

About the Author



Roosevelt C. Mosley,
FCAS, MAAA

Mr. Mosley is a Consultant in Pinnacle's Bloomington, Illinois office. He holds a Bachelor of Science degree in actuarial science and a Bachelor of Science degree in statistics from the University of Michigan in Ann Arbor. He has worked in the insurance industry since 1994.

Mr. Mosley is a Fellow of the Casualty Actuarial Society and a member of the American Academy of Actuaries. He currently serves the CAS as a member of the Committee on Professionalism Education, and is a member of the exam committee. He is also Vice President of the International Association of Black Actuaries Foundation.

Before joining the firm, Mr. Mosley was employed as a pricing actuary for State Farm Mutual Automobile Insurance Company and by Vesta Insurance Group, where he was the personal lines manager responsible for homeowner and private passenger automobile ratemaking and products. He has experience in the areas of personal lines ratemaking, including California auto sequential analysis filings and profitability analysis for private passenger automobile and homeowners insurance; insurance legislation pricing and analysis, including no-fault auto insurance pricing; evaluation of books of business for acquisition; rate filing and regulatory compliance; competitive analysis; reserving; catastrophe modeling, litigation support and financial modeling.



January 2005

Bloomington • Chicago • New Jersey • New York • San Francisco

“The Use of Predictive Modeling in the Insurance Industry”

By Roosevelt Mosley

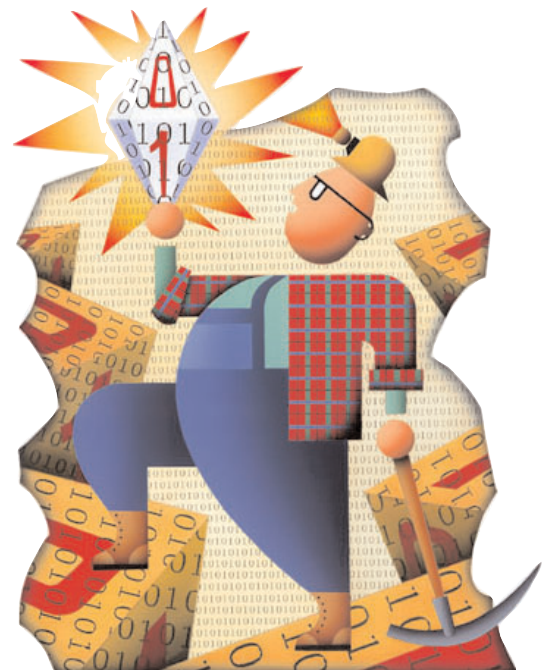
What is Predictive Modeling?

Many successful insurers today know that predictive modeling can assist in better identifying and segmenting insurance risks, which can lead to improved underwriting, pricing, and marketing decisions. There are many companies, however, that have not taken advantage of predictive modeling applications. Predictive modeling can help companies manage the insurance business smarter. Leaders no longer have to manage on instinct or “gut feel”, but can use factual data to assist in making better business decisions.

Predictive modeling is a form of data mining. Data mining is the “analysis of ... observational datasets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner.”¹ Predictive modeling takes these relationships and uses them to make inferences about the future.

What Can Predictive Modeling Do For Me?

First, it can help insurers improve their rating plans by identifying mispriced risks. By analyzing distributional relationships in insurance databases in a multivariate framework, predictive modeling can show



assumptions that can give misleading results. For example, when considering the relationship between insurance losses and age and the relationship between insurance losses and prior accidents, it is no surprise that younger drivers tend to cost more to insure, and that drivers with prior accidents cost more to insure. However, Exhibit 1 (*next page*) shows what this result does not account for, the fact that a larger proportion of younger drivers tend to have prior accidents. Knowing this could make a difference in the way that an insurer chooses to surcharge younger drivers for prior accidents.

Predictive modeling helps insurers define groups that are more homogenous for

¹ Hand, David, Heikki Mannila and Padhraic Smyth. Principles of Data Mining. 2001.

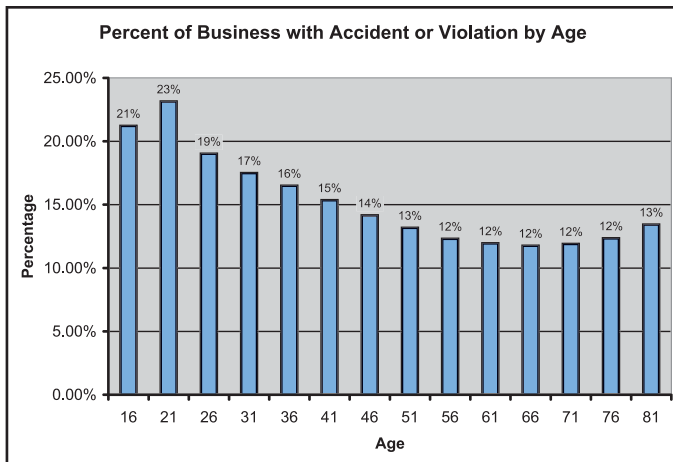


Exhibit 1

rating, underwriting, marketing, etc. For example, an insurance company may rate the city of Dallas and the surrounding areas in Dallas County the same. Predictive modeling may show that the risk of loss outside of the city of Dallas is considerably different than the risk of loss inside the city of Dallas. A company that successfully separates Dallas into more homogenous risk areas will gain a significant competitive advantage.

By identifying new variables or new relationships between variables, predictive modeling can also identify new ways to segment risks. The most vivid example of this in the last decade has been the increasingly widespread use of credit history in generating insurance scores. Insurance scores have been used not only for rating, but also for tiering, underwriting and marketing.

Stages of Predictive Modeling

There are several stages to predictive modeling. First, identify the specific questions you would like predictive modeling to help you answer. Second, find the appropriate data to help you answer the questions. Third, begin mining the data and developing models to help better understand the data. Finally, take the knowledge gained as a result of predictive modeling and apply it to the insurance function, generally through rating, underwriting, or marketing.

Defining the Application

It is very important to clearly define the purpose of the modeling project. Otherwise, a company can try to do too much at once and quickly become overwhelmed.

One of the first applications of predictive modeling has been to better analyze the rating and tiering of insurance business. Historically, the establishment of rating factor relativities for insurance has been based on one-way loss ratios or pure premiums. The problem with this approach was illustrated in the example of prior claim history and youthful drivers. There are many distributional biases in a dataset that cause the one-way approach to produce incorrect results. Using multi-variate predictive modeling methods will account for the

distributional overlap and correlation between risk factors, and ensure the rating and tiering factors being used properly account for differences in risk.

Developing credit-based insurance scores is another application of predictive modeling that is getting a lot of attention. Most companies using insurance scores today are using a score provided by a vendor. While relatively easy to implement, these scores do not factor in an insurer's unique book of business or underwriting philosophy. As a result, a general score developed based on data from several companies may not provide an optimal result for any one company. Developing a custom insurance score takes individual credit elements and uses them to determine a score that is based on the way a specific company does business. Even for small to medium sized companies, a custom insurance score can help provide a competitive advantage.

Customer response modeling (CRM) holds a great deal of potential for increasing profitability. Generally, the pricing of insurance is focused on the supply side of the economic equation, with little emphasis placed on the demand for insurance or the willingness of different market segments to insure at different prices. Given a set of risk characteristics, CRM looks at responses such as the likelihood of policyholder renewal and the likelihood of writing a new business policy. Understanding that the probabilities of renewal and new business conversion are going to be different depending on the characteristics of the risk can help a company boost profitability.

For claims department functions, there are several potential applications, including estimation of claim settlement value. Claims that are settled by insurance companies have characteristics associated with them, including claimant information, the nature of the injuries involved, the presence of an attorney, etc. A model can be developed from historical closed claims that estimates the value of the claim based on its characteristics. This model can then be applied to new claims to estimate the ultimate settlement value of that claim.

There are a number of other applications of predictive modeling that could be discussed here, including vehicle classification, fraud detection, agency evaluation, and loss reserve development factor modeling. It is important to stay focused because taking on too much actually can lead to getting little done. Meanwhile, carefully defining and then carrying out one of these applications will better prepare a company to handle the next application. (Note: future monographs will cover some of these applications in greater detail.)

Gathering and Mining the Appropriate Data

Once the application has been defined, it is essential to collect the data necessary for generating the models. A critical key to the success of any predictive modeling project is the quality of the data on which the model is based. The "garbage-in, garbage-out" rule certainly

applies here. So first, determine what data is needed. Next, you need to extract, verify and cleanse the data as necessary.

There is a wealth of information, internal and external, available to an insurer to be used in predictive modeling. Data sources include traditional internal sources from rating and underwriting; non-traditional internal sources, including agency, marketing, and billing information; and external sources such as credit and allowable demographic information. While it is important to be sure that the data being collected is relevant to the task at hand, being careful not to exclude potentially valuable information is also critical. Many times assumptions are disproved by the actual data, and dismissing data before modeling may undermine the process.

Once you have identified and extracted the data, the next step will be to ensure that it appears reasonable. One approach is to summarize the major statistics, such as premium, exposure, claim counts and claim amounts by the independent variables you are using. Doing so tests the reasonableness of the distributions of the independent variables. For example, there would be reason for concern if 95% of the drivers for the autos insured were coded as males. This process also forces the modeler to understand the levels of the independent variables being reviewed, which will be invaluable once it is time to interpret results of the models.

Developing the Model

There are a number of different types of models that can be fit to the data. The appropriate model will depend on the structure of the data as well as the application being developed.

One analysis method growing in popularity is **Generalized Linear Modeling (GLM)**. GLM allows users to fit a multivariate model with a flexible structure to a dataset, which enables a series of independent variables to predict the value of a dependent variable. This model is especially effective for determining the impact of class plan variables on loss costs, or the impacts of different claim characteristics on an ultimate claim settlement value.

GLM also gives you a framework for discovering the interactions of variables in an automated way. Interactions occur when two independent variables in a model do not have a constant relationship with each other. Exhibit 2 shows the difference in analyzing age and gender separately and then together in an interaction. Without considering the interaction, the model assumes that the difference between males and females is constant for all ages. However, once the interaction is considered, the facts show this is not the case. There are a number of relationships like this in a dataset, some that are intuitive, and some that are not. GLM assists in identifying these potential relationships and provides new insights for pricing and underwriting risk.

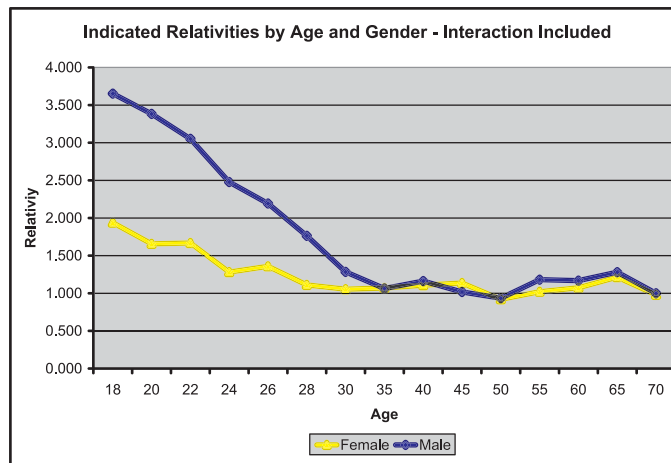
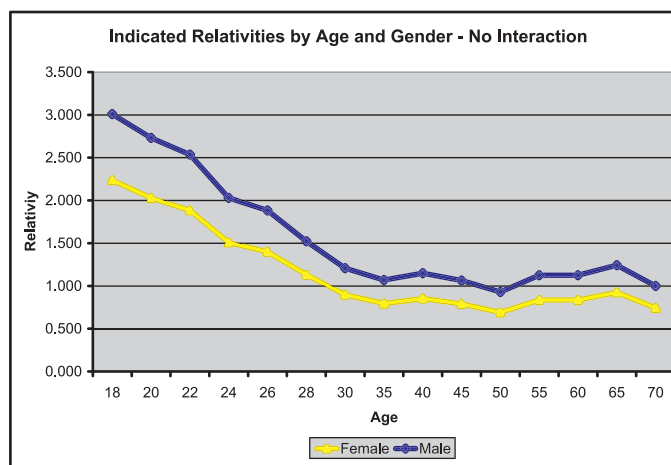


Exhibit 2

Decision tree analysis is a predictive model that attempts to separate a group of risks into homogeneous groups based on an identified response variable. The process begins by taking the entire population, and then analyzes each independent variable to determine which creates the largest degree of separation in the dependent variable. The dataset is then “split,” or branches off, into two or more groups based on this characteristic. Next, each branch is independently analyzed to determine which independent characteristic is most important in distinguishing between levels of the dependent variable for that branch. An example of this is shown in Exhibit 3 (*back page*), which identifies those claims more or less likely to settle for greater than \$25,000.

Logistic regression is used for determining responses to certain situations. Logistic regression generally attempts to model questions with a “yes” or “no” answer. Examples include: “Does the policyholder renew?” or “Does the applicant quoted actually purchase a policy?” Based on a set of independent characteristics, logistic models determine the likelihood of obtaining a “positive” response. For example, when a policyholder age 40 who has been insured with the company for 5 years comes up for renewal, what is the probability that he or she will renew?

Other models, such as neural networks, regression splines, and classification and regression trees can also help insurers glean new insights from data. Neural networks attempt to model human responses to a set of

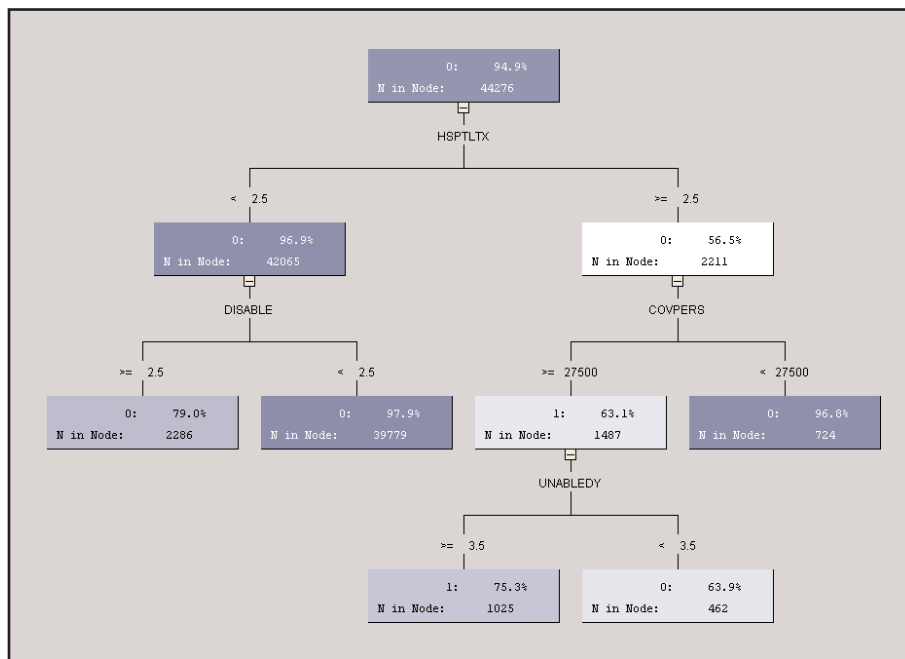


Exhibit 4

produce large indicated rate changes or disruptions that might make a company uncomfortable.

Many times, predictive model results suggest that insurers should be making significant changes in the way they do business. Often, computer systems are not able to handle certain types of changes. Potential systems impacts should be considered, and at times it may be necessary to adjust the application of the model results so that they fit within the framework of what is possible in the current situation. This can also facilitate discussion of what potential systems and infrastructure changes are needed for the future.

Another potential hindrance to the application of predictive modeling results is corporate culture. Many

stimuli, and regression splines build on multiple regression models by making the model structure more flexible. While there may be a variety of different model types, the predictive modeling process will be similar for whichever one is selected. It uses historical experience to attempt to predict future outcomes.

times, predictive models confirm that the assumptions a company is making about a risk are correct. However, there are also instances when the models go against conventional wisdom. The difficult choice of whether to follow the data or follow the “way we’ve always done things” will need to be made.

Before actually generating model results, hold back a random portion of the dataset for purposes of testing and validating the model once it has been developed. The size of the holdback will vary depending on the size of the dataset being used, but this holdback will help prevent over-fitting the model to the data. To perform this validation, take the model developed on the largest portion of the data and apply it to the holdback dataset. To the extent that the results are significantly different, there could be an over-fitting problem.

Lastly, public and regulatory acceptance must be considered. Just because data says the insurance industry should do something does not mean the public or the regulatory community will accept it. Therefore, explaining and implementing results should be done with caution. While the regulatory community has generally accepted predictive models, the results they have generated have not always been embraced. Educating and clearly communicating to regulators and legislators can help ease these concerns.

Interpreting and Applying Predictive Model Results

One of the most important parts of a predictive modeling project is the interpretation of the results. To understand the results, it is helpful to have many people available who understand the process being modeled and hold different points of view. For example, if modeling a rating and tiering plan, it would be helpful to have members of actuarial, claims, underwriting, marketing, and senior management professionals involved to interpret and apply the results. A number of perspectives can help apply professional judgment to model results and come up with a final product that is both powerful and practical.

This diverse team will need to consider many factors when applying modeled results to the real world. For starters, policyholder impacts can be a significant hurdle to implementing predictive modeling results. If an insurance company has not traditionally used these insurance pricing techniques, the modeling results can

Conclusion

Effective predictive modeling can and does enhance underwriting, pricing, and marketing decisions and boost insurer profitability. As companies continue to take advantage of predictive modeling applications, find new rating variables and sources of data, and apply the results in new and innovative ways, it will likely become a way of life for all successful companies, much as it is in other industries such as banking. Actuarial wisdom tells us that past experience is indicative of future experience. If this is true, then based on past successes with predictive modeling, the future of companies that take advantage of it can only be brighter.

For more information about the use of predictive modeling in the insurance industry, you can reach Roosevelt Mosley by phone at (309) 665-5010 or by email at rmosley@pinnacleactuaries.com.