



Predictive Analysis and the Business of Government

By Dr. Ramon Barquin

*President of Barquin International
&
BeyeNETWORK Government Expert*

Predictive Analysis and the Business of Government

By Dr. Ramon Barquin

We have all dreamed at some point of having a crystal ball that would allow us to predict the future. Who is going to be elected president in 2008? What is going to be the winning number in tomorrow's PowerBall lottery? It would be a lot of fun...and potentially a lot of profit.

For businesses, of course, being able to get a glimpse of the future can be extremely powerful from the point of view of corporate planning and resource allocation. More importantly, having an insight of what is going to happen allows a business to take action to prevent, augment or mitigate the results. If Delta Airlines could know today the number of passengers it is going to have for its 3:25 p.m. flight from Chicago-O'Hare to New York-La Guardia on February 21, 2007, it could better decide whether to assign a larger or a smaller plane for that flight or whether to allot it more or less frequent flyer seats. If Wal-Mart knew today how many Wilson tennis rackets it will sell at its Des Moines, Iowa store during May 2008, it would be able to stock precisely that exact number of rackets and/or take some marketing action if the number is below what it was planning to sell.

If you are in the public sector, you will not be as concerned with profits, of course, but the power of prediction is every bit as important. Just imagine how much easier it would be to run our Medicare programs if we knew exactly how many eligible beneficiaries would need to be treated for rheumatoid arthritis in the second week of November 2009 in Plano, Texas, and how many of those would also show diabetes complications. Or think how much easier the generals would sleep today if they knew that the Army's tally of new recruits in April 2007 was going to be 12.5% above their specific requirements. And then, there is that crucial post-9/11 objective to be able to predict where and when the next terrorist attempt is going to take place, and who is planning it, so we can preempt or prevent it.

While some of these predictions are still dreams in our wish list, we certainly have the tools today to do substantially powerful predictive analysis given a certain set of minimum requirements. I can recall that in the early days of data warehousing, there was a mantra that went: "First, be able to show how the business has done; second, be able to see real-time how the business is doing; and then, be able to predict how the business is going to do in the future." While it is important in business intelligence to not forget about the past (for compliance, enforcement, or recordkeeping) or the present (for management and operations), there is clearly maximum value in business intelligence to be able to get a glimpse of the future.

What are some of the tools that we have at hand today in the BI toolkit and how much can we expect to get out of them in government? In physics, prediction is best accomplished when we

can deduce, through observation and experiment, a specific formula that defines the relationships and interactions in a mathematical format. Hence we can “predict” that an automobile traveling at a velocity of 60 m.p.h. will take 2 hours to arrive at a destination that is 120 miles away, because we know that $T=D/V$ (where T is time; D is distance and V is velocity). We can also predict that the next total solar eclipse will occur in August 2008 and will be visible from Russia and Mongolia, since we now have very accurate formulas for planetary orbits.

Forecasting the behavior of economic or social systems, however, is often much more difficult because no “exact” formulas tend to exist. This is usually due to the complexity of the relationships, the interdependency of the variables, scarcity of metrics and many other complicating factors. The best we can usually do is to postulate some range of probabilities based on our analysis of existing data – data that has been collected about past events that permits us to reach certain tentative conclusions about what might occur in the future if our assumptions are valid. Life expectancy, a key input into insurance policies, is very dependent on the actuarial analysis of real events. People are born and they die and these events are recorded. From here, actuarial tables are developed that can be used for analysis. However, to determine how to price a specific policy for a person who smokes, or a person that suffers from untreated hypertension, or who works as a poisonous snake catcher, we must know a lot more about that individual and have data that allows us to analyze mortality rates for individuals in these specific classes as opposed to the general population. The more we know about our environment and the more relevant attributes we have that give us fodder for our analytical tools, the better our prediction will tend to be.

Data mining was usually premised on having software that is very good at working advanced statistical techniques and applying them to large amounts of data looking for patterns and anomalies. Neural nets are also often used as predictors by using them as a learning filter through which a specific model is refined. More recently, complexity theory, including chaos theory and genetic algorithms, are providing us with even more powerful predictive tools and techniques. In all cases, though, we have to depend on the data. Where we have no data from which to develop an initial model for prediction, there will be a period of data gathering and collection from which to start.

The U.S. government has the largest data collections in the world. It collects data on just about anything and everything. Using predictive analysis tools on much of that data could yield very useful results to assist public sector management and citizen service. Aside from the obvious area of national security and homeland defense, where much is going on right now, I would like to suggest three particular areas that should be explored as targets of opportunity in the near to medium term: healthcare, education and fraud detection.

One of the main challenges we face with the aging of America is attempting to estimate and plan for the healthcare needs of the retiring baby boomers. Being able to predict the number of cases of specific diseases, by geographic area, and their outcomes based on treatment could be a huge factor in mitigating the problems that will accrue. With the data we have from the CMS (Center for Medicare and Medicaid Services), the CDC (Center for Disease Control) and the NCHS (National Center for Health Statistics) as well as Census, Veterans Health Administration and

other relevant source data, we should be able to do a good job with the predictive analysis tools already available.

We have been flailing as a nation for the last few decades in terms of what works in education. If we can predict with reasonable accuracy the outcomes from the teaching of certain subjects, using specific techniques, by teachers with given attributes, to students with a known set of demographics, then we can design the best curricula and hire the right faculty to obtain the desired results.

“Fraud, waste and abuse” has been for years a very popular phrase, rightly or wrongly, used to characterize operations in government. Leaving aside the issue of legislative “pork” and intentional political earmarking or directions, there is much to be saved – billions of dollars – if we could identify and predict specific areas where fraud was occurring. Some of the juiciest environments to explore are in applications for government benefits from ineligible individuals (i.e., health claims, housing assistance, food stamps) and in government contracts (i.e., overpricing, false reporting).

While there are clearly going to be serious privacy considerations to address in all of these areas, predictive analysis is very much an area of opportunity for business intelligence in the public sector.

About the Author

Dr. Barquin is the President of Barquin International, a consulting firm, since 1994. He specializes in developing information systems strategies, particularly data warehousing, customer relationship management, business intelligence and knowledge management, for public and private sector enterprises. He has consulted for the U.S. Military, many government agencies and international governments and corporations.

Dr. Barquin is a member of the E-Gov (Electronic Government) Advisory Board, and chair of its knowledge management conference series; member of the Digital Government Institute Advisory Board; and has been the Program Chair for E-Government and Knowledge Management programs at the Brookings Institution. He was also the co-founder and first president of The Data Warehousing Institute, and president of the Computer Ethics Institute. His PhD is from MIT. Dr. Barquin can be reached at rbarquin@barquin.com.