Tapping Analytics to Fight Internal Insurance Fraud

White Paper

Tapping Analytics to Fight Internal Insurance Fraud

White Paper

August, 2010
www.mu-sigma.com

Tapping Analytics to Fight Internal Insurance Fraud

Financial businesses have traditionally been susceptible to fraud - in particular internal fraud. Internal fraud occurs when an employee exploits the vulnerabilities of the system to route benefits to either his or her self or an accomplice. Historical trends show an increase in internal fraud during economic crisis. Unfortunately, unlike external fraud, internal fraud is hard to detect and prevent at nascent stages. According to the Insurance Information Institute, currently, fraud costs the property & casualty insurance industry as much as $30 billion annually. Insurance claim investigation is a time-consuming process. Only a handful of claims are examined by internal investigators for fraudulent activity due to the operational burden involved to re-investigate each claim. From among millions of claims processed each year, a miniscule percentage is proven as internal fraud. Data-mining techniques leverage these historical instances of fraudulent claims and funnel out those with higher chances of being fraud, thus prioritizing claims for detailed investigation. Modeling for internal fraud can be complex due to the rarity of its occurrence.

This article discusses the data elements that need to be considered together with the various analytical approaches undertaken to monitor this "out of pattern" claim activity.

The insurance industry has evolved to better record and organizes data pertaining to customers, employees and agents. Transactions related to shopping, policy endorsements, claims, retention etc., including service level touch-points are now at their disposal. In particular, detailed claims data is captured from the very first notice of loss to final settlement; for example:

• Demographic characteristics of the claimant and employee such as Date of Birth, ZIP Code etc • Claim Characteristics such as Loss Facts (Time of Event, Type of Loss, etc), Damage Facts (Vehicle Total Loss, flooding of basement etc), Settlement Amount, etc.

The above structured data elements could be further transformed into the following:

• Geographic distance between claimant and employee or employee and repair shop (based on ZIP Code)

• The total number of supplements for an auto claim

• Difference between the allowable payment amount by an employee versus the actual amount paid
Data preparation (to model fraud) is the first critical step that translates the fraudulent instance into its component data elements. An illustrative fraud scheme might be:

• An employee (claims adjuster) issuing several payments to the same payee on different claims; or

• An employee with connections to a repair shop or providers and issuing multiple checks to the same payee.

In either of the above instances, "the number of payments made by an employee to payee" is the defining data transformation to interpret the fraudulent scheme. Similarly, various attributes can be suitably captured to define different fraudulent instances for modeling.

Understanding and classifying the target attribute is the key for modeling. For our purpose, each payment can be classified into two possible outcomes: fraud or non-fraud. One of the most popular data-mining techniques to predict dichotomous outcomes is Binary Response Modeling, often referred to as Logistic Regression. The goal of logistic regression is to correctly forecast the occurrence of fraud at each individual instance based on the data elements mentioned above. As mentioned, occurrence of internal fraud is extremely rare. Hence, weighting (1) fraudulent cases to achieve appropriate over-sampling (2) has been found to be effective to build better models to forecast fraud.

More recently, application of generalized linear models has been found to forecast fraud effectively. Note that in the approaches mentioned, only the claims investigated and proved as fraud can be used to identify key characteristics and predict fraud in future. Therefore, validation is yet another important step in the modeling process. Common resampling techniques for model validation include bootstrapping and jackknifing. These methods are used to improve model accuracy.

The response modeling approach to identify fraud has its limitations in the way companies consider an uninvestigated case as a non-fraud. There might be claims that could have been identified as fraud if they were flagged and investigated. Unsupervised learning techniques that classify payments into clusters and assign a target remediate this limitation. Inputs from domain experts should be used to define profiles of these clusters and build upon them. Predictive models, in combination with an ensemble approach to perform unsupervised learning and social networking analysis (3), is evolving.

Irrespective of the approach used, the models should be assessed for their effectiveness against the current strategies employed. It is implicit that the performance of the model needs to be monitored and revised as necessary on a continual basis. Efforts to investigate select payments at random should also continue as new fraud patterns emerge.

The analytical approaches discussed aid the detection of internal fraud. But the key to success of any model lies in understanding the business challenges and adopting the right analytical approach coupled with domain expertise.