



By Richard Boire

August 7, 2005

The Future of Data Mining in Marketing – Part 1

As human beings living in today's world, we are an extremely lucky group of individuals. Although problems do exist and are certainly magnified by our ever-incessant all-news type channels, the reality is that many of society's ills and problems have been mitigated over time. Yet, as individuals living in today's world, our largest challenge is to deal with changes in the world around us which is unprecedented when compared to previous generations. Given this observation, I'd like to focus the thrust of the next three articles on change and specifically how it will impact the world of data mining. Accordingly, the next three articles will deal with the future of data mining within marketing. The first article will deal with some general observations and commentary on what has happened over time. The second article will deal more with the specifics as to how technology, software, data and of course human beings will all interact within this future world. The third article will consist of interviews with leading-edge business practitioners in order to obtain their insights and views on how data mining will evolve in the future.

As we all know, the world today is experiencing tremendous change in virtually every facet of society. Political structures as evidenced by both the fall of communism and dictatorships, sports with their new collective bargaining agreements, businesses with their obsessions to become flatter and leaner all with the intention of reducing head count, and technology in its all-encompassing goal of making us more effective with our use of information, have all contributed to the dynamics of these vast structural changes. Nowhere is this more cleanly demonstrated than within the world of technology and specifically data mining. Having spent the last 22 as a data miner, I have witnessed how technology has impacted our data mining efforts of not only producing better solutions but also building them in a much quicker fashion.

Before discussing some of these implications within the marketing world, it is worthwhile to observe how data mining is being used in other areas. For example, in the area of health, increased access to data and information has facilitated the development of new drugs which seem to be produced at ever-increasing speeds. By being able to quickly analyze volumes of data from all kinds of different tests and during different time periods, data mining represents a critical cog within the drug development process. Besides the development of drugs, data mining efforts have also helped play a key role in the mapping of the DNA genome. The use of extensive computer firepower in conjunction with sophisticated data mining tools helped in determining the specific combination of genes within the DNA genome. What was once a few years ago almost considered impossible entered the realm of the possible through access to these

newfound capabilities due in large part to data mining. According to scientists, this breakthrough discovery of being able to map the DNA genome offers limitless possibilities in the development of new drugs now that specific genes can be isolated.

Another area that has taken advantage of data mining has been crime prevention. In Rudolph Giuliani's autobiographical book on 'Leadership', he cites what New York did within this area. Although the book is somewhat self-promoting, the facts do speak for themselves. In the late seventies, I remember the murder rate of New York City comprising approximately 2,200 murders a year. Overtime, under Giuliani's leadership, this rate was reduced to between 600-700 murders a year, roughly a third of what it was in the late seventies. Now, there were other factors at play besides data mining. Bill Clinton's bill to increase the number of police officers in the U.S. by 100,000 certainly helped in this area. The zero tolerance policy, initiated by Giuliani, represented another factor, which resulted in the full prosecution of even minor felonies such as the illegal placement of graffiti on public property. But Giuliani also used technology and data mining to analyze data concerning crimes in all sectors of his city. With data mining, they were able to analyze massive amounts of data that allowed them to uncover trends and patterns concerning future crime behaviour within high risk areas. This information led Giuliani to more effectively deploy his law enforcement activities to those areas which were deemed high risk. Sound familiar? This is not too dissimilar from the marketing world, whereby we identify key groups of customers and then deploy more of our marketing resources to those groups in the hope of optimizing profit within a specific marketing campaign.

Another current example of data mining's growing importance is within the public sector and in particular the U.S. government. The Total Awareness Project represents an initiative which has been funded by the U.S. Government at a scale involving billions of dollars. Under the auspices of the Pentagon, the mandate of the project is quite clear. Through the use of information and data collected by all persons, the Pentagon is attempting to utilize powerful data mining tools and technology that will allow it to develop better profiles of what constitutes a terrorist. The U.S. Government would be able to utilize all individual data relating to key areas such as credit card/banking activity, health-related activity, criminal activity, insurance behaviour and ultimately any other areas where one can capture individual-level data. Though this tremendous access to data in combination with powerful data mining tools, it is the hope that the government can be more proactive in identifying potential terrorists before they embark on their heinous activities. Obviously, this represents one of the many U.S. responses to 911 which have been in many cases unyielding and unrelenting in their efforts to make the U.S. a safer place. There have been many critics both within and outside the U.S. that have deep concerns about this specific initiative. The privacy advocates within the government have been quite vocal in their attempts to portray this initiative as a draconian measure which will severely restrict the freedoms and liberties of the individuals that have been a cornerstone of their philosophy of government. Currently, the debate ensues between the attainment of these tools and solutions that could further empower the anti-terrorist activities of their law enforcement officers versus the

restriction of basic civil liberties. At the present time, these initiatives are still being discussed and debated within the various levels of the U.S. Government.

Now let's examine to some extent how data mining has evolved within the marketing world. The development of data mining tools to help marketers actually began in the sixties. In the early eighties, which is when I began my career, data mining efforts for selecting customers were focused on using regression type techniques within mainframe computers that occupied the area of a typical 30-pupil classroom. The two largest constraints which prevented data mining from becoming a more all encompassing business tools were both storage and CPU.

The first limitation was the amount of data that could be stored within a PC. Traditionally, data miners have always been very cognizant of how much data could be captured within the hard drive of their computer. Knowledge of sample size theory was a paramount requisite for data miners in those days. Sample size theory consisted of determining what was a representative sample given the desired solution. It also consisted of balancing this objective against a volume of data that would produce a certain level of error. Smaller samples will produce larger errors but are more easily accommodated within a restrictive storage environment. Data miners in the early days were constantly striving to find the right balance of sample size that would provide an optimal solution. You can imagine the thrills, perhaps a strong euphemistic term here, when in the early to mid nineties, we received a PC that had a 3 gigabyte hard drive. We were considering 'cracking open the champagne' as our sample size thresholds could be significantly increased. In today's world, it is common for our staff to have hard drives with at least 60 gigabytes. This increase from 3 to 60 gigabytes in 10 years certainly provides one small example of exploding technological changes and its impact within the data mining world.

The second constraint of CPU prevented certain mathematical techniques from being practically deployed within the business environment. These techniques are commonly referred to as being "CPU-intensive." The nature of many of these mathematical routines involves the development of complex matrices which require the use of extensive interactive routines. In many cases, it is these exhaustive interactive routines which place an enormous burden on the CPU's capability. The increased advancements in computer technology within this area now make it possible for the deployment of these more sophisticated mathematical tools. Examples of these tools which are used in the business work environment include techniques such as neural nets, genetic algorithms, and machine type intelligence.

Besides the increased capability of working with larger amounts of data and more sophisticated techniques, data miners are also experiencing increased speeds in terms of data processing. Ten years ago, routines that used to take ten minutes to run can now be completed in under a minute. These capabilities now allow the user to explore a variety of other options. Rather than focus on approaches which maximize our effectiveness in processing data, the focus is now shifting on the data miner to both identify and explore other solution type options. Data miners can now run multiple

routines in simultaneous as well as sequential fashion. Results and information can now be observed by the data miner who now needs to be able to synthesize all this newfound access to information. At the same time, the data miner needs to exercise that innate human quality of thinking through the results in order to arrive at a solution or solutions that will meet a given business objective. It is this ability to think through all this information that will be in increasing demand especially as technology continues to provide both increased access and data processing capabilities. Technology will simply increase the demand for this skill set as the role of the data miner evolves to one that becomes more exploratory and open-ended in their approach to developing solutions. This shifting emphasis from being historically viewed as data processors to one of being viewed today as problem solvers represents yet another example of positive change. Businesses will continue to look at data mining as a critical area of investment which will require more and not less resources. Although we have all witnessed the many downsizings that are occurring within our world today, there are always emerging growth opportunities. As my grandmother used to say: "One door closes, another one opens." In the case of data mining, the door is definitely opening.

Richard Boire is a Partner with the Boire Filler Group, a database marketing consulting company that specializes in developing and implementing data mining strategies. He can be contacted at (905) 837-0005 or via e-mail at RichB@BoireFillerGroup.com.