An Executive Guide to Predictive Analytics

What Risk Managers need to know to implement a successful predictive analytics program

Joe Anderson, Director of Analytics, Helios

For many risk managers and claim professionals, the use of predictive analytics in workers' compensation is a complex subject, especially because there is a lack of standard methods and methodologies. The lines between true predictive analytics and data analysis and data access are often blurred. Terms are interchanged and performance is unsubstantiated. As a result, there is often confusion, which can hinder a risk manager's evaluation of an algorithm's efficacy. Becoming versed in key aspects of data and statistics, as well as knowing what to look for when evaluating a predictive analytics algorithm, can arm the risk manager with the information needed to properly evaluate, apply, and utilize predictive analytics.

Start with the Business Problem

The goal of predictive analytics is to generate information that will help to make better decisions. Therefore, it is important to build the predictive analytics algorithm with a specific business problem in mind. What is the nature of the business problem you are trying to solve? What is the outcome you hope to achieve? For example, if you are seeking an algorithm that when put into practice will help avoid claim losses greater that $1 million, it is not necessary to build a model to discern dollar value of losses less than that amount.

Keeping the business problem in mind right from the start can also help determine if more than one algorithm is necessary. For example, in many cases, it is important to measure the full lifetime value of a customer, not just the initial conversion to the business. Customers who are most easily converted may not be the most easily retained. In this case, it may be best to have two predictive models: one to predict conversion and another to predict retention.

Well-understood business problems can assure the algorithm(s) built will address the right questions.

Consider Your Data Set

Data is the foundation for predictive analytics. To qualify as "predictive analytics," the algorithm must rely upon a sufficient quantity of data to determine the best predictors for the most accurate predictions. Consider Google—the largest search engine. By analyzing the data entered in their search engine, they accurately predict trends. In fact, Google can estimate flu levels around the world by adding all flu-related search queries together. Year after year, they compare their model's estimate to traditional flu surveillance systems and refine their models to improve performance. Their findings have been published in the journal *Nature*, and one report suggested that Google can detect regional flu outbreaks 10 days faster than the Centers for Disease Control (CDC) (Google.org).

In workers' compensation, pharmacy data can be used to determine which injured workers are most likely to result in high pharmacy costs, or to have the longest duration of opioid use. By using historical pharmacy data, you can correlate each independent factor known about the injured workers to their long-term severity. With statistical modeling, these balance against each other to generate an accurate prediction.

Therefore, when evaluating a predictive analytics algorithm, always consider the data set because the more data the algorithm uses, the more accurate the predictions generated.

Geography Matters

Considering the geography is equally as important as the data set. In workers' compensation for example, every state has different regulations across the country. One state's solution or regulation may not work in another state for many reasons, specific to their geography. This is a very old problem in the use of applied statistics that still applies today. Franz Kafka, the famous author who was once a workers' compensation attorney, faced the problem of applying data from a different geography (Germany) as he tried to manage the Austrian system. In his case, he called the data from another geography "defective and inadequate" (Corngold, Greenberg, & Wagner, 2008).

Using data appropriate to your geography avoids what Kafka scholars call "a practice of calculating with dubious figures – a practice of the sort that Kafka fought against and that has become to be called Kafkaesque" (Corngold, et al).

Keep Time and Interactions in Mind

One of the greatest limitations in predictive analytics is the assumption that the future will always resemble the past. If you have built a model with data from 2010-2014, and you're trying to predict how long it will be before someone injured in April 2015 will return to work, there may be something unique about people injured in April 2015 that is different from those in your collected data. If the model had originally been built on 2010-2013 data and tested on 2014 data, this would at least show which effects from 2010-2013 are still valid through 2014 (however, this could overstate the significance of 2014). This helps to understand how predictors change over time. The data miners Michael Berry and Gordon Linoff call this type of approach an "out of time" test set (Berry & Linoff, 2009).

Another commonly encountered pitfall is using only current data. A business leader may think that the business has changed too much and therefore, data older than a certain point would be invalid. Most analysts prefer to use as much data as possible. Depending on how important the business cycle is to your predictions, you may want a model that includes at least two recessions in your dataset. In general, the time period of the data collection matters if the predictors in the model change during the time period. For example, in workers' compensation, the highest cost injury types do not change over time. Spinal injuries are more severe than foot injuries. This was true in 1980, and it is still true in 2015. But other predictors may not be the same, such as the regulatory environment. If this is the case, you may be overstating the length of time an injured worker will be out of work if the regulatory environment used to favor the injured worker, but now favors the employer.

On a related note, it is very important to be aware of interactions. In the aforementioned example, the analyst would need to know *what* in the environment has changed over time, and if there is something inherent in the type of injuries in this state that is different from the rest of the country. This could overstate or understate injury severity if the mix of injuries are different, or are treated differently.

Evaluate Fully

If you have ever taken a statistics course, you may remember the term *r-squared*, which isan overall statistical performance measure for a regression model. The higher the r-squared value, the better it is. You may also remember *significance tests*, which allow you to be 95% confident that something is accurate. These concepts, *r-squared* and *statistical significance*, are important however, they're not the final test for determining accuracy. As you evaluate a model's accuracy, don't stop here.

Traditional statistical measures like r-squared assume that the decision makers are acting at random. In workers' compensation, for example, it would assume that the risk managers and/or claims professionals have no idea which injured workers need help. This is almost never the case. Instead, compare your current decision-making process to what would be done without a predictive analytics algorithm. In pharmacy, for example, preventing opioid misuse and abuse can be attempted simply by finding the injured persons already having the highest current opioid use, measured in morphine equivalents. A predictive analytics model would therefore need to demonstrate that it is better at identifying long-term use of opioids than these simple metrics. This can usually be quantified: the "simple version" is an improvement over choosing "at random." Then, how much better is the predictive model than the simple version? For example, if you're showing me an r-squared for a predictive model, what's the r-squared for what I'm already doing today?

In the best case scenario, businesses are able to invest in experiments to test the predictive models in a controlled environment prior to deploying them across the business. If you're fortunate enough to be able to do this, you should make sure to take the time up front to design the experiment correctly. The most important point from an analytical perspective is to have a sufficient control group. Control groups should have the best decision-making tools available outside of the predictive analytics algorithm. Often, this isn't just good methodology; it's ethically mandatory, as described in *Design of Comparative Experiments* by R.A. Bailey: "For a given illness, if there is already a standard drug which is known to be effective then it is not ethical to give no treatment in a clinical trial of a new drug for that illness. The control treatment must be the current standard drug" (Bailey, 2008). This is similar to the point about why r-squared is insufficient; r-squared typically measures the model's performance against random decision-making. However, hardly any business processes are random. The new algorithm must be compared to the old algorithm.

At the end of the experiment, the "control group" can be compared to the group where the predictive analytics algorithm was applied. This is usually done through a test of "statistical significance." The standard methodology is to measure the outcomes for the control group against the ones in the program you're trying to measure. Ideally, the outcome metric (for example, reduction in costs or duration of claim) can be quantified for both the control group and the group in the program, and they can be compared through a significance test. One of the most critical evaluators, the US Food and Drug Administration (FDA), generally requires a study at the 99.875% confidence to pass a clinical trial for a new medication (Chin & Lee, 2008). Typically, businesses are comfortable with 80-95% confidence.

Focus on Comparisons

When you hear "Our model has a statistical significance of 95%," what does that really mean? Always remember, a number by itself is never statistically significant. It must be compared to something else to be significant.

Discussions about statistics can be challenging for risk managers, but discussions about comparisons usually are not. Therefore, ask what the comparison is for the 95% significance. An appropriate answer might be something like, "Our outcomes are statistically significant because our model resulted in claims with lower costs than the control group with 95% confidence." This brings the discussion away from numbers into something that generally non-statisticians are more comfortable talking about such as how the control group was structured and what metrics were analyzed in each group. If these answers are satisfactory, then the risk manager can feel more comfortable with the significance test.

Continually Test Performance

While most of the time analysts will not be able to run a real-world experiment on their predictive models, there are still best practices that analysts can follow. Today, analysts use very large datasets to build predictive models. To test the algorithm, an analyst can *partition* this data set into two or more datasets: one (or more) to build the algorithm, and a final dataset to test it. The concept here is that all the statistical measures used to generate the algorithm are all applied before using the final dataset. Then, after building the model, it is tested on the remaining data. This simulates a "real world" application of the model and helps avoid the problems of "over fitting."

It is critical that this testing does not just happen once. Rather, algorithms should be continually tested and updated when necessary. For example, if a model predicts an average claim cost of $100,000 but your actual average is $75,000, it is probably worth looking at why the model is constantly generating higher predictions on average. You may find that something in your business has changed, and that the model is telling you to focus resources on a problem in the business that has already been solved.

Confirm Benefit

Predictive analytics and other algorithms are tools to help risk managers and claims professionals make better decisions. Their use provides the ability to identify trends earlier and position for action sooner. Knowing what to look for when evaluating their use and performance helps determine if your business is not only reaping all the benefits of the latest advances in technology but appropriately answering the business problem at hand.

Works Cited

Bailey, R. A. (2008). *Design of comparative experiments.* New York: Cambridge University Press.

Berry, M., & Linoff, G. S. (2009, September 30). *Data Mining Techniques: Theory and Practice Course Notes.* Retrieved from www.Dataminers.com: http://www.data-miners.com/companion/BDMT_notes.pdf

Chin, R., & Lee, B. Y. (2008). *Principles and Practice of Clinical Trial Medicine.* London: Elsevier.

Corngold, S., Greenberg, J., & Wagner, B. (Eds.). (2008). *Franz Kafka: The Office Writings.* (E. Patton, & R. Hein, Trans.) Princeton: Princeton University Press.

Google.org. (n.d.). *Google Flu Trends*. Retrieved from Google: http://www.google.org/flutrends/about/how.html