "Developing a data analytics-enabled compliance program for the real world," *CEP Magazine*, April 2020. https://bit.ly/2zuTfSc

This publication is only available to members. To view all documents, please log in or become a member.

Become a Member Login