



Article from

Risk Management

December 2018

Issue 43

The Perils and Prospects of Predictive Analytics

By Mark Griffin

Many life and health insurers are beginning to use Predictive Analytics (PA) for a variety of purposes. This article addresses the unique challenges and potential disruption PA brings to the risk manager, as well as the opportunities.

First the high-level challenges:

- Typically, PA models are built by a recently hired group of data scientists. While data scientists tend to be very analytically adept, they may have little understanding of the insurance business and little understanding of model governance protocols and techniques within a financial institution.
- PA models are often built using open source code. It will be easy and tempting for model builders to import blocks of code from libraries or other sources to do specific tasks. There should be specified protocol on this topic.
- Insurance risk managers typically have familiarity, if not hands-on experience, with traditional insurance asset, liability and finance models. With respect to PA models, risk managers are less likely to have hands on experience or any level of familiarity. In fact, there may not be anyone within the company (apart from those who built the models) who can peer review or challenge the models. There may be a fairly limited number of people even outside the company who can do this effectively.

Additionally, one must not forget the standard risk management concerns for any large systems development project. For example, project management and data security should be in place, the pros and cons of using the cloud should be considered, etc.

Risk managers must first recognize that Predictive Analytics is a very dynamic field. Kaggle is a very popular platform for PA competitions, bringing parties with problems and data together with willing data scientists. A number of insurance companies have sponsored projects on Kaggle. Last December, the SOA released case studies from their Kaggle Involvement Program.¹ So familiarizing oneself with Kaggle is an easily accessible first step. Also,

Google recently announced they will offer a service to share their expertise in machine learning and Artificial Intelligence.

Other specific considerations and suggestions include:

With respect to data:

- What were the sources of data? Have the legal and compliance teams been involved? Have the sources and uses of the data been properly documented? How will the use of the data be controlled?
- How were any data quality issues or missing data issues addressed? How were perceived outliers treated? Were they deleted or changed? Was this documented? Will this “cleansed” data be used for other purposes within the company for which the treatment of outliers may not be appropriate? ASOP 23 on Data Quality may be helpful.

With respect to the model:

- The risk manager should have knowledge of the high-level decisions and modelling choices. Was a General Linear Model (GLM), or a machine-learning approach used? With GLM, the modelling process often starts with a hypothesis. Typically a target variable is established. These are both steps the risk manager will want to understand. If a lot of data is available, the model builders may try to determine a smaller set of effective “features,” which are combinations of variables that appear to influence the target variable. Machine-learning approaches allow the computer to search for dependencies within the data with much less human interface. While this approach may be more effective in finding relationships that weren’t previously recognized or properly understood, (and therefore be more “independent” of current approaches and thinking) the approach may be drawn to relationships that have less data supporting them.
- Many PA techniques amount to using computing power to test if the model’s result is stable when different subsets of data are used. To do this, the data are typically split into three groups, training, validation, and test data, often in the proportions 70 percent, 20 percent and 10 percent. There may not be enough data to afford this luxury. Any temporal dimension to the data will also limit this process. Actuaries may want to consider tests of statistical significance as another perspective on this same issue. The article “Is Credibility Still Credible?” in the August 2017 issue of *Risk Management* can provide a perspective on this.
- Regardless of the method selected, were all of the relevant subject matter experts involved?

- There is a long list of model governance considerations. Is version control done automatically by the software, or does it rely on developers? Is there a style guide for coding and documentation? Will there be an “audit trail” produced?
- Can the model be explained to users? To management? Do users understand the purpose and results, and will they use the model appropriately, recognizing its limitations?
- How will the impact and success of the model be measured? Some may want to measure the speed of the model, some may want to measure the number of times it is used, some may want to measure the change in financial results.

PA modelling provides the opportunity for significant disruption of traditional approaches. A good example is in the area of streamlined underwriting for term insurance. Companies have sought to provide an easier, faster process for the purchase of term insurance. Removing the collection of fluids from the underwriting process is the fastest way to do that. PA modelling is therefore being used to build underwriting models using both data sources other than body fluids that have historically been collected, in addition to other data sources that may have predictive power with respect to mortality. These new data resources might include credit scores, pharmacy records, etc. The resulting models are then tested against historical records to see how well they reproduce the results of classic underwriting processes that included the collection of fluids.

Like any disruptive force, it can be very advantageous for companies who adopt [Predictive Analytics] early and effectively, but very detrimental to those who are left behind.

It is worth taking the time to envision the evolution that could follow. As different companies change their underwriting to utilize, or even just reweight, old and new data sources, agent and customer behavior would also change, redistributing business. Eventually, those companies maintaining a traditional approach may lose most or all of their market share to—and even ultimately be selected against—companies using better risk selection in new models.

Many insurers have focused their PA efforts in the marketing area, trying to get ahead in the disruption which is well underway. At a minimum, traditional marketing and sales channels and approaches can be made more efficient.

In the P&C industry, where the use of PA has been prevalent for decades, the underwriting process has been transformed into one where in many cases there are no underwriting classes at all, just a unique price for each customer. In areas like term insurance—where the feedback loop between underwriting and results is slower—the evolution may be slower but would naturally move in the same direction.

In addition to these challenges, there are opportunities related to PA from the risk manager’s perspective. A couple of examples are:

1. PA techniques allow a deeper understanding of dependencies between variables, particularly some that don’t meet the naked eye. What better tool to help the risk manager design early warning triggers or Key Risk Indicators?
2. Insurance companies have begun to appreciate the value of data. For example, in the earlier example of building streamlined underwriting models, not only are new data sources purchased, but older underwriting records may finally be digitized to provide more data points on the ability of the model to reproduce historical underwriting. With digital underwriting information, the effectiveness of the historic model can be properly evaluated and undoubtedly improved.

The advanced analytics teams (referred to at the beginning of the article) may be part of the risk team. Whether they are or not, their success is in the best interest of the risk management effort, and a close working relationship is essential. Diversity of thought and approach is always the risk manager’s friend!

The emergence and evolution of PA approaches will destabilize the marketing, underwriting, and undoubtedly many other aspects of the current life and health insurance environment. Like any disruptive force, it can be very advantageous for companies who adopt early and effectively, but very detrimental to those who are left behind. In addition to the many considerations listed in this article, it would seem wise for risk managers to periodically assess their company’s use of advanced analytics relative to their competitors. ■



Mark Griffin, FSA, CERA, is senior managing director and CRO financial risk at TIAA. He can be reached at mark.griffin@tiaa.org.

ENDNOTE
1 <https://www.soa.org/predictive-analytics/kaggle-program/>