



SOCIETY OF ACTUARIES

Article from:

Health Watch

May 2014 – Issue 75

Operational Considerations in Predictive Modeling

By Christine Hofbeck

The advent of “big data” and predictive analytics has led to a deluge of information on how to model. We are taught how to select the appropriate distribution, methods and steps in univariate and multivariate analyses, discovery of interactions, refinement of splines and bins, and validation techniques using appropriate statistical tests. Modeling teams are developed and nurtured, and optimal predictive models are created.

But building the perfect predictive model does not necessarily guarantee a successful outcome (and often it’s the modelers who are blamed for a fail, even if they’ve built something great). There are operational considerations that must be addressed to optimize the results of any predictive model. These considerations are sometimes obvious, but often overlooked.

Some considerations are described below—but there may be many more based on your particular project, company, staff, knowledge and philosophy. The important thing to remember is that the project starts long before the modeling starts, and the project ends long after the modeling ends. If you ignore the before and after, and don’t strategically manage the middle, you could be setting yourself up for disappointment.

1) EXECUTIVE AND CROSS-FUNCTIONAL SUPPORT

Top management has funded the modeling project, so you’re all set, right? Wrong. The directive to build a predictive model may not mean that all parties will support the final product. And if the target users don’t support the model, they likely will resist using it. It is business critical to gain both executive and cross-functional support for a modeling project prior to model build. This step can be difficult for several reasons:

- Resistance to change
- Concern that model results will highlight current deficiencies
- Lack of understanding of predictive models.

Consider the following examples (see chart on page 32) of typical concerns and possible response.

2) TIME/VALUE CORRELATION

The process of building and implementing a model can typically be quite lengthy—longer than most people expect. If management is looking for a quick fix (i.e., three to 12 months to execution), then they must understand that the study will be simpler and the results more conservative. A more comprehensive and thorough exploration may provide a greater payoff, but will require additional time and development cost. What, exactly, takes so long?

• **Data**—Internal data is often stored in multiple systems, and it is almost never fully accurate and complete. Extracting and combining datasets is a challenge. There will be multiple discussions with data stewards to ensure that the data is properly understood and utilized. Are the data stewards available whenever we need them? Do any data fields include randomly input values whenever the actual values are unknown? Missing data must be imputed. Inaccurate data must be corrected. All data must be converted to the appropriate exposure unit and matched together by a primary key.

Now let’s add external data. It could take a month or two or more to simply finalize a contract with an external data provider. Then it could take two to four months to receive the external data. Sometimes external data must be secured in consecutive steps (versus all at once). For example, credit data must be scrambled due to regulatory constraints; that data must be collected after all other internal and external data has been compiled, appended and scrubbed.

Now let’s consider how we should bin (or group) the data. For example, will we consider each age separately, or group together everyone over the age of 80? 85? 90? What other variables should be binned? Do we group zero values with missing values? Why or why not? Remember that if we are working with hundreds of variables, this seemingly simple step can take weeks.

Or ... do we simply use whatever data we can easily access and assume it is “mostly correct” in order to accelerate the modeling process and minimize development costs?



Christine Hofbeck, FSA, MAAA, is an independent analytics consultant at Centroid Analytics, LLC in Glen Gardner, N.J. She can be reached at christine.hofbeck@centroidanalytics.com.

CONTINUED ON PAGE 32

| Function | Concern | Possible Response |
|-------------------|---|---|
| Underwriting | <p>My position will be eliminated if a model is now used to select risks.</p> <p>My expertise must not be important to the company.</p> | The model will make decisions for the simpler risks, which will allow our highly skilled UWs more time to address the most complex cases. |
| Underwriting | You call it an outlier, but I call it wrong. If the model is calculating the wrong result for one risk, it is useless. | No model is correct 100 percent of the time. It is not a crystal ball; there will always be outliers. However, xx percent of the time the model provides an accurate result. "All models are wrong; some are useful." |
| Underwriting | The model could be contrary to anecdotal evidence, so I won't believe it. | An UW leader will be engaged throughout the model build to ensure the results make sense. Sometimes results can be surprisingly informative |
| Actuarial Pricing | The model will highlight areas in which my current pricing algorithm is incorrect, which will reflect poorly on my performance/reputation. Therefore I will not support the initiative. | Our current pricing algorithms are the best solutions based on tools available until now. New tools are now available to refine current practice. We will always continue to improve. |
| Marketing | I already have an established marketing plan. I know who our target customer is. | A model will help us refine targets. Marketing reps will be engaged throughout the model build to provide valuable insight and feedback, to ensure the target customers identified by the model are appropriate. |
| Marketing | I don't know how to explain this to a broker or agent so I don't want to use it. | Thorough training will be provided so you fully understand how to interpret the results and can explain and support anything counterintuitive. |
| IT | I don't have enough staff to implement scoring engines and user interface in the allotted time. My workflow will double (triple). | IT resources will be fully addressed at project scope. IT will be engaged throughout the model build to ensure the proposed solution is possible from an IT perspective. |
| All functions | We've always done it this way, and it's worked. I don't see a reason to change anything. | We will strive to continue to improve and refine results based on cutting-edge technologies. This new technology will allow us higher profitability, improved operational efficiencies, and the ability to stay ahead of the competition. A transparent model will be built so results are easy to understand and convey. Thorough training will be provided for all users. |



- **Model Build**—To develop a model thoroughly, we may investigate different modeling techniques and software. We will consider many different combinations of variables. We will develop and refine splines to smooth results. We will examine interactions in-depth. We will use multiple statistical tests to optimize results. We will continually review results with the ultimate model users to ensure our proposed formula makes intuitive sense. We may go back and revise data more than once.

Or ... do we simply run our data through the available modeling software, skip or minimize interactions, and allow for generous binning? Simplicity in model structure is necessary if we have minimal time for implementation, and may be desired if the goal is simply to develop a better general sense of our customer, pricing accuracy, marketing technique, or long-term strategies.

- **Implementation**—Now that our model is built, how will we use it to improve results? IT will need to build a scoring engine. Multiple variables, variable conversions, bins and assumptions in a model will complicate implementation. If individual external data will be requested at time of calculation, then a delivery pipeline must be created. Once the model output is calculated for a given risk, how is this information delivered to

the user? Such interface will be required. In order to track the impact of the model, reporting packages for various audiences must be developed.

Or ... perhaps we will calculate simple results in an Excel worksheet and track results on an ad hoc basis?

Ultimately, management will need to understand that you get what you pay for. A basic, swift model build is certainly possible and is definitely appropriate in some situations. However, a simple study probably will not deliver the same impact as a more thorough exploration. The appropriate combination of time, value (cost and return), and comprehensiveness must be understood and approved prior to build to ensure an expected result.

3) STRATEGIC MODELING PROCESS

To optimize usage and impact, it is important for the modeling process to be executed strategically, while allowing flexibility. Consider the following:

- **Target Prediction and Use**—It is critical to define the target prediction clearly, and ensure that it is appropriate for your intended use of the results. In health insurance, are you modeling the individual risk, or the family as a whole? Are you

CONTINUED ON PAGE 34

Modeling is not just for modelers. Building a model collaboratively and cross-functionally is critical for optimal usage and results.

modeling what will happen in the next year, or expected outcome in a lifetime? If you are modeling medical providers, are you interested in the doctor, the doctor group, or the associated hospital? There is no right or wrong answer to these questions. Just make sure that your target prediction and intended model usage are compatible.

There are many interesting topics to investigate and ideas to prove—when developing the target prediction, try to focus on that which is actionable and will impact the company’s profitability, operations, or overall strategy and goals. For example, you could predict how many people of a certain region, age, or insurance coverage have Disease X. However, a more impactful strategy might be to predict the onset of Disease X for an individual based on particular characteristics. (Researchers from the Johns Hopkins University School of Medicine in Baltimore, Md. have recently developed a model to detect early-onset dementia, allowing for possible improvements in the treatment and progression of Alzheimer’s disease.)

- **Statistical Significance vs. Ultimate Impact**—Similarly, the most statistically significant model may not be the most impactful. Often the value to the company hinges on a model’s ease of implementation (distribution) and repeatability. While an intricate, multifaceted or unique model may be ideal in certain situations, parsimonious models that allow for easier implementation and automatic or straightforward updates may be deemed more useful, less costly to maintain, and ultimately more powerful to the business.

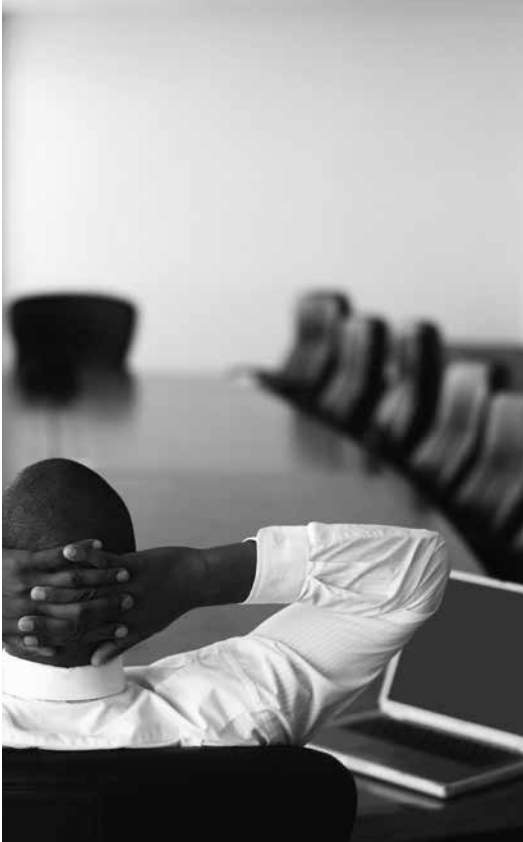
It seems an oxymoron that delivering the less complex (parsimonious, distributable, repeatable) model could be a greater challenge. However, it is not uncommon for health actuaries to build unique models that are not repeatable through time. We are capable of highly intellectual, creative, in-depth solutions, and may become frustrated by our company’s failure to understand our insights. For some actuaries, identifying when “less is more” to deliver a more usable solution could be an ideal area of targeted personal growth.

- **Flexibility**—Although a specific goal may be targeted, significant unexpected insights could occur during the modeling process as a result of the fresh review of data, leading to unanticipated business-critical changes in strategy or process. Flexibility is key—in some cases it may make better business sense to modify the target mid-stream. Consider the development of penicillin, which Sir Alexander Fleming found by accident while studying *Staphylococcus*, or iodine, which Bernard Courtois discovered while processing seaweed ash for gunpowder, or Viagra, which was originally studied for its cardiovascular benefits. In these cases, insights gained from the journey ultimately proved more important than the planned goal.

4) CROSS-FUNCTIONAL INVOLVEMENT THROUGHOUT MODEL BUILD

Modeling is not just for modelers. Building a model collaboratively and cross-functionally is critical for optimal usage and results. Modelers should not disappear into the back room and emerge weeks or months later with the final product. They shouldn’t want to, and they shouldn’t be expected to. This is an especially important consideration if a model build is outsourced. There must be continuous involvement by company data, product and IT experts, legal advisers and model users. Why?

- The modelers are (hopefully) experts in building models, but they may not be experts in all other functions. Underwriters, marketing teams, adjusters and IT can provide valuable insight. Perhaps there is a variable that must be (or must not be) included for some reason. Let’s not wait until the model is fully built to figure this out. For example, if we are building a stop loss model, we must include the attachment points that the company is willing to sell. If the availability of aggregate stop loss attachment points has certain business rules surrounding it, then those rules should be reflected as well.
- The modelers may discover interesting and important information during a model build that should be shared with other function areas. For example, perhaps the modelers are surprised to learn that



sold-versus-needed rate ratios for one broker's customers are consistently lower than others. If a marketing representative is included in the project, then the representative can look into potential marketing differences. If an actuarial rep is included, then potential pricing issues can be further investigated.

- Besides improving the model, collaboration by stakeholders eases training and implementation. Leaders in each function will support the model to their respective teams.
- Modelers may not be aware of planned changes in company strategy that would affect model usage or data availability. Conversely, the model could suggest expansion to an area that runs counter to existing or proposed philosophy.
- Legal considerations for particular internal or external variables could exist. For example, variables may be deemed as unacceptable or create a risk of breaching individuals' private health information, and may be discouraged for use in a model. Even if a particular variable may be includable under the law, the legal team may wish to exclude it to avoid potential future litigation that could lead to poor company reputation or perception. Please note that different variable decisions may be made by a given company's legal team based on the particular line of business or risk being modeled.

5) THOROUGH TRAINING

The model isn't done when it's done. The users must understand exactly how to interpret and correctly apply the model results, to ensure full functionality. Often overlooked is the question, "Who will provide the training?" Or perhaps, "Who is most appropriate to provide the training?"

The modeling team may not be your best choice to provide training, as training may not be their specialty. However, internal teams dedicated to generic employee training may not fully understand a predictive model—why it is built this way and how to interpret and handle the results. You could consider including a member of the training team throughout the model build for maximum payoff.

If outsourcing a model build, the consultants are generally hired to build the model only. But then who (internally) fully understands it, and can provide training? And are the model formula and results proprietary, or may they be shared/distributed? You could consider expanding the scope of the consulting project to include full training (by the consulting team) of all users.

Training in marketing applications is especially difficult. We want to be able to express to brokers and agents the characteristics of our target customers, without revealing the proprietary modeling formula. How will we do that?

There is never one clear answer on how to train and who should perform the training. What is clear is that this step must be thoughtfully considered and appropriately executed in order to reap the full benefit of the model you just built.

6) REPORTING

Let's remember that the modeling project was initiated to improve something. Perhaps we wanted to improve customer retention, increase quote volume, improve risk selection based on profitability, or optimize pricing. We won't know how close we come to our goals (or how far we exceed them!) unless

CONTINUED ON PAGE 36

we develop comprehensive reporting packages to convey the results over time. Multiple reporting options should be developed to address the different needs of various users:

- Executives may wish to see the highlights in aggregate, by location, business unit or product.
- The marketing team may wish to see results broken down by broker/agent, or by location.
- Actuaries may wish to see more detailed results by predictive variable, by region, or by rate group.
- Underwriting leaders may wish to see results by underwriter, as a performance measure.
- Will each of these reports update weekly? Monthly? Quarterly? Annually?

You will also need to consider whether the reporting package will be pre-programmed for automatic generation of results, or if they will be individually calculated each month/quarter/year or as needed.

7) MODEL UPDATES

There must be a plan in place for updating the model. Will modelers rebuild every 18 months?

Three years? Five years? Does an “update” mean recalibrating coefficients for the current variables, or rebuilding the model from the ground up? It sounds almost ridiculous to think about V2 when V1 hasn’t yet started. But this is a necessary conversation, as management may not support a two-year build for a model that may only last for 18 months. Remember, too, that the model update schedule must mesh with your existing product or business strategy.

Conclusion

Modeling exercises require a deep commitment of staff, time and dollars. To avoid wasting these resources and to fully optimize the results of the project, a comprehensive range of operational considerations must be addressed. Only when “modeling” is accepted as a complete business strategy (versus simply a mathematical process) will it be as successful, profitable, efficient and impactful as possible. ■