Data Mining Tools: Which One is Best for CRM? Part 1

Information Management Special Reports, January 2006

Robert A. Nisbet

It has been several years since I ventured forth to answer the question, "How to Choose a Data Mining Tool Suite." That article was organized around two central concepts:

1. There is no "best" tool; that is, no best tool for everyone.
2. The most useful tools are those that facilitate well the greatest number of tasks in the kind of data mining applications that you need to perform.

Major Data Mining Tasks

In the past, data mining tool development has focused primarily on providing powerful analytical algorithms. However, the analytical "engines" handle only a small part of the complete task load in a data mining project. As most data miners know, 70 to 90 percent of a data mining project is consumed with data preparation. Development of tools for data preparation has taken the backseat in most data mining tool evolution. Finally, you must be able to evaluate models properly, in order to compare models, and commend them to marketing staff.

DATA PREPARATION TASKS

Common data preparation tasks include:

- Data assessment to determine:
  - Missing values (blanks, spaces, nulls)
  - Outlier values
  - Collinearity assessment (related to correlations between predictor variables)
  - Frequencies of multiple codes in a given variable;
- Merging multiple datasets;
- Mapping metadata (field names and types) from various input formats into a common format for analysis;
- Transforming contents of similar variables into a common format;
- Changing data types from numerical to categorical data types (by binning and classification) and from categorical to numerical data types for use with algorithms with specific input requirements;
- Splitting variable codes into separate fields, and combining of multiple fields into a single field; and
- Deriving new variables from existing variables. Most data miners discover that some of the most predictive variables are those that they derive themselves.

Most data mining tool sets only "minor" on these important data mining tasks. This evaluation will "major" on the ability of common data mining tools to facilitate these tasks.

In addition to providing tools for doing important tasks of preparing data for modeling, a good data mining tool for direct marketing should include tools for evaluation of the models created by the modeling exercise.

MODEL EVALUATION TOOLS

In analytical theory, the best model is one that has the greatest accuracy in predicting all classification states of the target variable and is acceptably robust in its agility to perform well on the validation data set. That means we must consider the combined accuracy of predicting responders and nonresponders. This approach is called the Global Accuracy method. Most data mining tools use this method to identify the "best" model. However, there is a "fly" in this ointment. Embedded in the theory behind the Global Accuracy evaluation method is the assumption that the costs of all types of classification errors are the same. This approach works well in the classroom, but it does not work well in CRM data mining operations, particularly those that drive direct mail (DM) campaigns. In fact, this is one of the major reasons why many CRM initiatives to support DM campaigns have failed to produce much business value in the past. Models have been evaluated largely on a basis that is only partly relevant to the only things that marketers care about: maximizing positive customer response and minimizing the cost of doing so. Most data mining tools focus on the combined accuracy of prediction but ignore the cost element entirely.

In DM campaigns, the cost of mailing to a prospect that does not respond (referred to as a "false-positive" error) is rather small; but the potential cost of *not* mailing to a prospect that would have responded ("false-negative" error) can be rather large (reflected in the lifetime value of membership fees not paid and other services not purchased). This means that DM model evaluation methods should focus on minimizing the false-negative errors, rather than the false-positive errors. Because marketers care only about response rates and costs, a mailing to the top three deciles that hits 60 percent of the responders is likely to satisfy both concerns. Mailing to the non-responders (false-positive errors) in the top three deciles is an acceptable cost to the direct marketer for the sake of contacting 60 percent of the total responders available in the target area. This situation represents a 100 percent lift over random expectation and is much more cost-effective than a mass mailing approach.

Most data mining tools employ the global accuracy method for model evaluation. You may be forced to accept this method to identify the "best" model using the tool's reporting capabilities. The best model among many performed with different algorithms should not be evaluated by comparing the accuracy reports of each tool. Rather, evaluation should focus on how well the model clusters the positive responders in the top deciles of a scored list sorted on the prediction probability. Even classification algorithms can output classification probabilities. The actual classification (e.g., 0 or 1) is a highly summarized expression of the classification probability (e.g., <0.5 = 0; ≥ 0.5 = 1). Here lies a lot of the true "gold" hidden in the capability set of the tool. The naive CRM data miner will focus on the classification and accuracy thereof, but the true "gold" of CRM data mining must be expressed in terms of *probabilities* for retention, purchase and new customer acquisition.

A cumulative lift table (e.g., Table 1) must be inspected to determine how effective the model is in clustering true-positives in the upper deciles. This table can be created by:

1. The prediction probabilities are sorted in descending order.

2. The sorted list is divided into 10 segments (deciles).

3. Count the number of actual hits (actual responders in the modeling dataset) in each decile.

4. Calculate the random expectation per decile by dividing the total number of actual responders by 10. This means that 10 percent of the total responders are expected in each decile. If the percentage of hits exceeds the random expectation, the model provides a lift in that decile (over random expectation).

Decile	Hits	TP	FN	TN	FP	Random Hits	% of Total	Cum % of Total
1	81	81	0	0	735	27.5	29.45	29.45
2	43	18	25	411	361	27.5	15.64	45.09
3	39	0	39	777	0	27.5	14.18	59.27
4	30	0	30	785	0	27.5	10.91	70.18
5	11	0	11	805	0	27.5	4.00	74.18
6	7	0	7	808	0	27.5	2.55	76.73
7	11	0	11	804	0	27.5	4.00	80.73
8	21	0	21	794	0	27.5	7.64	88.36
9	16	0	16	799	0	27.5	5.82	94.18
10	16	0	16	798	0	27.5	5.82	100.00

Table 1: Lift Table with Coincidence Counts

True-Positives (TP): the number of correctly predicted responders

False-Negatives (FN): the number of incorrectly predicted responders

True-Negatives (TN): the number of correctly predicted non-responders

False-Positives (FP): the number of incorrectly predicted non-responders

The analysis of the lift table shows that the incremental lift (percentage of total in the eighth column) declines below the random expectation (10 percent per decile) after the fourth decile, containing over 70 percent of the total responders. This crossover to negative lift can be seen graphically in an incremental lift curve (Figure 1) below.

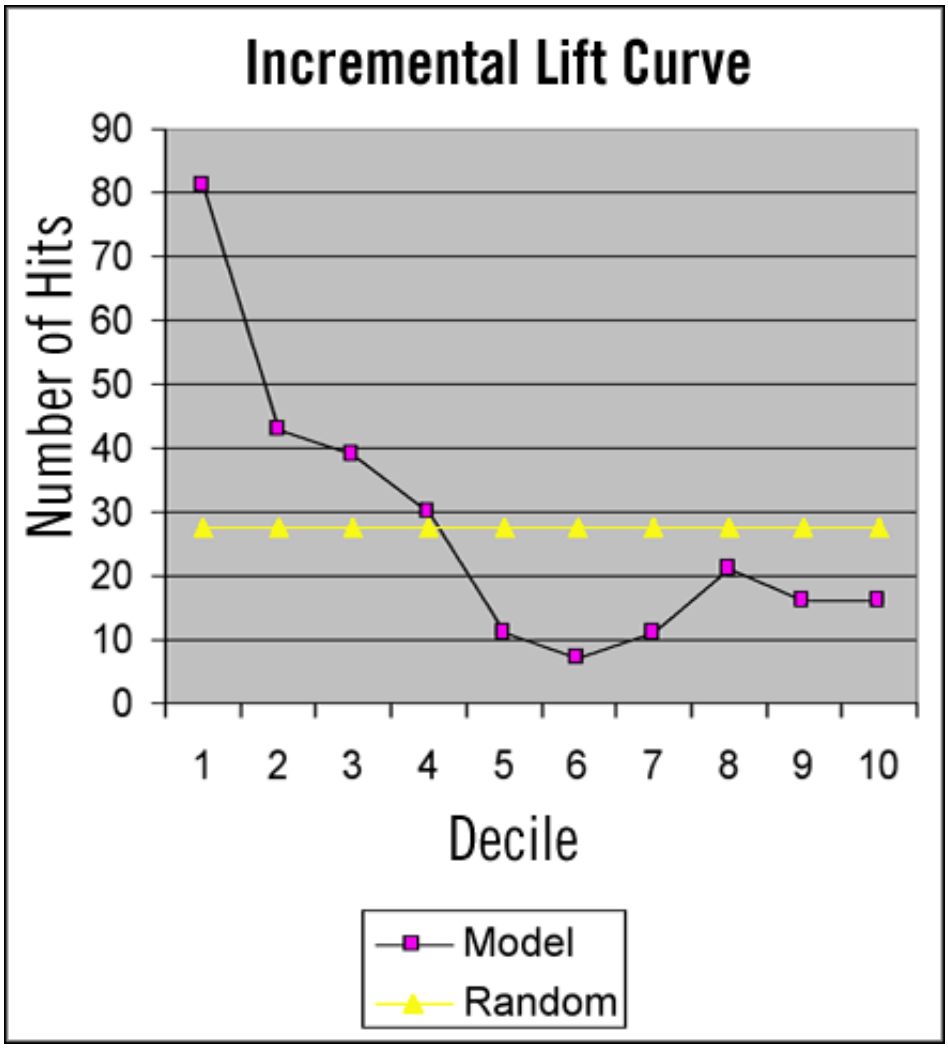

Figure 1: An Example of an Incremental Lift Chart

The incremental lift curve graphs the number of hits in each decile. In Figure 1, the curve crosses the random expectation line (10 percent of the total of 275 positives = 27.5 per decile) after the fourth decile. Presentation of the results in tabular and graphical forms will communicate the necessary information to market managers, no matter how they think. These model evaluation tools can be used by marketers to set the number of customers to mail to. Table 1 shows that a marketer could mail to the top four deciles (40 percent of the total scored list), and expect to hit over 70 percent of the potential responders.

Now that we have a clear understanding of how to evaluate DM models properly, we can look closer at the business processes that must be coordinated by data mining tools that can leverage model results to increase corporate profitability. These business processes include:

1. The data mining process,
2. The knowledge discovery process,
3. Business process management (BPM) programs,
4. Knowledge management systems, and
5. Business ecosystems management processes.

The Data Mining Process

Eric King maintains that the most important aspect of data mining is the journey, not the destination in "How to Buy Data Mining: A Framework for Avoiding Costly Project Pitfalls in Predictive Analytics" that appeared in *DM Review* in October 2005. He defines this journey

as the "process" of data mining. He describes the major elements of this process as:

1. A discovery process,
2. Has a flexible framework,
3. Proceeds from a clearly defined strategy,
4. Contains numerous checkpoints,
5. Includes periodic assessments,
6. Permits adjustments that function in feedback loops, and
7. Organized into an iterative architecture.

Process Models

Vendors of several data mining tool packages have simplified the process for the sake of clarity. SAS has collapsed the data mining process into the five stages: Sample, Explain, Manipulate, Model, Assess. One metaphor that has been used the past to describe the data mining process is a recirculating water fountain. Water (data) flows onto the first level (phase of analysis), forming eddies (refinements and feedbacks) until enough "processed" water accumulates to spill over to the next lower level. The "processing" continues until it reaches the lowest level, where it is pumped back to the top, and the "process" begins again. Data mining is a lot like this iterative cascading process. Even the internal processing of many data mining algorithms like neural nets is accomplished through many runs (epochs) through the data set until the "best" solution is found. (Insightful Miner) have built versions of a simple process model into their user interfaces. Such integration of the data mining process into the tool interface helps the user to organize the necessary data mining tasks in proper processing order.

The problem with the water fountain analogy is that there is no reflection of the feedback loops that often occur in the data mining process. For example, data assessment might uncover some anomalies that require extraction of additional data from source systems. Or, after modeling, it may become apparent that additional data records are needed to adequately represent the parent population.

One attempt to address this problem was embodied in the CRISP process model created by a consortium of Daimler-Benz, ISL (developer of Clementine) and NCR. The CRISP is an integral part of the Clementine tool design (now owned by SPSS). CRISP comes closest to encompassing the entire data mining context.

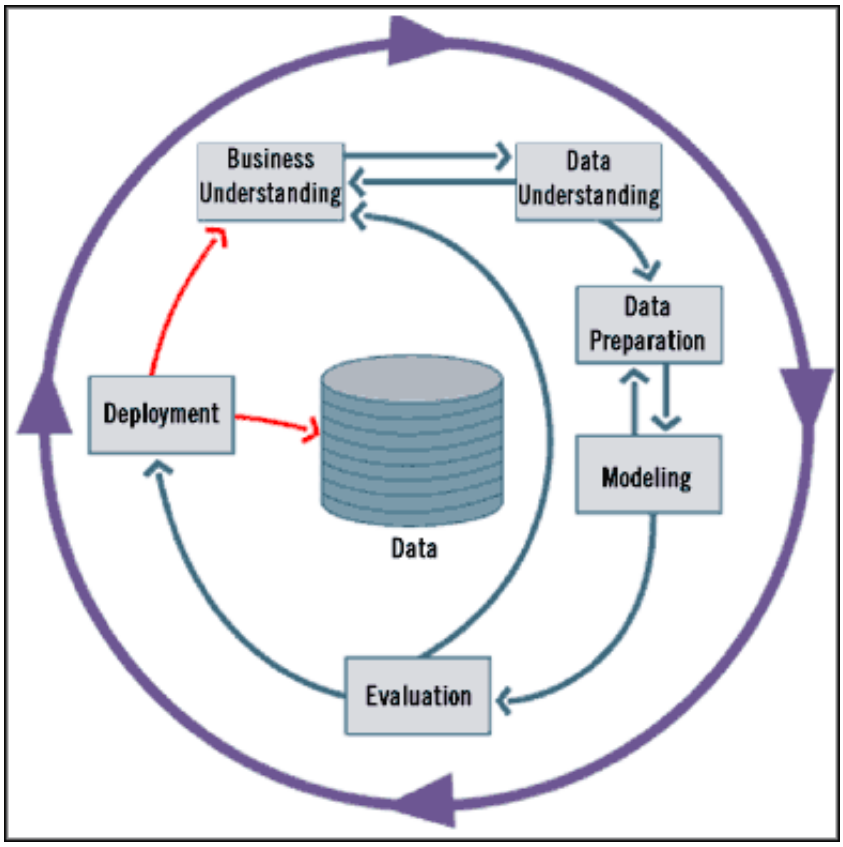

Figure 2: The CRISP Diagram

Modeling with data is much like modeling with clay or marble. The artist starts with a lump of material, and with many rounds (iterations) of manipulation and inspection, the art piece gradually reaches its final form. Modeling with data is complicated by the additional problem of not sufficiently knowing the nature of the modeling medium until midway through the modeling process. Eric King observes (rightly) that the data mining process is circular (like the CRISP process diagram above shows it to be), rather than being a linear process. The operation of the circular data mining process might remind you of the Wankel rotary automobile engine. The engine goes round and round (instead of up and down), pumping out kinetic energy in the form of rotory motion used to move the car. Likewise, the data mining process goes round and round and pumps out information that can be used to accomplish business goals. This information is the "energy" used to fuel business. There are many feedbacks to previous stages in the process (e.g., acquisition of additional data after preliminary modeling is done).

There is one element missing in the CRISP process, though - the element of feedback to the data warehouse or source data systems. Results from previous CRM campaigns should be entered into the data warehouse to provide insights for subsequent modeling operations and permit tracking of trends across campaigns. These feedbacks are superimposed on the CRISP process as dotted lines (Figure 2).
This structure of the data mining process gives us some of the necessary tasks that a data mining tool must do, but there are some that are missing. After the data mining results are available, what do you do with them? And, how do the actions that the data mining results spawn affect subsequent data mining activities? Among the other things that a data mining tool should facilitate are:

1. Model export to a number of database structures,
2. Model export in a format easy to import to other applications for decision support and business action,

3. Data feeds from one modeling algorithm to another (meta-modeling), and
4. Comparison of results between algorithms.

Part 2 will continue this discussion.

Robert A. Nisbet, Ph.D., is an independent data mining consultant with over 35 years experience in analysis and modeling in science and business. You can contact him at Bob@rnisbet.com or (805) 685-0053.

Robert A. Nisbet, Ph.D., is an independent data mining consultant with over 35 years experience in analysis and modeling in science and business. You can contact him at Bob@rnisbet.com or (805) 685-0053.

For more information on related topics, visit the following channels:

- Business Intelligence (BI)
- Customer Relationship Management (CRM)
- Data Mining