

Next-level data analytics for fraud prevention and detection

Prepared by:

Chris Fitzgerald, Manager, Financial Investigations and Disputes, RSM US LLP
chris.fitzgerald@rsmus.com, +1 212 372 1904

Kyle Marinko, Supervisor, Financial Investigations and Disputes, RSM US LLP
kyle.marinko@rsmus.com, +1 212 372 1420

March 2020

Today's organizations generate and maintain tremendous amounts of data, including detailed transactional logs, in-store and online traffic data, subscription information, geographic and demographic datasets, social media activity, and more. Organizations with strong business intelligence functions often reach beyond internally generated data to mine third-party databases to further enrich analysis and derive meaningful insights.

Such approaches are increasingly prevalent, in part due to the ease of use and adoptability of new, powerful analytical tools and technologies such as Alteryx, Power BI, Tableau, R Analytics and SAS. Increased speed and efficiency can lead to significant cost reductions, and deeper insights can drive the development of new products and services; optimize operations and financial performance; and identify indications of fraud, waste and abuse.

The spectrum of data analytics ranges from descriptive to predictive to prescriptive, with each approach generally delivering more value (with greater implementation effort and complexity) than the last. These three approaches can be described as follows:



Descriptive analytics answer the question "what has happened and why?" This type of analytics looks at historical data in an effort to determine the reasons for past successes, failures or missed opportunities. Such analytics are often utilized by the financial planning and analysis functions of organizations to summarize financial reporting and key performance. Organizations can leverage dashboards and customized reports to understand performance and inform management decisions.

Predictive analytics answer the question "what will happen?" These analytics assess data patterns, trends and relationships among variables to identify possible future outcomes, and set goals for the organization. Companies can use statistical and machine learning algorithms, sentiment analysis, data mining, forecasting, simulations, regressions and pattern identification to help design an effective forward-looking business strategy.

Prescriptive analytics answer the question "how can we make it happen?" As the logical next step to predictive analytics, prescriptive analytics uses a combination of business constraints and data to analyze possible outcomes,

and suggest the best outcomes for the business based on simulations. Tools such as mathematical programming, optimization, decision trees and scenario analysis can have a major impact on business growth and efficiency.

Using data analytics to identify fraud, waste, abuse and corruption

Data analytics should be a key weapon in every organization's fraud protection arsenal to guard against loss and avoid being vulnerable to regulatory enforcement. Former Deputy Assistant Attorney General Matthew S. Miner recently made the following statement as it relates to the U.S. government's use of data analytics to identify fraud and the importance and expectation of companies proactively and reactively utilizing data analytics to detect and deter fraud:

[F]raud prosecutors are able to leverage data analytics to identify fraud indicators within Medicare claims data. This use of data analytics has allowed for greater efficiency in identifying investigation targets, which expedites case development, saves resources, makes the overall program of enforcement more targeted and effective. Health care isn't the only arena where data can signal indicators of fraud... we are now approaching enforcement, particularly in the commodities arena, around a data-driven approach... Whereas we are able to identify indicators and anomalies from market-wide data, companies have better and more immediate access to their own data. For that reason, if misconduct does occur, our prosecutors are going to inquire about what the company has done to analyze or track its own data resources—both at the time of the misconduct, as well as at the time we are considering a potential resolution.¹

This caliber of analytics requires expanding beyond traditional internal control audits. Whether conducting typical internal audit procedures, or in response to a suspicion of fraud, auditors often rely on a statistical sampling methodology—by definition, auditors do not look at all the data, and often focus on the largest transactions. This approach translates to vast amounts of unanalyzed data. Fraudsters smart enough to conduct improper activities in smaller steps, and in ways that do not appear inherently risky, could avoid detection for years and misappropriate large sums as a result.

In contrast to traditional methods, data analytics relies on the use of forensic techniques to analyze the entirety of the population of data, and to look for connections or other unusual characteristics that might indicate fraud, enabling high-risk transactions to be flagged for further examination.

The following types of transactions have inherently higher levels of risks and are candidates for closer analysis:

- Transactions that vary widely from norms—that could mean a transaction for an amount that differs from an expected or usual amount or to a different vendor
- Use of new attorneys, accountants, consultants or other professionals with no prior relationship to the company, including payments out of the norm relative to historical patterns
- Payments to high-risk vendors, including reimbursement expenses to business development personnel who deal with government officials
- Payments made from and to foreign bank accounts
- Suspicious payment transactions, such as for services rendered, donations, lavish entertainment expenses, gifts, facilitation expenditures and trips with undocumented or unclear business purposes

Real world application:

Simple descriptive analytics such as summarizing a large volume of claims by doctor or service can identify anomalies for investigation. In a recent case, an acupuncturist recruited Amtrak employees to participate in a fraud in which Amtrak paid over \$3 million for services that were not provided or were medically unnecessary. A high-level summary of claims and average cost of service, by doctor, helped identify this anomaly.² This example highlights the need for a comprehensive analytical approach, incorporating simple to advanced analytics.

1. <https://www.justice.gov/opa/speech/deputy-assistant-attorney-general-matthew-s-miner-delivers-remarks-6th-annual-government>

2. <https://www.justice.gov/usao-cdca/pr/acupuncturist-pleads-guilty-charges-scheme-caused-millions-dollars-losses-amtrak-s>

By employing data analytics, companies can identify potential fraud, waste and abuse, and develop proactive steps to minimize the risk of future misconduct. Advanced analytic techniques organizations may use to bolster anti-fraud programs include geospatial analytics, affinity analysis and automated online data extraction techniques as follows:

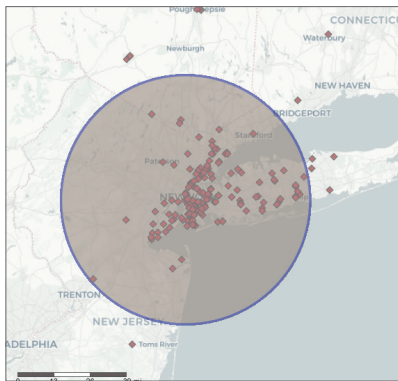
Geospatial analytics

As the world of technology becomes smarter and the internet of things (IoT) proliferates through the average consumer's everyday life, the volume of available geotagged³ and georeferenced⁴ data has increased dramatically. In recent years, organizations have utilized this data to gain insights into operational performance and customer behavior patterns.

For example, shipping companies may use GPS data streaming from delivery trucks to monitor their drivers' route efficiency. Companies like Google use customers' geographic location to optimize search results and target advertising. As this data accumulates, companies can draw further insights into consumer and operational behavior to [make more informed strategic decisions](#).

Geospatial analytics are often paired with demographic analysis to gain further insights. To do so, geographic coordinates are mapped to census tracts or other third-party demographic databases to infer expected income level, age, crime rate and education level, among other characteristics.

Company location and employee addresses



Organizations may slice their transactional records into these categories to identify trends and make strategic decisions. For example, a company may target price increases to customers in locations with higher expected income levels and apply more moderate price increases to customers in lower-income areas in order to maximize profit and customer retention.

In the context of monitoring and investigating fraud, waste, abuse and corruption, organizations can use this type of data to detect anomalies that may indicate a heightened risk for misconduct. For instance, if a company is concerned with payroll fraud and potential ghost employees, it may utilize geospatial analytics to identify employees with outlier home addresses.

To accomplish this, the company would plot the address of each of its locations, along with the addresses of the employees assigned to each respective location. The company would then assign an expected radius, an expected drive time (e.g., 90 minutes) or straight-line distance (e.g., 45 miles), and run its model to identify outliers.

Street view



Organizations may also embed street view or aerial imagery in their dashboards to aid in investigation of suspicious or unusual parties. In the following example, a vendor's address appears to be a check-cashing store or a pawn shop. This may raise a red flag and warrant further investigation.

3. Geotagging is the process of adding "geocoded" coordinates to the metadata of a digital file. According to Google, geocoding is the process of converting a street addresses into geographic coordinates (latitude and longitude), which can be used to place markers on a map, or position the map. See <https://developers.google.com/maps/documentation/geocoding/start>
4. Per the U.S. Department of the Interior, georeferencing means that the internal coordinate system of a map or aerial photo image can be related to a ground system of geographic coordinates. The most visible effect of georeferencing is that display software can show ground coordinates (such as latitude/longitude or UTM coordinates) and also measure ground distances and areas. See <https://www.usgs.gov/faqs/what-does-georeferenced-mean?>

Affinity analysis and association rule learning

Affinity analysis and association rule learning have been staples of retail strategy since the popularization of computer-based point-of-sale (POS) systems decades ago. Commonly referred to as “market basket analyses” in the retail industry, these analyses seek to identify correlation between the purchases of goods. Analysis of a grocery store’s POS data, for example, may indicate that when a customer buys hamburger buns, 95% of customer transactions also included hamburgers, and 45% of transactions included ketchup. These insights aid companies in forecasting sales, ordering product and arranging goods within retail locations.

Tech companies and online retailers also take advantage of these analytics. Companies such as Amazon use such algorithms to provide recommendations to customers when browsing through products (e.g., “frequently bought together,” “customers who viewed this item also viewed” and “compare with similar items”).

While likely not the first thought when considering this type of analysis, affinity analysis can be effectively employed in the context of monitoring and investigating fraud, waste, abuse and corruption. Similar to predicting that consumers will buy ketchup along with their hamburger buns, organizations can apply such algorithms in an anti-fraud context to identify typical business patterns and flag outliers for investigation.

In data mining and predictive analytics, algorithms become increasingly more accurate and insightful as the size of the dataset increases and becomes more standardized. Therefore, industries with vast volumes of standardized data and a high frequency of fraud and abuse, like health care or banking, are particularly well-suited for this type of predictive analysis.

Real world application:

Why the health care industry could benefit from affinity analysis:

Volume: The recent consolidation of the Michiana Health Information Network and the Indiana Health Information Systems created a large technology network of health information linking over 100 hospitals in 38 health systems, over 50,000 providers, and over 18 million electronic patient records.

Standardization: Insurance companies and self-insured companies receive large volumes of medical billing records containing standardized billing codes and forms.

Fraud and abuse: In FY 2018, the Department of Justice opened over 2,000 new health care fraud investigations.⁵

Algorithms can analyze medical data to assess what the expected medical billing line items are for a given procedure, or the likelihood of a procedure or diagnostic test given a health diagnosis. In doing so, such analytics will identify outliers that may be indicative of fraud or abuse.

Online data extraction

A foundational component of any investigation is background research. When investigating potential asset misappropriation or corruption, an investigator needs to identify as much information about the target as possible, including asset holdings, known relatives and associates, corporate ownership, physical and mailing addresses, and phone numbers among other data points. Collecting such information and compiling it into a structured format, capable of cross-referencing and further analysis, can be time consuming.

This burden can be eased through data mining methods such as web scraping using automated headless web browsing, robotic process automation (RPA) and application programming interfaces (APIs).

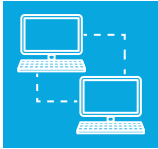
5. <https://oig.hhs.gov/publications/docs/hcfac/FY2018-hcfac.pdf>



Automated headless web browsing: Headless browsers are used to control a web page using scripts or code (as opposed to traditional manual browsing). Web developers often used these tools to test functionality of web pages. In an investigative context, if an investigator frequently accesses the same websites or databases to extract data, a developer can write a script that mimics this exercise in the background without the need for manual browsing and data extraction. For example, if one wanted to check corporate filings related to a list of individuals from ten states, a tool such as Selenium could be used to execute a Java or Python script to query each of the 10 secretary of state web pages and download any filings responsive to a provided list of individuals.



RPA: RPA is a technology in which a user can configure software to access applications and perform tasks such as inputting or capturing data. For example, an investigator could use RPA to extract online information, load and analyze data. A benefit of RPA, as compared to headless web browsing, is that RPA is not limited to web browsing. RPA software, such as Automation Anywhere, can be trained to operate virtually any computer program.



API: An API is an interface in which two computers or parts of a computer can communicate and exchange information. Many online databases have public or subscription-based APIs that allow extraction of data in an automated and structured manner. Examples of useful APIs in the financial investigations and dispute space include:

- Securities and Exchange Commission (SEC) EDGAR filing system
- U.S. Census Bureau
- U.S. Bureau of Labor Statistics
- Credit reporting agencies such as Equifax
- Dun and Bradstreet

Using an API increases efficiency on the front end by eliminating the need to manually browse web pages or database interfaces, and on the back end by providing structured data that requires minimal transformation for analysis.

Wrap-up: Tips for effective implementation of data analytics strategies

As the fraud landscape becomes more complex, and fraudsters more complex in their approach, it's critical to integrate technology and data analytics to assist in preventing, detecting and mitigating fraud. The following tips can help organizations make the move forward into next-level fraud analytics:

Know and optimize your data. Organizations should evaluate the data available within their information systems, including the connection, or lack thereof, between key databases. For example, linking data from a company's customer relationship management (CRM) system to transactional information within the company's financial accounting software will enrich analyses and allow for greater insights. Organizations should work to integrate their information systems to help set the foundation for a strong analytics program.

Validate your data. Data should be validated for quality, standardization and completeness prior to performing substantive analysis to avoid false positives or misleading results. Performing periodic proactive data checkups of key company databases (such as master customer listings) will save time and money when the time comes to analyze an organization's data. Key data tables such as customer and vendor lists should be periodically reviewed to avoid duplication of records.

Use your data. Once armed with relevant and complete data, a company will be able to more accurately understand the drivers of costs and revenues, and go beyond the bounds of typical analysis. Such advanced analytics can help influence more educated and strategic business decisions and create a stronger internal control environment to protect against and detect fraud, waste, abuse and corruption.

+1 800 274 3978
rsmus.com

This document contains general information, may be based on authorities that are subject to change, and is not a substitute for professional advice or services. This document does not constitute audit, tax, consulting, business, financial, investment, legal or other professional advice, and you should consult a qualified professional advisor before taking any action based on the information herein. RSM US LLP, its affiliates and related entities are not responsible for any loss resulting from or relating to reliance on this document by any person. Internal Revenue Service rules require us to inform you that this communication may be deemed a solicitation to provide tax services. This communication is being sent to individuals who have subscribed to receive it or who we believe would have an interest in the topics discussed.

RSM US LLP is a limited liability partnership and the U.S. member firm of RSM International, a global network of independent audit, tax and consulting firms. The member firms of RSM International collaborate to provide services to global clients, but are separate and distinct legal entities that cannot obligate each other. Each member firm is responsible only for its own acts and omissions, and not those of any other party. Visit rsmus.com/aboutus for more information regarding RSM US LLP and RSM International.

RSM, the RSM logo and *the power of being understood* are registered trademarks of RSM International Association.

