

Previews of course books are provided as an opportunity to see the quality of material used at TDWI educational events. The selected pages are representative of the entire course book. These pages cannot be printed.

The previews are intended to help you to select the courses that best fit your needs.



Predictive Analytics

*An Intensive Overview of Strategy, Application,
and Best Practices for Data Mining*

Tony A. Rathburn, Senior Consultant
tony@the-modeling-agency.com
(281) 667-4200 ext 221

Eric A. King, Business Development
eric@the-modeling-agency.com
(281) 667-4200 ext 210

Copyright © 2000 – 2008 The Modeling Agency

Notes

It's a Game... Play to Win!



- Games
 - You choose to play
 - You know the rules
 - You know how to keep score
- Goal: Improve your score!!!

2

Copyright © 2000 – 2008 The Modeling Agency

Notes

Businesses are created for a purpose. ...A purpose defined by their founders, enumerated in mission statements, and implemented through the policies and procedures implemented by boards of directors and management.

The procedures adopted differ from company to company, based on the judgments of those with responsibility. They establish internally defined “right” and “wrong” ways of doing things -- rules for playing their game.

Ultimately, the success of those procedures, and the careers of those charged with creating them, are evaluated based on a set of performance metrics: performance metrics that, like the internal rules for playing the game, are determined by the management of the organization.

As a participant in the game we call business, we must be aware of the rules by which our organization operates. If we are to succeed, we must understand how our organization keeps score.

Our goal, is to improve the score.

Math as Strategy

Formulas consist of three components:

- Variables
- Weights associated with variables
- Mathematical operators

$$Y = X_1(w_1) + X_2(w_2) + \dots X_n(w_n)$$

Every combination of variables, weights and mathematical operators results in a specific level of performance.

We build formulas searching for a strategy that enhances performance based on custom performance metrics.



3

Copyright © 2000 – 2008 The Modeling Agency

Notes

A manager collects data about a variety of factors. S/he develops their personal methodology for determining the priorities of the factors considered, and reaches a decision.

Mathematical formulas do exactly the same thing, in a more consistent manner.

Formulas use a combination of variables. Weights are associated with those variables and mathematical operators to reach a conclusion.

The major contribution of the application of mathematical formulas to decision making is the consistency of the decision making process.

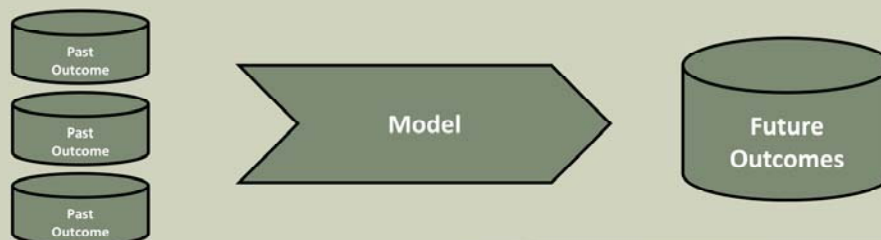
When we apply quantitative decision processes, we have the ability to evaluate the specific decision making capabilities. Additionally, we have the ability to make adjustments in the decision making process, and specifically evaluate the results of that change based on our specific performance metrics.

Every time we make a change to the variables included -- or to the weights associated with those variables, or to the mathematical operators -- we have a different decision process, with it's own specific performance score that can be evaluated.

There is nothing inherently right or wrong, good or bad, about any formula. The only thing we can evaluate is how it performs in the context of the decision environment in which we are operating. The formula is simply a strategy for making a decision.

Why Build Models

- Analyze data from past outcomes, searching for patterns
- Build models that identify groups displaying a behavior of interest at a rate different from the population as a whole
- Allocate resources based on the model's segregation of group behavior
- Continue to search for grouping strategies that enhance organizational performance.



Notes

Predictive Analytics works from the assumption that the best indicator of future performance can be found in the experiences we have captured in our historical data.

Predictive Analytics is the goal-directed analysis of large data sets to enhance organizational decision making.

Today's organizations collect vast amounts of data, organize it, explore it, visualize it, and analyze it in a variety of ways.

Predictive Analytics organizes data with a particular purpose in mind. Based on the decision process we are addressing, we identify an indicator of a behavior of interest and a variety of variables having potential impact in anticipating and evaluating that behavior.

We then build mathematical formulas to assist in reaching decisions regarding future behaviors in similar circumstances.

Predictive Analytics has a large number of techniques available to develop formulas, each achieving a specific level of performance or explainability. Those formulas are evaluated based on their relative performance. In Predictive Analytics, our goal is to find a formula with a level of performance on a specific decision process that significantly outperforms the current decision process.

Understanding the Environment



- Physical Systems have 'right' answers...
- Human Behavior is individual choice...
 - Inconsistent behaviors
 - Behavior changes over time
 - Probabilistic expectations by group description

5

Copyright © 2000 – 2008 The Modeling Agency

Notes

The application of Predictive Analytics in a business environment is generally related to the anticipation of human behavior. Unlike physical systems, there is no "right" answer in human behavior.

Human behavior is choice. People with very similar characteristics decide to act differently. The same individual displays inconsistent behavior patterns. Behaviors change over time.

Humans do not behave as physical systems do. Water always boils at 212 degrees. It doesn't say "I just did that yesterday" or, "I'll wait ten minutes for the break before I do that." As such, we can not accurately anticipate the behavior of individuals.

In Predictive Analytics, we use the variables in our model to identify a group. That group displays a behavior of interest at an expected rate. Our goal with individuals then, is determining whether or not they are a member of this group. If they are, by association, we have an expected probability of that individual displaying a behavior of interest -- as opposed to an individual who is not a member of that group -- or one who is a member of a different group.

This approach allows us to achieve the most common task in business decision making... it allows us to allocate resources.

Why Statistics Isn't Enough

Statistics Assumes:	Human Behavior Data:
Normal distribution of data	Not a normal distribution
Stable mean & standard deviation	Behavior changes over time
Consistency of characteristics	Behaviors are inconsistent
Data availability	Missing data
Causal relationships between variables and behavior	Behavior is choice
Orthogonal input variables	Incremental information content

Predictive Analytics gives us the opportunity to develop models that more accurately estimate real world behaviors

Notes

The seasoned practitioner in Predictive Analytics understands the importance of traditional statistics. There is nothing “wrong” with statistics, any more than a carpenter would find something wrong with a hammer or a screw driver. Statistical techniques are simply one in a set of tools that make up our toolbox.

To be effective, it is important that we understand the strengths and limitations of each tool available to us -- know when to use it, when not to -- and how to combine our tools in such a way that we achieve the desired goal.

Traditional Statistics Works...

- When we meet the assumptions and constraints of the tools.
- Human Behavior modeling violates many of the constraints and assumptions.
- Predictive Analytics acknowledges the reality of human behavior and provides the “opportunity” to develop more accurate models and achieve higher levels of performance.



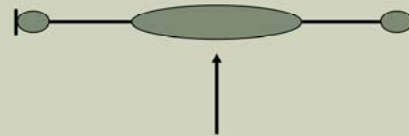
Notes

Predictive Analytics utilizes a variety of tools, each intended for the development of mathematical formulas based on historical data. By selecting tools and techniques that are appropriate for decision process we are supporting and, that are consistent with the characteristics of the data with which we are working, we have the opportunity to develop models that achieve higher levels of performance.

Shifting Our Focus

Statistics

- Concentrates on Central Tendency
- Gives us an “In General” perspective
- Required First Step



Predictive Analytics

- Focuses on the Tails of the Distribution
- Identifies Low Incidence / High Impact Occurrences

Notes

Statistics is typically our starting point. Statistical analysis tends to concentrate on the “in general” characteristics of our decision space. We focus on the central tendency, usually the average, for our population. It is critical that we have this perspective before we proceed.

In Predictive Analytics, we are looking for sub-groups that display our behavior of interest at a rate different from the population as a whole. In many cases, what we find are the low incidence, high impact groups. These groups may be relatively small, but their behavior may have a significant impact on our level of performance. By successfully identifying these groups, and the individuals who comprise them, we can adjust our allocation of resources to enhance performance.

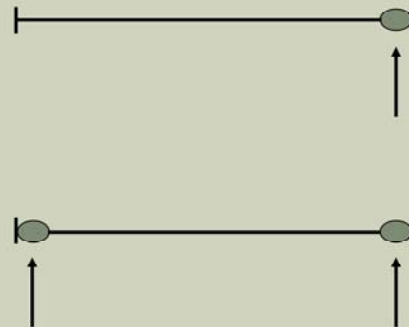
Predictive Analytics Models

Model Any Behavior

- Address “what-if” type questions
- Best performance in the tails of the distribution
- One tail, or both
- Identify groups that display behaviors at different rates
- Allocate resources appropriately

Classification

- Use a tool that matches reality
- Understand what you ‘don’t know’



Notes

Predictive Analytics models may focus on sub-groups at either end of the distribution. The behaviors displayed by these groups may significantly enhance performance or significantly hurt performance. In either case, it is usually a very straightforward resource allocation once we have a way of reliably identifying member of the group.

This approach also offers us the ability to recognize that there are going to be circumstances that we can not reliably classify. In these circumstances, we have the ability to fall back on the characteristics of the population as a whole and achieve the levels of performance that we have achieved prior to adopting a Predictive Analytics approach to decision support.

Predictive Analytics

- Is goal driven analysis of historical data sets
- Is a search for a set of variables, weights and operators that result in better performance in unseen data on a business decision process of interest
- Is not a completely automated process
- There are no silver bullets! It is a *process* of applying 'statistics on steroids' to adapt more closely to real-world applications and enhance your performance.
- Strategy is driven by an industry-standard process; Tactics are facilitated by modern software solutions



10

Copyright © 2000 – 2008 The Modeling Agency

Notes

In today's world, there are still a large number of purveyors of magic cures.

Predictive Analytics is not magic. Used appropriately, it is a logical mathematical approach to deriving sound decision making strategies based on the specific goals of your organization. Predictive analytics has many published cases of significant return to the organizations who have properly implemented the process.

Organizations that effectively apply predictive modeling best-practices will leverage highly valuable information hidden within the glut of data they've amassed. They will move from a reactive search and explore posture to a proactive decision support system with greater accuracy and efficiency.