

Predictive Analytics White Paper

Charles Nyce, PhD, CPCU, API
Senior Director of Knowledge Resources
American Institute for CPCU/
Insurance Institute of America

Copyright AICPCU/IIA and subject to all copyright laws.

PREDICTIVE ANALYTICS WHITE PAPER

Copyright AICPCU/IIA and subject to all copyright laws.

Charles Nyce, PhD, CPCU, API

Senior Director of Knowledge Resources

American Institute for CPCU • Insurance Institute of America
720 Providence Road • Suite 100 • Malvern, PA 19355-3433

Phone (610) 644-2100, ext. 7245 • Fax (610) 644-7387

www.aicpcu.org

© 2007

American Institute for Chartered Property Casualty Underwriters/Insurance Institute of America

All rights reserved. This paper or any part thereof may not be reproduced without the written permission of the copyright holder.

Unless otherwise apparent, examples used in AICPCU/IIA materials related to this paper are based on hypothetical situations and are for educational purposes only. The characters, persons, products, services, and organizations described in these examples are fictional. Any similarity or resemblance to any other character, person, product, services, or organization is merely coincidental. AICPCU/IIA is not responsible for such coincidental or accidental resemblances.

This material may contain Internet Web site links external to AICPCU/IIA. AICPCU/IIA neither approves nor endorses any information, products, or services to which any external Web sites refer. Nor does AICPCU/IIA control these Web sites' content or the procedures for Web site content development.

AICPCU/IIA specifically disclaims any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials.

AICPCU/IIA materials related to this paper are provided with the understanding that AICPCU/IIA is not engaged in rendering legal, accounting, or other professional service. Nor is AICPCU/IIA explicitly or implicitly stating that any of the processes, procedures, or policies described in the materials are the only appropriate ones to use. The advice and strategies contained herein may not be suitable for every situation.

Contents

| | |
|---|-----------|
| Introduction | 1 |
| Overview of Predictive Analytics..... | 1 |
| Drivers of Insurers' Use of Predictive Analytics..... | 2 |
| Technological Advances | 2 |
| Data Availability..... | 2 |
| Insurers' Desire for Growth In Slow-growth Markets | 2 |
| Insurers' Search for Competitive Advantage | 3 |
| Insurers' Use of Predictive Analytics | 4 |
| Marketing | 4 |
| Hit Ratio | 4 |
| Retention Ratio | 5 |
| Underwriting | 5 |
| Claims | 5 |
| Identifying Potentially Fraudulent Claims | 5 |
| Prioritizing Claims | 7 |
| The Predictive Analytic Process | 7 |
| Data Mining..... | 9 |
| Model Development..... | 9 |
| Regression Models | 9 |
| Advanced Models | 11 |
| Model Validation..... | 11 |
| Predictive Analytics' Advantages for Insurers..... | 11 |
| Predictive Analytics' Disadvantages for Insurers | 12 |
| Frequently Asked Questions Regarding Widespread Insurer Use of Predictive Analytic Techniques..... | 14 |

Copyright AICPCU/IIA and subject to all copyright laws.

Foreword

The American Institute for Chartered Property Casualty Underwriters and the Insurance Institute of America (the Institutes) are independent, not-for-profit organizations committed to expanding the knowledge of professionals in risk management, insurance, financial services, and related fields through education and research.

In accordance with our belief that professionalism is grounded in education, experience, and ethical behavior, the Institutes provide a wide range of educational programs designed to meet the needs of individuals working in risk management and property-casualty insurance. The American Institute offers the Chartered Property Casualty Underwriter (CPCU®) professional designation, designed to provide a broad understanding of the property-casualty insurance industry. CPCU students may select either a commercial or a personal risk management and insurance focus, depending on their professional needs.

The Insurance Institute of America (IIA) offers designations and certificate programs in a variety of disciplines, including the following:

- Claims
- Commercial underwriting
- Fidelity and surety bonding
- General insurance
- Insurance accounting and finance
- Insurance information technology
- Insurance production and agency management
- Insurance regulation and compliance
- Management
- Marine insurance
- Personal insurance
- Premium auditing
- Quality insurance services
- Reinsurance
- Risk management
- Surplus lines

You may choose to take a single course to fill a knowledge gap, complete a program leading to a designation, or take multiple courses and programs throughout your career. No matter which approach you choose, you will gain practical knowledge and skills that will contribute to your professional growth and enhance your education and qualifications in the expanding insurance market. In addition, many CPCU and IIA courses qualify for credits toward certain associate, bachelor's, and master's degrees at several prestigious colleges and universities, and all CPCU and IIA courses carry college credit recommendations from the American Council on Education.

The American Institute for CPCU was founded in 1942 through a collaborative effort between industry professionals and academics, led by faculty members at The Wharton School of the University of Pennsylvania. In 1953, the American Institute for CPCU merged with the Insurance Institute of America, which was founded in 1909 and which remains the oldest continuously functioning national organization offering educational programs for the property-casualty insurance business.

The Insurance Research Council (IRC), founded in 1977, helps the Institutes fulfill the research aspect of their mission. A division of the Institutes, the IRC is supported by industry members. This not-for-profit research organization examines public policy issues of interest to property-casualty insurers, insurance customers, and the general public. IRC research reports are distributed widely to insurance-related organizations, public policy authorities, and the media.

The Institutes strive to provide current, relevant educational programs in formats and delivery methods that meet the needs of insurance professionals and the organizations that employ them. Institute textbooks are an essential component of the education we provide. Each book is designed to clearly and concisely provide the practical knowledge and skills you need to enhance your job performance and career. The content is developed by the Institutes in collaboration with risk management and insurance professionals and members of the academic community. We welcome comments from our students and course leaders; your feedback helps us continue to improve the quality of our study materials.

Peter L. Miller, CPCU
President and CEO
American Institute for CPCU
Insurance Institute of America

INTRODUCTION

While accurately forecasting factors such as operations, budgets, supplies, or product demand is crucial to any organization's success, insurance organizations are particularly reliant on predicting future activities. An insurer's ability to forecast a policy's ultimate cost determines how accurately it prices its product and, in turn, the extent to which it can avoid adverse selection.

Insurance has always relied on forecasting. Initially, insurers simply guessed appropriate premiums. Subsequently, they determined premiums by analyzing a single factor, such as the age of an insured building or the piloting history of an insured ship's captain (univariate analysis). As insurance operations became more technologically advanced, multiple factors such as the age of the insured building, its type of construction, its usage, and so forth (multivariate analysis) were used to determine an appropriate premium. Today, insurers use techniques known as predictive analytics to determine additional information such as credit scores or local economic conditions that may be relevant or correlated with a potential insurance outcome.

The use of predictive analytics has quickly become an insurance industry best practice. Insurers use predictive analytic techniques to target potential clients, to determine more accurate pricing, and to identify potentially fraudulent claims. This white paper discusses the foundations of predictive analytics, the drivers of its growth, its uses in the insurance industry, the implications of its widespread use, and some of its technical aspects.

OVERVIEW OF PREDICTIVE ANALYTICS

Predictive analytics is a broad term describing a variety of statistical and analytical techniques used to develop models that predict future events or behaviors. The form of these predictive models varies, depending on the behavior or event that they are predicting. Most predictive models generate a score (a credit score, for example), with a higher score indicating a higher likelihood of the given behavior or event occurring.

Predictive Models: Credit Score

The most prevalent examples of predictive models are those used by the three credit bureaus (Experian, Equifax, and TransUnion) to develop credit scores for individuals. Each credit bureau uses a variety of information about an individual (income, credit history, outstanding loan balances, and so forth) to develop a credit score that predicts the likelihood that he or she will repay current and future debts. The higher the credit score, the more likely the individual is to pay his/her debt.

Data mining is a component of predictive analytics that entails analysis of data to identify trends, patterns, or relationships among the data. This information can then be used to develop a predictive model. Predictive analytics, along with most predictive models and data mining techniques, rely on increasingly sophisticated statistical methods, including multivariate analysis techniques such as advanced regression or time-series models. Such techniques enable organizations to determine trends and relationships that may not be readily apparent, but still enable it to better predict future events or behaviors.

DRIVERS OF INSURERS' USE OF PREDICTIVE ANALYTICS

Though businesses have employed predictive analytics techniques for many years, several drivers have increased their prevalence throughout the insurance industry. These drivers include the following:

| Insurance Industry Drivers of Predictive Analytics |
|--|
| • Technological advances |
| • Data availability |
| • Insurers' desire for growth in slow markets |
| • Insurers' search for competitive advantage |

Technological Advances

The statistical techniques used in predictive analytics are computationally intensive. Depending on the amount of data they use, some require performing thousands or millions of calculations. Advances in computer hardware and software design have yielded software packages that quickly perform such calculations, allowing insurers to efficiently analyze the data that produce and validate their predictive models.

Data Availability

The validity of any predictive model depends on the quality and quantity of data available to develop it. While most insurers today have a sufficient amount of data (quantity) to develop their predictive models, many store policyholder information on legacy systems that may not be compatible with systems running predictive analytics software. Converting data on these legacy systems to a usable format can be time consuming and costly.

In addition to the insurers' proprietary data, there are numerous third party sources of data that insurers can use to develop predictive models. These sources include rating bureaus, regulators, advisory organizations, rating agencies, predictive modeling companies, and other data gathering organizations.

Insurers' Desire for Growth in Slow-Growth Markets

Although the property-casualty insurance industry generates more than \$400 billion in premiums annually, premiums have grown at a rate of less than 5 percent per year over the last ten years.¹ In many lines of insurance, including personal lines (where predictive analytics are used most frequently), the growth rate has been substantially lower. This slow growth rate has driven insurers to look for other ways to expand their market share. Insurers can use predictive analytics to develop more accurate pricing and to improve how they target their services. Thus, insurers that use predictive analytics can claim market share from their less efficient competitors.

¹ Source: Author's calculations, Insurance Information Institute Web site, www.III.org.

Insurers' Search for Competitive Advantage

The final driver of the use of predictive analytics, insurers' search for competitive advantage, is related to insurers' desire for growth because their desire for market share leads them to seek advantages over their competitors. The use of predictive analytics may provide insurers with information about applicants that their competitors do not possess. If an insurer's predictive model's rating or score for applicants accurately forecasts behavior, the insurer can more efficiently define the target market, more accurately develop pricing, and more efficiently handle claims, all of which provide it with a competitive advantage over competitors who do not use predictive analytics. As more insurers use predictive analytics, those not doing so will be increasingly exposed to adverse selection because their market will be limited to a subsection of the general population that has worse-than-average loss ratios.

Adverse Selection

Consumers with the greatest probability of loss are those most likely to purchase insurance. This phenomenon is known as adverse selection.

Appropriate insurance pricing requires that the insurer gather information about the applicant sufficient to adequately assess and price a particular policy. Although much information about an applicant is available from the applicant and other sources, it can be expensive for insurers to collect. After taking into account all of the information that can be collected cost-effectively, a portion of the information about the applicant remains unknown to the insurer. In economic terms, this is called information asymmetry and it occurs when one party has information that is relevant to the transaction (the applicant) that the other party (the insurer) does not have. Although this information is important to insurers and would enable them to appropriately price their insurance products, the benefits to the insurer of appropriate pricing do not outweigh the costs of obtaining the additional information.

When an insurer charges an average rate because it cannot differentiate between a low risk and a high risk, high-risk individuals have an incentive to buy the insurance because the premium is too low. Conversely, low-risk individuals do not want to buy the insurance because the premium is too high. This results in adverse selection. Therefore, despite offering the average rate, the group of people the insurer has insured is not an average group—it is worse than average because of adverse selection. This group of insureds would tend to have more accidents and higher claims than the average group because they are poor risks. Avoiding adverse selection is one of the main functions of underwriting.

INSURERS' USE OF PREDICTIVE ANALYTICS

The following discussion of applications of predictive analytics focuses on the three core insurer functions—marketing, underwriting, and claims. While policy pricing may be considered an actuarial function, it is included in the discussion of underwriting.

| Insurance Industry Use of Predictive Analytics | |
|--|--|
| Marketing | Property-casualty insurers can use predictive analytics to analyze the purchasing patterns of insurance customers. This information can be used to increase the marketing function's <i>hit ratio</i> and <i>retention ratio</i> . |
| Underwriting | Insurers can use predictive analytics to filter out applicants who do not meet a pre-determined model score. This type of screening can greatly increase an insurer's efficiency by reducing the employee hours it may have spent researching and analyzing an applicant who ultimately is not a desired insured. If an applicant's model score is sufficient for consideration, then the model score can be used as a rating mechanism on which the insurer can base a variety of price/product points. |
| Claims | Insurers can use predictive analytics to help identify potentially fraudulent claims. It also can be used to score claims based on the likely size of the settlement, enabling an insurer to more efficiently allocate resources to higher priority claims. |

Marketing

Insurance marketing has often relied on referrals and other traditional marketing approaches. Predictive modeling in insurance marketing represents a revolutionary approach to what has commonly been perceived as a relationship-based business.

Predictive analytics is used in the marketing of many products and services. Financial services organizations use predictive analytics to identify potential customers for mortgages, annuities, loans, and investments. Property-casualty insurers can use predictive analytics to analyze the purchasing patterns of insurance customers. This information can be used to increase the marketing function's hit ratio and retention ratio.

Hit Ratio

Hit ratio is a measure of how often the marketing function generates a sale for each contact made with a potential customer. If an agent makes one sale for every ten potential clients, his or her hit ratio is one in ten (10 percent). Predictive analytics used to analyze purchasing patterns may allow the agent to focus on customers more likely to buy, thereby increasing his or her hit ratio. For example, if predictive analytics identifies the two customers out of every ten

potential customers who are least likely to purchase a policy, the elimination of those potential customers from an agent's sales agenda will raise his or her hit ratio to one in eight (12.5 percent).

Retention Ratio

Similarly, an insurer can use predictive analytics to attempt to retain the business of customers who are more likely to purchase insurance from another provider. If it succeeds, the insurer will increase its retention ratio, which is a measure of how many insureds renew policies relative to the total number of insureds. For example, if nine out of ten insureds renew their insurance policies, the insured's retention ratio is 90 percent.

Underwriting

The use of predictive analytics in underwriting is more evolutionary than revolutionary. Underwriting has always attempted to accurately predict future losses and price the products that protect against those losses. Predictive analytics represents the next generation of underwriting tools available to achieve those goals.

Predictive models can be used in expert underwriting systems to remove the human error factor from the underwriting process by streamlining the "normal" underwriting cases and only referring the "exceptions" to the human underwriters. Insurers can use predictive analytics to filter out applicants who do not meet a pre-determined model score. This type of screening can greatly increase an insurer's efficiency by reducing the employee hours it may have spent researching and analyzing an applicant who ultimately is not a desired insured. If an applicant's model score is sufficient for consideration, then the model score can be used as a rating mechanism on which the insurer can base a variety of price/product points.

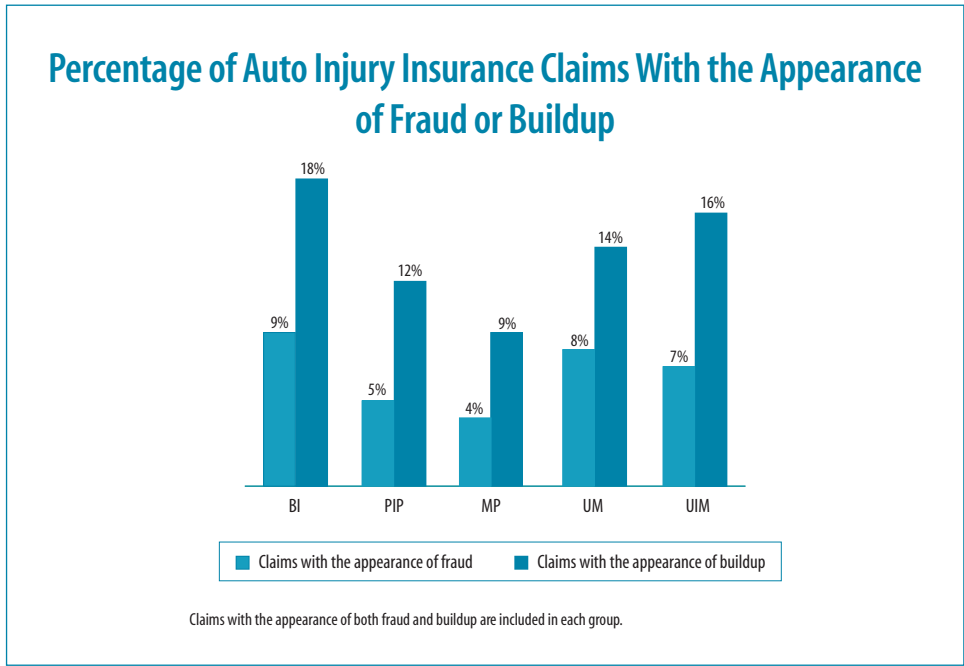
Claims

Predictive analytics is more of a revolutionary concept in claims handling than it is in underwriting. Insurers can use predictive analytics to help identify potentially fraudulent claims. It also can be used to score claims based on the likely size of the settlement, enabling an insurer to more efficiently allocate resources to higher priority claims.

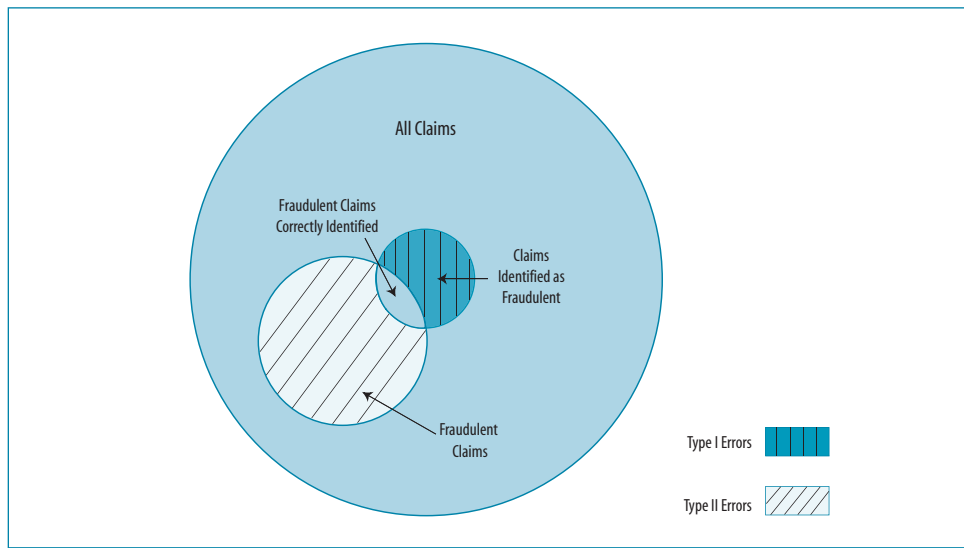
Identifying Potentially Fraudulent Claims

Insurers have struggled for years to develop methods to identify potentially fraudulent claims. While the Insurance Information Institute estimates that insurance fraud costs property-casualty insurers over \$30 billion annually, it is difficult to estimate the actual percentage of all claims that are fraudulent. Fraud can take many forms, from staged accidents to the padding or building up of claims (inflating the value of the claim) for accidents that have already occurred. The

Insurance Research Council's study, *Fraud and Buildup in Auto Injury Insurance Claims: 2004 Edition*, estimates that nearly one in five auto injury insurance claims may have the appearance of fraud, illustrated as follows:



Property-casualty insurers traditionally have had difficulty identifying the relatively small number of fraudulent claims (the needle) made among the millions of claims filed every year (the haystack). Predictive analytics can help insurers more accurately determine claims that need additional review for fraud by increasing the likelihood of discovering fraudulent claims and helping it to refine the claims marked for review. This is known as limiting the occurrence of type I and type II errors. A type I error occurs when a legitimate claim is identified as possibly fraudulent. A type II error is the failure to identify a fraudulent claim and paying it as if it were legitimate, illustrated as follows:



Copyright AICPCU/IIA and subject to all copyright laws.

Both types of errors can be costly. Identifying legitimate claims as fraudulent may anger policyholders and result in litigation or accusations of bad faith in claims practices. Failing to identify fraudulent claims results in higher claims costs and therefore higher premiums for all insureds. Any tools that can aid in the accurate identification of fraudulent claims reduces both types of errors and significantly improves the claims process.

Prioritizing Claims

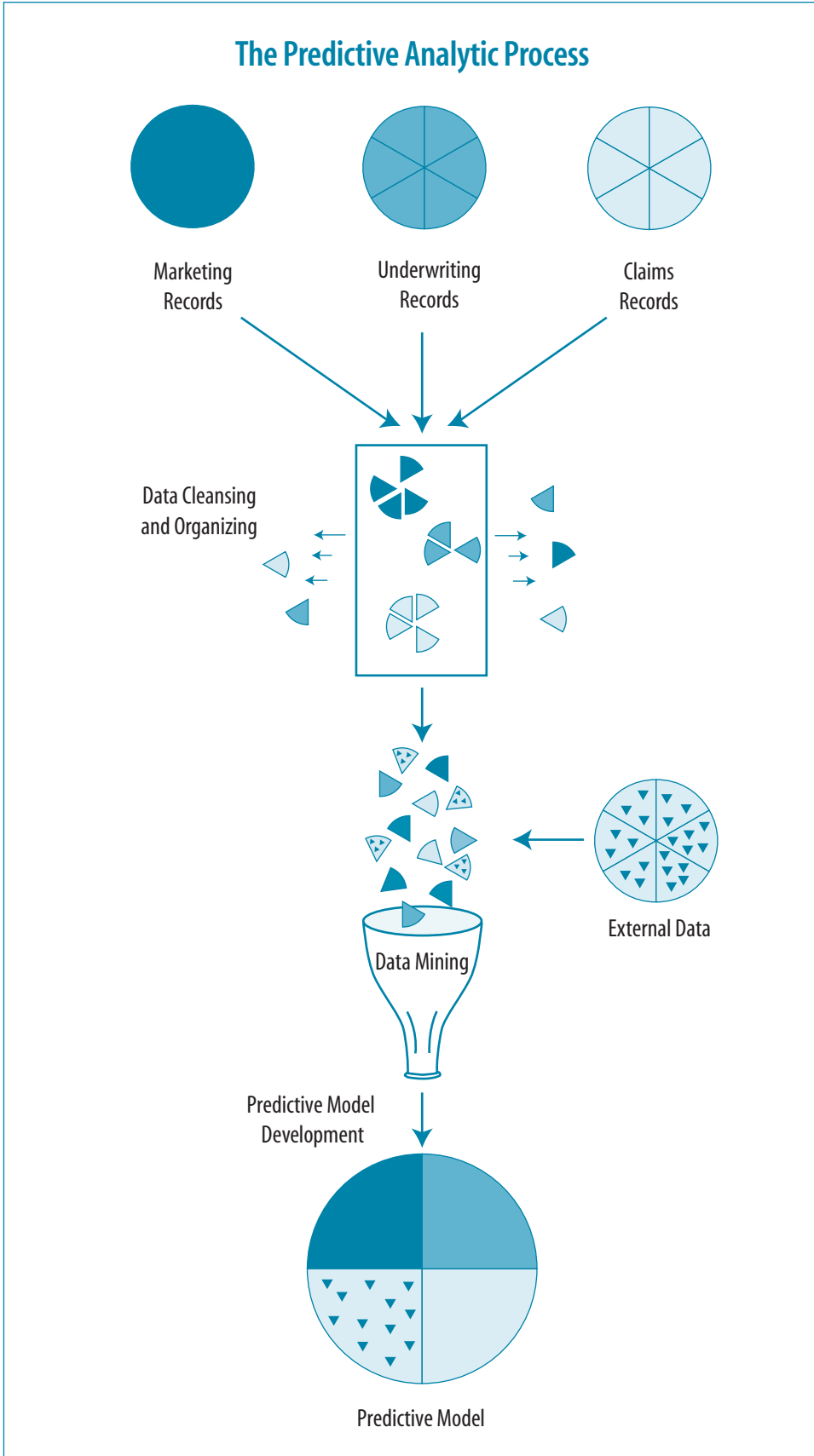
A second use of predictive analytics in the claims process is the prioritizing of claims for handling. Predictive analytics can help identify claims at an early stage that are likely to be settled for higher values. These higher-value claims can then be identified as high-priority claims. More accurately identifying high-priority claims helps the claims function operate more efficiently. High-priority claims can then be handled internally, while the handling of lower-priority claims can be outsourced.

THE PREDICTIVE ANALYTIC PROCESS

When using predictive analytics, an insurer starts by aggregating and “cleansing” its data for use in the analytics software. Cleansing entails scouring records to identify those with missing or incomplete data. Records with missing or incomplete data can have an impact on the accuracy of the predictive model. These records must be completed and/or corrected so that the ultimate predictive model is as accurate as possible.

For example, an insurer developing a predictive model for auto insurance claims would start with its own marketing, underwriting, and claims records for auto policies it has sold. Some of the records may have missing information, such as the age, sex, or marital status of the insured, or may not contain the complete details of the claim (such as not noting whether a police report was filed or subrogation efforts were made). For some insurers, determining this missing information may involve a long and costly process. Multiple legacy systems that may have accumulated through an insurer’s mergers and acquisitions add to the difficulty of converting the data to a usable format.

Once the data have been aggregated and cleansed, sound statistical practices dictate that they be divided into an in-sample group that will be used to develop the predictive model and an out-of-sample group that will be used to test the model.



Data Mining

Data mining is the analysis of data to identify underlying trends, patterns, or relationships. It is a necessary first step in predictive analytics, because the data that the mining process identifies as relevant can then be used to develop the predictive model. One can think of data mining as gathering knowledge about relationships, and the resulting predictive analytics model as applying that knowledge. One distinct advantage to data mining is that it catalogs all relationships (or correlations) that may be found among the data, regardless of what causes that relationship. For example, data mining may discern a relationship between age and gray hair, or age and number of auto accidents, but it does not imply that age causes gray hair or auto accidents.

Model Development

Predictive models can assume many shapes and sizes, depending on their complexity and the application for which they are designed. This section introduces some of the statistical methods that may be used to develop a predictive model.

Unlike data mining, many of the statistical procedures that are employed in predictive models search for one specific relationship. This may require, for example, specifying during model development that age does cause gray hair or that age may reduce the likelihood of auto accidents (at least up to certain ages).

Regression Models

| Regression Predictive Models and Their Uses | |
|---|---|
| Linear Regression | Can be adapted to a wide variety of data types, including time series, cross sectional, pooled, or panel data |
| Partial and Stepwise Regressions | Measures how one independent variable and the dependent variable are related after determining the effect of all the other independent variables in the model |
| Logit or Probit Regressions | Allow one to predict a discrete outcome (yes or no) from a set of variables that may be continuous, discrete and/or dichotomous |
| Regression Splines | Allow different regression models to model data over different regions of the dependent variable |

Regression modeling mathematically describes the relationship between the dependent variable (in predictive models, the variable to be predicted) and independent, explanatory variables given sample data. Regression implies some causation, unlike correlation since modelers must specify a relationship beforehand and then test how well the regression model fits or models the specified relationship. For example, suppose a regression model was designed to examine the relationship between an insured's auto physical damage loss ratio and a few independent variables such as age, type of car driven, and driving history. The model may show a negative relationship between age and loss ratio. In other words, younger drivers have more losses than older drivers. In this case, it's safe to assume that being younger causes more losses, not that more losses causes drivers to be younger.

Regression Basics

Ideally, to explain relationships between variables (such as age and losses in auto), insurers would examine the entire population (in this case, the entire population of drivers). However, insurers can only draw data from a sample of the population (such as only the drivers they insure), so they must do the following:

1. Build the best model they can to determine the “true” relationship between variables.
2. Analyze confidence in the model (mathematical description of its accuracy).

Predictive models may use a variety of regression models. The most basic regression models are linear regression models such as ordinary least squares (OLS) regression. Some of the most complex regression models are multivariate adaptive regression splines called MARSplines.

Regression Techniques Used in Predictive Models

Linear Regression

In a linear regression model, the dependent variable (the variable the model is attempting to predict or explain) is a linear function of several explanatory variables. The most common linear regression models are OLS regressions. Linear regression models can be adapted to a wide variety of data types such as time series, cross sectional, pooled, or panel data.

Partial or Stepwise Regressions

Partial (or stepwise) regression is a regression procedure in which the modeler does not need to specify all of the explanatory variables beforehand, but instead allows the regression procedure to iteratively add variables to the model based on the partial correlation of that variable. For example, after accounting for the age of the driver, how much does car type affect the probability of a car accident? Partial correlation measures how one independent variable and the dependent variable are related after determining the effect of all the other independent variables in the model.

Logit or Probit Regressions

Logit and probit are two examples of a larger class of “generalized linear models.” This broad class of models includes ordinary regression and ANOVA (analysis of variance), as well as multivariate statistics such as ANCOVA (analysis of covariance) and loglinear regression. Logit and probit regressions allow one to predict a discrete outcome (for example, group membership or a fraudulent claim) from a set of variables that may be continuous, discrete and/or dichotomous. Generally, the dependent or response variable is dichotomous (for example, will/will not or success/failure).

Regression Splines

Regression splines allow different regression models to model data over different regions of the dependent variable. For example, assume that for all drivers the probability of auto accidents ranges from .001 percent to 45 percent for a given year. Perhaps the variables that best model the probabilities between .001 percent and 5 percent are age, type of vehicle, and location, while the probabilities between 5 percent and 45 percent are best modeled by type of vehicle, location, number of moving violations and gender. Regression splines allow the model to be created piecewise over various portions of the dependent variable’s distribution.

Advanced Models

In addition to regression models, more advanced models such as neural networks may be used to create predictive models. Neural networks are nonlinear statistical modeling tools. In general, neural networks can handle many more variables than the regression techniques discussed previously and also address some of the other limitations associated with regression techniques such as statistical concerns regarding dimensionality.

Model Validation

To ensure that a predictive model is as accurate as possible, it must be validated through out-of-sample testing. Out-of-sample testing divides data into in-sample data (the data used to develop the model) and out-of-sample data (the model testing window), which includes only data not used. For example, suppose an insurer has twenty-four months' worth of data on the frequency of homeowners' claims. To properly construct and validate a predictive model using the data, the modeler may choose use the first eighteen months' worth of data. Once the model has been developed, data from the final six months could then be used to validate it.

Sound statistical practices dictate that multiple in-sample and out-of-sample data windows be used to develop and validate a predictive model. In the homeowners claim example, the predictive model would not have been properly validated if a major catastrophe had occurred during the six-month out-of-sample testing window. Using multiple out-of-sample testing windows would minimize the influence of such a single event.

PREDICTIVE ANALYTICS' ADVANTAGES FOR INSURERS

If knowledge is power, then the advantages of predictive analytics are clear. Predictive analytic techniques allow insurers to better understand their data and how to use it to predict future events. Proper implementation of predictive analytic techniques can improve an insurer's consistency and efficiency in marketing, underwriting, and claims services by helping to define target markets, increasing the number of policy price points, and reducing claims fraud.

| Predictive Analytic Techniques' Advantages for Insurers |
|--|
| <ul style="list-style-type: none"> Helps marketing department more precisely identify potential policy sales through analysis of customer purchasing patterns |
| <ul style="list-style-type: none"> Reduces the employee hours underwriters may have spent researching and analyzing an applicant who ultimately is not a desired insured |
| <ul style="list-style-type: none"> Provides predictive modeling scores for applicants that can be used as a rating mechanism for determining a variety of policy price/product points |
| <ul style="list-style-type: none"> Helps identify potentially fraudulent claims |
| <ul style="list-style-type: none"> Scores claims based on the likely size of the settlement, enabling an insurer to more efficiently allocate resources to higher priority claims |

PREDICTIVE ANALYTICS' DISADVANTAGES FOR INSURERS

While nearly all insurers will find that the benefits of predictive analytics outweigh its costs, the techniques entail disadvantages, including the following:

- Inherent inaccuracy of the predictive model
- Cost of implementing predictive analytic techniques
- Resistance to change within the organization
- Need for clean, accurate data

A model's potential for inaccuracy is an important consideration for insurers relying on predictive modeling. First, by definition, even a correctly specified model cannot always be 100-percent accurate. An error term represents the portion of the model that is unexplained. This error term can be substantial even in well-specified models and can result in significant variation between predictions and actual outcomes.

Errors in predictive models may also be caused by errors in model specification. A model may include factors that are not significant predictors or factors that may be significant predictors may be excluded, or unobserved.

The final source of error in a predictive model stems from the model's assumption that significant parameters will remain stable from its development period through the period that it may be used. Significant changes in parameters may invalidate a model. For example, a significant economic downturn may substantially change the number of fraudulent claims insurers receive. If the development period on which a model designed to predict fraudulent claim frequency is based does not include any economic downturns, it may not properly reflect the expected frequency of fraudulent claims during that period. Together, these three sources of model errors may be significant.

In addition to a predictive model's inaccuracy, an insurer's use of predictive analytics entails additional disadvantages, many of which are associated with the operational changes that using predictive analysis techniques require. First, an insurer may find that investing in the hardware and software necessary to facilitate predictive modeling constitutes a costly investment. Poor record keeping and multiple legacy systems often indicate that the insurer does not have the clean, accurate data necessary to support a successful predictive modeling platform, which creates the need for further investment. Finally, as with any substantial change in operations, an insurer may encounter resistance from within to the incorporation of predictive analysis techniques that streamline operations and reduce the demand for human resources, particularly from employees who may feel their jobs are being marginalized.

Predictive Analytic Techniques' Disadvantages for Insurers

- Inherent inaccuracy of the predictive model
- Cost of implementing predictive analytic techniques (for example, investments in new software and hardware)
- Resistance to change within the organization, particularly from employees whose job functions are streamlined or marginalized by the use of predictive analytic techniques
- Need for clean, accurate data

FREQUENTLY ASKED QUESTIONS REGARDING WIDESPREAD INSURER USE OF PREDICTIVE ANALYTIC TECHNIQUES

How do insurers avoid the “garbage in—garbage out” modeling conundrum?

Statistical models are only as good as the data used to develop them. Insurers with insufficient data may need to enlist third-party data providers (for example, rating bureaus, data vendors, and modeling organizations) to supplement their own data to ensure that the output from the predictive models is relevant and accurate.

What happens to the value of the relationship between the insurer and the insured as insurance becomes less of a “people business” and more automated?

Studies have demonstrated that the longer an insured remains with the same insurer, the more profitable the relationship is for the insurer. However, automation of the insurance transaction reduces the transaction costs associated with switching insurers, increasing the likelihood that an insured will shop for insurance and ultimately switch providers. This could diminish the value of long-term insured/insurer relationships. However, automation may also help the insurance industry shed the image that who a customer knows is more important than the risk he or she presents.

Does widespread insurer use of predictive analytics increase the rate of commoditization of insurance products? (That is, do insurers now only compete on price?)

Insurers have historically tried to position themselves in the marketplace based on both quality and price. Recently however, as customers have become increasingly price-sensitive, more personal lines insurance marketing campaigns have focused solely on price. Predictive modeling may increase price competition in the market for insurance. If that is the case, predictive modeling may actually speed along commoditization through a reduced focus on differentiating the quality of the coverage being offered.

Is the insurance industry’s use of predictive analytics revolutionary or evolutionary? Hasn’t the industry always been based on trying to predict future losses?

In the context of an insurer’s three major functions—marketing, underwriting, and claims—predictive analytics is both revolutionary and evolutionary. Predictive analytics is evolutionary to underwriting, and revolutionary to marketing and claims. In underwriting, predictive analytics increases efficiency through improved technology. However, because underwriting has always focused on predicting the future, predictive analytics in underwriting is just another tool for performing more accurate risk analysis and price determination. It is more revolutionary in marketing and claims. In both of these functions, predictive analytics can greatly improve efficiency by applying technology where little has been invested before.

How do insurers justify the switch from risk pooling to risk identification?

One of the core benefits of insurance is its ability to pool a group of homogeneous exposure units and offset the losses of a small subsection of that pool with the non-losses of the remainder of the pool. This pooling reduces the variance of losses within the pool and actually removes societal risk. As predictive models become more refined and more insurers use them, the tradeoff between identifying smaller and smaller subsections of the population to use as “homogenous exposure units” versus the ability to pool and reduce risk becomes a more significant issue. For many lines of business, such as auto and homeowners insurance, large insurers insure enough individuals that the added refinement of the predictive model (the increase in the number of groupings) does not reduce the advantages of pooling. Each of these groups will still be large enough to benefit from pooling. However, smaller insurers may see a reduction in the benefits of pooling in some of their smaller groups since they may have fewer insureds in the group and see wider variation in the smaller groups’ performances.

What are some of the social or regulatory implications of predictive analytics?

One of the advantages of predictive modeling is that it may detect relationships among the data or predictors/indicators of potential losses/claims that may not be readily apparent to individuals or that may not be readily explainable. However, an insurer must be able to justify charging differential premiums to customers based on a predictive model output. For example, some consumer organizations and regulators have resisted insurers’ use of an insurance score or credit score as a pricing factor for policies. Insurers initially could not explain why the relationship between credit score and loss ratios existed, thus making it difficult to justify using the relationship to price policies. While such use of an insurance score is becoming more widely accepted, this type of resistance may become more likely if the predictive model factors used to justify pricing are not intuitive. For example, what if a predictor of losses is not just the education level a potential insured attained, but the high school he or she attended, or the hospital where he or she was born? How insurers justify the factors a predictive model uses may be just as important as discovering the relationships.

Copyright AICPCU/IIA and subject to all copyright laws.