

People Make Predictive Modeling Work

*By Rick Ezell
Senior Consultant, Wheaton Group*

Original version of an article that appeared in the November 29, 1999 issue of “DM News”

A student of the role of technology in direct marketing might characterize the past three to five years as the period of the ascendancy of software. This is especially true if focusing on trends in predictive modeling. Much software promises superior results based on the strength of the algorithms. This focus on software has led some to believe that buying the right software tool is all that is needed to produce effective predictive models.

This is a dangerous misconception caused by viewing predictive modeling simply as the discovery of an algorithm or equation within a mass of data. This is the goal, but achieving it requires a synergy of people and software. The software is, after all, only a tool. As with any tool, the quality of what is produced depends critically on the skill and intelligence with which it is used.

Before looking at how human skill and intelligence combine with the software to produce effective predictive models, let me define what I mean by “effective.” A predictive model is effective to the degree that it enables marketers to maximize, through improved selection, the return on their marketing investments.

Typical marketing investments aim to acquire new customers, generate revenue from existing ones or ensure the continued flow of an existing revenue stream by reducing attrition or churn.

People play a critical role in determining the effectiveness of predictive models in three primary areas: the planning of the analysis, insight in creating derived variables to include in the analysis and evaluation.

Producing an effective predictive model begins with developing an analytical plan. This is the blueprint for how to build the data set the software will process. The plan addresses a number of crucial questions, questions that only people – usually a combination of marketing, domain and analytical experts – can answer. Some of the questions it must address are what behavior to model, how to define it in the data, who to include in the model sample and how to build the sample.

Marketing experts typically determine what behavior to model and who to include in the sample because these issues are closely tied to the marketing goals. Consider a simple example from cataloging. In modeling the response to a campaign, should response, sales or profit be modeled? The answer will depend on a number of factors such as the variability in order size and the correlation between response and order size.

Domain and marketing experts and statisticians typically address how to define the behavior in the data and how to construct the sample.

Modeling attrition for a credit card issuer who charges no annual fee presents a challenging example of how to develop an effective definition of attrition. Marketers recognized that most attrition was not from customers canceling their cards. Rather, it was silent attrition from customers who simply tossed their cards in a drawer and stopped using them.

Marketers identified the characteristics of attrition, but it required working that definition out in the data to achieve a solution. It further required checking that definition by ensuring that a relatively large proportion of customers whose behavior satisfied that condition did not resume using their cards at a later date. Only people, not software, can propose meaningful answers to marketing questions.

A frequently overlooked area where people can greatly enhance the effectiveness of models is in the creation of derived variables. Derived variables are built from a company's existing data. They are typically time-ranged variables, ratios, deltas and other combinations and divisions of data elements. One of the greatest mistakes modelers can make is to fail to create and examine such variables for their utility.

Customer data presents almost endless opportunities to create variables, but domain and marketing experts will be able to identify those variables that have the most potential to improve the model. While the software is indispensable for creating these variables, people identify which ones to use.

People also play a critical role in the evaluation process. In this process, humans interact most closely and critically with the modeling software. Good software tools provide automated procedures for conducting a number of well-defined, repetitive tasks. For example, it will provide meaningful descriptive statistics on each variable that may potentially be used for prediction. It should provide this in easily-readable tabular form.

Some software tools also have good visualization capabilities. It is up to people, however, to evaluate this output, determine which variables have unreasonable or suspect distributions and assess the severity of missing data on a variable by variable basis. It is also up to people to determine how to repair data, when possible, and how to deal effectively with missing data. The software can provide the information, but people must make the assessments and decisions.

Good software will also automate the exploratory data analysis process. This process looks at the relationship of every potential predictor variable to the target variable. Tabular and graphical output as well as statistical measures aid the analyst in assessing the strength and form of these relationships.

Evaluation is critical. For example, the data analysis may reveal some counter-intuitive relationships. Marketing, statistical, and domain experts must determine whether the revealed relationship is correct or an artifact of a coding error. Other, more subtle cases require determining

whether a variable, although potentially a useful predictor, is unsuitable for use because of operational issues surrounding its capture and storage.

Finally, we come to what much of the software on today's market does best – discovering the algorithm. Numerous software developers have produced elegant packages that completely automate this process.

Depending on one's philosophy about predictive models, evaluation will play a greater or lesser role. At a minimum, however, evaluation is required to estimate the effectiveness of the model. This is done most typically through the joining of the gains chart with financial measures and marketing goals in order to estimate the economic "lift" provided by the model.

So, take advantage of the numerous packages that automate building predictive models. Many have delivered on their promise of taking technical experts out of this process. Remember, however, that they've done this for only one part of a complex process. The most effective predictive models will continue to be those built from the synergy of people and software.