Data Mining

AN EXCLUSIVE INTERVIEW WITH

DR. AMJAD ZAIM, CEO, COGNITRO ANALYTICS

WHAT IS DATA MINING?

Data mining refers, in general, to the entire process of gaining various insights from large volume of data for the purpose of making informed decisions. Because critical information is usually buried deep within the data, the analogy to gold mining symbolizes the challenge overcome by the data mining technology. Predictive Analytics adds another dimensionality to Data Mining, enabling one to fast forward in time and forecast future trends and predicting complex patterns. Both Data Mining and Predictive Analytics borrow from several scientific disciplines such as math, statistics and even psychology. The term Business Intelligence, or BI, extends the definition of data mining as a technology to include all the necessary software tools and business applications used to effectively mine the data and obtain hidden "intelligence" of business value.

The word "Analytics" is also becoming a familiar term. In general, the word Analytics describes "science of analysis". It places data mining concept in the main context of either the application or the industry in focus. For example, banking analytics is the application of data mining in the banking and finance sector in general, whereas CRM analytics encompasses the process of mining customer data to uncover consumer behavior across several industries including the banking industry.

HOW DOES DATA MINING HELP BANKS?

Data mining helps banks profit from their customer and financial data.

Banks are constantly struggling with global competition and market positioning, adapting new strategies to help distinguish its core market, build and expand customer base, and manage market uncertainties. To guide their strategy, they must continue to sift through their data and discover new drivers for success. Depending on what type of data is available to a bank, data mining can help answer several common questions that are known to have a direct impact on a bank's bottom line, such as which customers are likely to default on loans? which are likely to be profitable and loyal customers? and which are likely to engage in fraudulent transactions?. All these questions are related to credit, customer and operational areas, respectively, and are of common high concerns to banks. From a strategic point of view and in accordance with the Basel II Accord, data mining can help financial institutions meet industry standards by demonstrating improved ability to minimize risk especially in the aftermath of the current economic crisis. For example, by reducing risk weighting, a factor that is used to calculate capital reserve requirements, banks can free up some additional capital for profit-making activities such as loans and securities.

There are also other less common yet more complex problems related to investment decisions that can be solved with Analytics, such as portfolio optimizations and stock performance predictions.

CAN YOU GIVE AN EXAMPLE?

There are numerous examples and case studies that show how banks have been rewarded generously for investing in the data mining technology. For example, the National Bank of Canada was able to increase customer retention rate by 40% by implementing an intelligent system that enabled customer service to identify and retain potentially churning customers. The Royal Bank of Canada also saved over $15m after the implementation of a fraud prediction measures. The Bank of America also reported a net reduction rate of 17% in loan defaults which allowed them to effectively manage and control credit risk. The long term return in all these cases clearly far outweighs the initial investment in data mining technology and resources.

WHAT TYPE OF DATA DOES A BANK NEED TO HAVE TO ENGAGE IN DATA MIN?

Depending on what type of questions a bank is looking to get answers to. If it's looking to improve customer relations and/or enhance its branch performance, it then needs to have customer characteristics (age, income) as well as behavioral data (ATM and bank transactions, fees, purchased instruments, balances, tenure, etc.). If a bank is looking to improve its lending process, it may need to examine a long history of the demographics and financial profiles of defaulters, collect econometric data reflecting credit market, as well as a broader base of credit data collected from credit bureaus or credit agencies. Surveys data can also be used to understand customer perception, measure their level of satisfaction and loyalty, and assess the quality of services at different branches and points of transactions.

WHAT ARE THE MAIN OBSTACLES FOR DATA MINING IN BANKING?

The main obstacle in most cases is the data. With data being the building block of knowledge, it's hard to imagine that any meaningful insight can be mined from inaccurate or inconsistent pool of data. Flawed and corrupt data seem to be a fact of life. This is evident by the fact that data cleansing and preparation, which is a painstaking process that is far from intuitive, usually comprises a substantial 40-60% portion of the project. Therefore, data quality does pose a significant challenge to data mining and can diminish confidence in the reported information if not addressed accurately.

Another big obstacle is lack of understanding of what data mining really is. You can think of a data mining models as the "brain" that drives the knowledge discovery process. The main challenge then becomes selecting the best and the most efficient combination of resources (people and technology) and methodologies (algorithms and heuristics) necessary to produce the best model and carry out effective data mining projects. Many companies rush to invest in expensive software and IT solutions that end up being greatly underutilized. Data mining initiatives should originate from organizational business units that have a clear and well-defined business goal in mind. The main effort should then be directed at developing a data mining model that is tightly matched to the business goal. Clearly, the IT component is not a major constituent, and IT systems should only be viewed as vehicles that allow the discovered knowledge to be properly channeled and reported to the designated managers and analysts.

Knowledge barrier is also another major challenge to data mining. Applying data mining usually requires people with a good background in math and statistics and sufficient understanding of somewhat complex algorithms and convoluted heuristics. This is usually found in people with an advanced degree in applied math and statistics. Also, developing an analytics intuition and domain-specific expertise also requires years of experience and exposure to best-practices. All this makes data mining competency anything but abundant, and is the reason why data mining skills are amongst the highest in demand in the US and around the world.

WHAT ARE THE MAIN REQUIREMENTS FOR A BANK TO IMPLEMENT BANKING ANALYTICS IN THEIR OWN BUSINESS? AND WHAT KIND OF TIMELINE IS INVOLVED IN GETTING ANSWERS FROM THESE KINDS OF DATA MINING IMPLEMENTATIONS?

Aside from good data, there is not much requirements for a bank to engage in any banking analytics project. Once the data mining model is developed, tested and verified, the results can be graphically presented in structured reports with visual aids and illustrative figures, or integrated into the bank's IT environment in the form of dashboards or scorecards. Most business intelligence solutions, such as Oracle BIEE, are capable of importing and reading data mining models and publishing the discovery electronically.

In terms of timeline, each project/client is unique and there is no unified "one-size-fits-all" solution. Depending on the scope, every project is assessed independently during the initial business understanding phase, but most of our projects have ranged from 4-6 months of time window. In general, factors that influence the duration of a data mining project include the intricacy of the business question, the quality and integrity of the available data, as well as the level of commitment of the project stakeholders.

IF YOU COULD LIST THE CRITICAL SUCCESS FACTORS FOR DATA MINING, WHAT WOULD THEY BE?

The most important factor is to have a clear and unified consensus amongst all stakeholders on the ultimate business problem to be addressed. Questions that are best suited to be answered by data mining include describing an uncertain and complex business phenomenon or making an inference about future events or behavior from a wide array of variables. Questions that require summarization or simple description across a few variables are usually answered more efficiently and effectively using traditional and simple descriptive statistics, in which case employing data mining may not be necessary.

Another critical success factor is planning and management of data mining projects. All too often, data mining projects fail because they rely wholeheartedly on technology and do not put enough emphasis on strategy. Even the most successful solutions are doomed to fail without proper preparation and analysis. Simply harvesting every byte of company information and feeding it into a technology solution will not deliver the results needed to make better business decisions. As advanced as today's technology is, no piece of equipment or software will determine what type of information you need to enhance business operations. For that, you need a select group of seasoned data mining experts working as a team under the su-

pervision of a project manager towards solving the business problem.

CAN YOU TELL US MORE ABOUT COGNITRO ANALYTICS? WHAT KIND OF SECTORS AND BUSINESS QUESTIONS DOES COGNITRO ANALYTICS COVER AND TARGET?

Cognitro Analytics was founded in the US in 2004 by a group of scientists working at the forefront of analytics and pioneering advanced artificial intelligence methodologies. Today, Cognitro Analytics is to enable companies to profit from their data and compete on those advanced analytics. Our company helps institutions, including banks, to better manage risk, optimize marketing, uncover fraud and retain customers by maximizing the value of data to make more insightful and informed business decisions.

We mainly cover telecom, financial, retail and customer-centric companies with a revenue scale that is sufficient to realize substantial return on investment. We're witnessing a growing interest amongst other sectors to embrace data mining such healthcare, manufacturing and transportation. Across all these industries there are some common business questions that are relevant to marketing, operation and even HR. For example, some common questions include: who my loyal customers are, what are the main drivers for customer retention and attrition, and what is the lifetime value of a customer to the bottom line.

CAN YOU GIVE AN EXAMPLE OF A DATA MINING PROJECT IN BANKING?

A US retail bank wanted to gain a better understanding of its customer perceptions and their impact on the bank revenues. A customer survey data was administered, collected and used to construct a data mining customer retention model. A set of customer satisfaction metrics were then developed and measured against the waiting time in three different segments of line waiting time: 1:05 hr, 1:30 hr, and 2:00 hr. We found that customers in districts with long wait times will remain customers, but few will open new accounts or recommend the bank to others. The findings were used to reshape the bank customer relation management strategy and improve customer satisfaction and loyalty.

WHAT GEOGRAPHICAL AREAS DOES COGNITRO ANALYTICS COVER? HOW ABOUT LEBANON AND THE MIDDLE EAST

Our operation is mainly based in the US, but we're growing worldwide. We're also in the process of establishing a regional office in the Middle East. ACT (Automated Computer Technology) is one of our premier local technology partners that is helping us introduce this new technology in Lebanon. By combining analytics with IT, we're better positioned to offer a comprehensive business intelligence services and deliver complete end-to-end service solutions to organizations in a more seamless manner.

WHAT KIND OF ACTIVITIES DOES COGNITRO ANALYTICS HAVE IN LEBANON?

We already started a number of analytics projects with a number of financial institutions and we are optimistic about its future in the area. Over the next few months we will be offering a number of training programs that include seminars, courses and workshops aimed at increasing awareness in the field and elevating the analytics skills in the region. We are also planning a regional seminar entitled "The Road to Successful Banking Intelligence" to be held late this year in close coordination with our partners in Lebanon such as ACT and Oracle.

WHAT IS THE FUTURE FOR DATA MINING AND FOR COGNITRO ANALYTICS?

The data mining has been undergoing a rapid growth since the start of the new Millennium. While the whole global software market declined by 8% in 2009 compared with the previous year, data mining and BI did not share the same fate and actually showed solid double-digit growth, even in 2009. As the economic revival continues, BI will further stimulate the growth of the software market, and data mining will mature in many ways. Banks will soon be able to run complex data mining and predictive analysis in real-time. The use of data mining as a service will also allow widespread use by enterprises at various scales. The introduction of data mining technology into microprocessors will create a new generation of "smart" appliances and hardware. It's hard to envision a plateau point where the technology will begin to level off but we believe that we're still in the early days of the technology.

Cognitro Analytics is constantly searching for new business phenomena to be addressed with innovative analytics and is committed to pushing the boundaries of data mining beyond the obvious classical applications. Working with researchers from world-class institutions, we're able to bring the latest advancements in data mining to business in no time. Cognitro Analytics is also teaming up with other BI solution providers to validate, integrate and benchmark new analytical models and leverage their software functionalities. In the Middle East, we're also expanding our network of strategic alliances that will enable us to bring this new emerging technology to a relatively raw and untapped market.