Using Data Mining to Detect Insurance Fraud

Improve accuracy and minimize loss

Highlights:

- Combine powerful analytical techniques with existing fraud detection and prevention efforts

- Build models based on previously audited claims and use them to identify potentially fraudulent future claims

- Ensure your adjusters focus on claims that are likely to be fraudulent and have the greatest potential for adjustment

- Deploy results to the people who can use that information to eradicate fraud and recoup money

Introduction

Every organization that exchanges money with customers, service providers or vendors risks exposure to fraud and abuse.

Insurance companies around the world lose more and more money through fraudulent claims each year. They need to recoup this lost money so they can continue providing superior services for their customers. Agencies can now combine data mining with existing fraud detection and prevention efforts to improve accuracy, decrease manpower and minimize loss.

To ensure adjusters target claims which have the greatest likelihood of adjustment, many insurance companies have incorporated data mining into their investigating and auditing processes. Data mining combines powerful analytical techniques with your firsthand business knowledge to turn data you've already acquired into the insight you need to identify probable instances of fraud and abuse.

Discover how to prevent loss due to fraud

How does your organization determine which of its thousands or millions of claims are legitimate? Perhaps your adjusters tend to target claims or payment requests that represent inconsequential adjustments, while missing the ones that offer significant amounts of money to recoup. What if you could:
- Discover small subsets of claims with a high percentage of recoverable fraud?
- Isolate the factors that indicate a claim or payment request has a high probability of fraudulence?

Summary:

Insurance companies lose millions of dollars each year through fraudulent claims, largely because they do not have a way to easily determine which claims are legitimate and which may be fraudulent. To ensure that adjusters target claims which have the greatest likelihood of adjustment, many insurance companies have incorporated data mining into their investigating and auditing processes. This report describes how data mining techniques can enable you to improve accuracy and save time, money and resources.

- Develop rules and use them to flag only those claims or requests most likely to be fraudulent?
- Ensure your adjusters could review claims or requests that are not only likely to be fraudulent but also have the greatest adjustment potential?

If your company could accomplish these goals, you could use your resources more efficiently and more effectively prevent and reduce fraudulent activity. Then your department could reduce the substantial amount of money lost to fraudulent claims each year.

Capitalize on existing data

Your previously audited claims hold the key to recouping money in the future. By creating models from historical information, you can accurately pinpoint fraudulent claims out of the millions of claims you receive each year. These data mining models lower the cost of fraud and abuse while saving your adjusters time. Data mining empowers a variety of insurance providers with the ability to predict which claims are fraudulent so they can effectively target their resources and recoup significant amounts of money.

The following scenario demonstrates how one insurance provider – in this instance a medical insurance provider – used data mining to build models based on previously audited claims to identify potentially fraudulent claims. With these models in place, the provider's claim audit selection will be more exact, generate more money through claim adjustments and save time and manpower.

Building models to find fraudulent claims

A large insurance provider needs to accurately determine which claims are fraudulently filed so it can concentrate on preventing revenue loss. Over the years, this organization collected audit results on insurance claims. The organization seldom used historical records to identify probable fraudulent claims in the future. Their previous methods often missed opportunities to collect money and adjusters spent too much time reviewing legitimate claims.

Data mining now enables this company to predict which insurance claims are likely to be fraudulent. This gives adjusters the power to determine what returns they should target, thereby, recouping millions of dollars otherwise lost and saving adjusters many hours of valuable time.

The insurance company's fraud detection office used IBM SPSS Modeler*, the leading data mining workbench, to get results. Modeler examines each line entry on claims, compares the line entries against the amount of fraud dollars detected, ranks claims in the order of likely fraudulence and displays the results back.

* *IBM SPSS Modeler was formerly called PASW® Modeler.*

The following scenario describes how the agency created models and predicted which claims might be fraudulent.

Understand your data

Modeler's visual programming interface makes examining and modeling the audit records straightforward. Records used to model medical claims might include medical insurance billing records with detailed information such as recipient/provider codes and county of residence, diagnostic codes, admission source, length of stay and total charges claimed.

Determine your population makeup

An important step in the data mining process is to continually ensure you are using the right data to solve your business problem. You must ensure data doesn't disproportionately represent any provider or exclusively associate any provider with one particular diagnostic code.

Discover relationships in your data

Modeler easily adds new variables to each record in the dataset. Then we can route data to a node that will build a web graphic to examine how frequently (and for which diagnostic codes) each provider filed claims for out-of-county services. This information might prove useful further along in our analysis.

Build a model

In this step, we model the total charges on an insurance claim, using the admissions source, length of stay and diagnostic codes as inputs. We chose a modeling procedure called rule induction because it is easy to understand. When we insert the model into the stream, the model will read the inputs (admissions source, length of stay and diagnostic code) for each record, and then produce a projected value for total charges. We'll use this new value later on.

Use the model against actual records

To examine the difference between the actual charges recorded on each insurance claim and the charges that our model projected, we will graph one against the other in a plot graphic. As part of the graphic, we could add a graph of the line, y=x. If the actual charges equal the projected charges on a particular record, that records point should fall on this line. On the other hand, if the record's actual charges were greater than the model projected, its point would be above this line.

Segment your data

To further drill down on the differences between actual and projected charges, we derive a new variable to add to our data. This new variable, "miss," is then graphed in a histogram. It shows that the majority of records in our data had a miss value clustered very close to $0. However, few records have a miss value that extends to $10,000 and beyond. Now we can narrow our attention to only those latter records.

Compare your subset to the entire population

Suppose that two providers had disproportionately high claims values. Exploring the types of claims submitted might uncover very suspicious activity for one of the claimants. They filed claims exclusively for one diagnostic code. The business analyst can easily identify this highly suspicious behavior and investigate the claims. Data mining proved useful for two reasons. First, it yielded valuable information about potential cases of fraud in the records currently on file. Although by no means an open-and-shut case of fraudulent claims submittal, the evidence we gathered can now be passed to investigators.

Plus, the model we built can be applied to future claims. The model will compute total projected charges on incoming claims. These projections can be compared against actual charges, and the system will "flag" questionable claims for investigation. With the information uncovered through data mining, adjusters can focus on claims that may yield larger adjustments, and are less likely to waste time investigating legitimate claims. With data mining, your adjusters can focus on recovering money so your organization's bottom line is less affected by fraud.

Although this insurance agency's fraud detection office used data mining for provider fraud, you could also use it for:

- Eligibility fraud
- Auto insurance fraud
- Credit fraud
- Online fraud
- As well as many other types of fraud

And, because circumstances change over time, you can periodically review the models and update them so they continue to be effective and deliver the best results. Modeler is the only data mining product that empowers organizations to continually modify the data mining process to keep decision makers updated.

Strategically deploy your data mining results for optimum success

Once you have models that predict fraudulent activity, you need to strategically deploy your results to the people who can use that information to eradicate fraud and recoup money. Strategic deployment means integrating models into your company's daily operations. Strategic deployment empowers you to put timely, consistent information into the right hands. Everyone in your organization is on the same page and can act more quickly to recoup the most money for your organization.

IBM® SPSS® Decision Management enables you to score values against new claims and payment requests, then deploy your model and distribute the results (for example, a list of claims most likely to be non-compliant). Depending on your needs, you can display these results over an intranet, through e-mail or via hard-copy reports.

For companies that have local or branch offices, strategic data mining deployment provides an additional benefit. The central office can store and mine data for the entire organization and deploy the data mining results to local offices, which are often charged with stopping fraud and abuse. Deploying data mining results to local offices can stretch scarce resources, empowering you to consistently share information throughout the organization.

Not sure where to start?

You can begin by tracking and solving critical business problems using the best practice approach to data mining, CRoss-Industry Standard Process for Data Mining (CRISP-DM). CRISP-DM is a comprehensive data mining methodology and process model that makes large data mining projects faster, more efficient and less costly. SPSS, an IBM Company, subscribes to this best practice approach to data mining – which it co-authored with several other leading companies – and brings it to your organization to deliver actionable results. The CRISP-DM model offers step-by-step direction, tasks and objectives for every stage of the process, including business understanding, data understanding, data preparation, modeling, evaluation and deployment. This methodology can provide an excellent starting point for your data mining efforts by helping you:

- Assess and prioritize business issues
- Articulate data mining methods to solve them
- Apply data mining techniques
- Interpret data mining results
- Deploy and maintain data mining results

For more information about CRISP-DM, see www.crisp-dm.org.
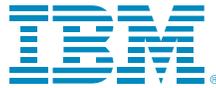
Data mining makes the difference

Discover patterns that indicate which claims have a high probability of fraudulence when you apply sophisticated data mining techniques to your past claims data.

IBM SPSS solutions help you use data and technology better and improve your ability to recoup significant amounts of money. Our experts team with you to merge deep analytical and technical knowledge with your business expertise. Along the way, they educate your staff, provide recommendations and build a repeatable process so your organization can apply the skills and tools to easily proceed on your own.

Analyze your data using a variety of techniques, from simple reports to advanced methods for predicting provider, client or vendor behavior. It's important to use multiple analytical methods so you can work with many types of data in many applications and always get answers that lead to the best chance to realize substantial claim adjustments.

About SPSS, an IBM Company

SPSS, an IBM Company, is a leading global provider of predictive analytics software and solutions. The company's complete portfolio of products - data collection, statistics, modeling and deployment - captures people's attitudes and opinions, predicts outcomes of future customer interactions, and then acts on these insights by embedding analytics into business processes. IBM SPSS solutions address interconnected business objectives across an entire organization by focusing on the convergence of analytics, IT architecture and business process. Commercial, government and academic customers worldwide rely on IBM SPSS technology as a competitive advantage in attracting, retaining and growing customers, while reducing fraud and mitigating risk. SPSS was acquired by IBM in October 2009. For further information, or to reach a representative, visit www.spss.com.

IBM®

SPSS®
AN IBM® COMPANY

IMW14283-USEN-00