Visualytic Solutions™
Improving Tomorrow with Today's Information

Karen Haddad Ziton, MBA
10000 Ridge Line Drive • Cleveland, OH 44133-3100
office 440.582.2971
cell 440.390.0441
fax 440.582.1131
karen.ziton@visualytic.com
www.visualytic.com

Predictive Modeling
A Brief Overview

He [Wayne Gretzky] claimed that the reason he was so good at hockey was that he didn't go where the puck was; he went where he felt the puck was going to be.

Expert intuition and natural ability are a rare combination of talents. For the other 99.9% of us to compete successfully, predictive modeling helps provide that insight.

Predictive Modeling forecasts the ***probability*** of future outcomes based on the past performance history of its predecessors (previous clients, portfolios, claimants) or a peer population. Each outcome being modeled (dependent variable) has a relationship to multiple characteristics (independent variables) which define the population. Defining the performance timeframe is significant because the outcomes most resemble the predictions at the same point in time at which the development population was measured. Model results should be validated with a hold out sample after development and before implementation and then regularly at intervals that resemble the performance timeframe.

Businesses utilize predictive models as a guidance tool in the relationship with their customers (or possibly employees). Three of the more common types of models are:

- **Response Models** are most often designed for marketing initiatives. These models rank the likelihood of potential customers to respond to an offer and thereby limit the number of prospects solicited and the associated marketing costs.
- **Risk Models** can predict the likelihood of financial losses or claims, allowing businesses to be selective in their approval processes, pricing, or monitoring.
- **Behavioral Models** are commonly used to determine account management strategies. These models recommend actions based on outcomes and timelines.

Steps are similar for each of the primary types of predictive models.

1. **Design Scope –** Determine project goals and timeline including type(s) of predictive models and data needed, define success and failure criteria (dependent variable 'good' and 'bad') and the performance timeframe.
2. **Data Collection –**
   - Identify and organize internal data (application data, account history, claim detail, performance history)

- Purchase and collect external data (demographic, census or economic information; credit bureau data if it's for an admissible purpose, or block level zip +4 credit information; industry statistics)
- Determine how to handle rejects and missing data
- Aggregate data into the modeling dataset and hold out a portion for validation testing

3. **Segmentation** – Since major subsets of a population can perform differently, it must be determined if multiple models are needed for a more complete solution. Modeling populations can be segmented subjectively based on experience or through statistical methods which measure splits such as CART decision trees or cluster analysis.
4. **Rank Variables –** Implement statistical procedures to rank the independent variables so that a more manageable pool of variables (less than 100) can be analyzed for modeling.
5. **Exploratory Data Analysis** – Conduct bivariate analysis to look at the relationship between each independent variable and the dependent variable. Any variables which show significance will carry into the final modeling dataset.
6. **Modeling –** Build the model using a multivariate analysis, usually a type of regression (linear, logistic, etc.), which will calculate the relationship of each independent variable to the dependent variable. After numerous iterations, the most significant variables are maintained and those that are highly correlated are removed until a final equation is determined.
7. **Score –** Represent the probability that is being predicted by the model with a chosen scale (1-10, 1-100, 1-1000 or other). Based on the final equation and scale, each observation is scored.
8. **Validation -** Group the scores in brackets (generally percentile distributions) to conduct statistical testing on the validation population and compare to the modeling population. A stable model will have very close results between the modeling population and the validation population. A standard test used in the industry is the Kolmogorov-Smirnov test, referred to as K-S. It measures the separation between the 'good' and 'bad' distribution, determining the extent that any model ranks above random.
9. **Implementation –** The score report is used to forecast, set approval levels, target customer segments, set collection strategies or whatever the goal of the modeling project. The scorecard must be maintained for use on future populations either internally (by the client) or externally (third party) and audited for accuracy. Routine validations should be performed periodically after implementation to ensure the continued validity and stability of the modeling system.

Predictive modeling is continually influenced by outside factors and is not a guarantee of success. However, it is a useful tool for objective, advanced decision-making. In closing, consider another quote attributed to Wayne Gretzky . . .

You miss 100% of the shots you never take.

