Incorporating unstructured data in fraud analytics

Businesses must tap the power of unstructured data to better detect fraud, says Gary Pan

CORPORATE fraud in Singapore is reported to be on the rise. The 2014 KPMG Global Fraud Survey report suggests that 22 per cent of the survey respondents experienced fraud incident within their organization over the past two years, up from 20 per cent in the 2013 report.

Interestingly, some of these respondents in the survey highlighted that fraud incidents were first detected by data analytics. This seems to suggest that as data analytics becomes more prevalent, the use of advanced analytical technologies to identify fraud may increase. The survey also highlights that even the most sophisticated frauds are driven to a large extent by the manipulation of unstructured data. There is a need to incorporate data analytics to better detect and prevent fraud.

Inadequate awareness and expertise in fraud analytics

Traditionally, corporate fraud detection tends to be more technology-centric so as to make better use of advanced data analysis. These include methods such as matching and reporting data in an efficient manner, as this will enable businesses to not only validate and analyse the data and turn it into meaningful information, but also to identify data anomalies and similarities when they are communicating confidentially to their colleagues and when they are processing large amounts of data. Therefore, fraud detection has to evolve to embrace more technology-centric solutions.

Applying unstructured data in fraud detection

The advancement in analytical tools plays a key role in identifying fraud analytics. Several of these analytical tools are built on the premise that the source data is stored in a structured form, for example, spreadsheets and databases that can be queried using structured query language (SQL). These analytical tools strive to make better use of advanced data analysis. These tools and applications can include employees' electronic mails, telephone conversations and many others.

Therefore, fraud detection has to evolve to embrace more technology-centric solutions.

Inadequate awareness and expertise in fraud analytics

A major reason that contributes to this misalignment is the fact that there could be a false sense of security among businesses towards their risk exposure to fraud. This is possible as companies may not evaluate their internal controls regularly, hence overlooking or underestimating the sophistication of new fraud schemes and countermeasures.

In addition, being unfamiliar with the fraud analytical tools that are available in the market could also suggest businesses may not know what they are missing. Understanding the potential risks and what can be done to address them becomes imperative for companies to know more about fraud analytics and to familiarize themselves with the fraud analytical tools that are available in the market. This will significantly improve the overall effectiveness of fraud monitoring and detection in businesses.

In terms of the availability of fraud data, many businesses may not have the required advanced data mining skill sets required for performing fraud analytics techniques. Therefore, it will be useful to have someone in the fraud investigation team who has big data and other advanced forensic technology skill sets to assist with gathering, analysing and validating and analysing the data and turning it into meaningful information. Ideally, this person should be someone who possesses expertise in management and data analysis skills, in addition to a fraud and accounting related skills.

Furthermore, it will be useful to instil a fraud data analytics culture within the business and encourage employees to report suspicious data in an efficient manner, as this will go a long way in establishing an effective fraud monitoring and detection programme.

Approximately 80 per cent of the company’s data is unstructured. Therefore, it is important to incorporate as much unstructured data as possible in fraud detection programmes. This will help improve detection accuracy, as it will lend itself to conventional analysis. These unstructured data may include employees’ electronic mails, telephone conversations and many others. For instance, one may leverage data from electronic mail conversations and social media sites, and combine insights gained from these unstructured sources with official (structured) data sources to build more robust fraud detection models.

Another example would be businesses may analyse unstructured data to investigate collusion. Unstructured data may reveal mismatches between the language people use when they are communicating confidentially to their colleagues and when they are processing large amounts of data. The analysis of unstructured data is much more nuanced in the interpretation of language than is possible in the case of structured data.

In summary, it is important to incorporate as much unstructured data as possible in fraud detection programmes. This will help improve detection accuracy, as it will lend itself to conventional analysis. These unstructured data may include employees’ electronic mails, telephone conversations and many others. For instance, one may leverage data from electronic mail conversations and social media sites, and combine insights gained from these unstructured sources with official (structured) data sources to build more robust fraud detection models.

Allocating resources to mine unstructured data

As with any other type of financial fraud, the major decision lies in whether to allocate resources to mine unstructured data, and how to perform fraud data visualisation, statistical analysis and risk modelling among other things. Traditionally, many businesses may have allocated resources to mine unstructured data, for instance, electronic mail conversations and social media sites, and combine insights gained from these unstructured sources with official (structured) data sources to build more robust fraud detection models.

Another example would be businesses may analyse unstructured data to investigate collusion. Unstructured data may reveal mismatches between the language people use when they are communicating confidentially to their colleagues and when they are processing large amounts of data. The analysis of unstructured data is much more nuanced in the interpretation of language than is possible in the case of structured data.

In summary, it is important to incorporate as much unstructured data as possible in fraud detection programmes. This will help improve detection accuracy, as it will lend itself to conventional analysis. These unstructured data may include employees’ electronic mails, telephone conversations and many others. For instance, one may leverage data from electronic mail conversations and social media sites, and combine insights gained from these unstructured sources with official (structured) data sources to build more robust fraud detection models.