Building Profitable Customer Relationships With Data Mining

Herb Edelstein, President
Two Crows Corporation

You've built your customer information and marketing data warehouse – now how do you make good use of the data it contains?

Customer relationship management (CRM) helps companies improve the profitability of their interactions with customers while at the same time making the interactions appear friendlier through individualization. To succeed with CRM, companies need to match products and campaigns to prospects and customers — in other words, to intelligently manage the customer life cycle.

Until recently most CRM software has focused on simplifying the organization and management of customer information. Such software, called operational CRM, has focused on creating a customer database that presents a consistent picture of the customer's relationship with the company, and providing that information in specific applications such as sales force automation and customer service in which the company "touches" the customer.

However, the sheer volume of customer information and increasingly complex interactions with customers have propelled data mining to the forefront of making your customer relationships profitable. Data mining is a process that uses a variety of data analysis and modeling techniques to discover patterns and relationships in data that may be used to make accurate predictions. It can help you select the right prospects on whom to focus, offer the right additional products to your existing customers, and identify good customers who may be about to leave you. The result is improved revenue because of a greatly improved ability to respond to each individual contact in the best way, and reduced costs due to properly allocating your resources. CRM applications that use data mining are called analytic CRM.

This white paper will describe the various aspects of analytic CRM and show how it is used to manage the customer life cycle more cost-effectively. Note that the case histories of these fictional companies are composites of real-life data mining applications.

Data mining

The first and simplest analytical step in data mining is to **describe** the data — for example, summarize its statistical attributes (such as means and standard deviations), visually review it using charts and graphs, and look at the distribution of values of the fields in your data.

But data description alone cannot provide an action plan. You must **build a predictive model** based on patterns determined from known results, then **test** that model on results outside the original sample. A good model should never be confused with reality (you know a road map isn't a perfect representation of the actual road), but it can be a useful guide to understanding your business.

Data mining can be used for both classification and regression problems. In classification problems you're predicting what category something will fall into – for example, whether a person will be a good credit risk or not, or which of several offers someone is most likely to accept. In regression problems you're predicting a number such as the probability that a person

will respond to an offer.

In CRM, data mining is frequently used to assign a score to a particular customer or prospect indicating the likelihood that the individual will behave in the way you want. For example, a score could measure the propensity to respond to a particular offer or to switch to a competitor's product. It is also frequently used to identify a set of characteristics (called a profile) that segments customers into groups with similar behaviors, such as buying a particular product.

A special type of classification can recommend items based on similar interests held by groups of customers. This is sometimes called collaborative filtering.

The data mining technology used for solving classification, regression and collaborative filtering problems is briefly described in the appendix at the end of the paper.

Defining CRM

Customer relationship management in its broadest sense simply means managing all customer interactions. In practice, this requires using information about your customers and prospects to more effectively interact with your customers in all stages of your relationship with them. We refer to these stages as the customer life cycle.

The customer life cycle has three stages:

- Acquiring customers
- Increasing the value of the customer
- Retaining good customers

Data mining can improve your profitability in each of these stages through integration with operational CRM systems or as independent applications.

Acquiring new customers via data mining

The first step in CRM is to identify prospects and convert them to customers. Let's look at how data mining can help manage the costs and improve the effectiveness of a customer acquisition campaign.

Big Bank and Credit Card Company (BB&CC) annually conducts 25 direct mail campaigns each of which offers one million people the opportunity for a credit card. The conversion rate measures the proportion of people who become credit card customers, which for BB&CC is about 1% per campaign.

Getting people to fill out an application for the credit card is only the first step. Then BB&CC must decide whether the applicant is a good risk and accept them as a customer. Not surprisingly, poor credit risks are more likely to accept the offer than are good credit risks. So while 6% of the people on the mailing list respond with an application, only about 16% of those are suitable credit risks, for a net of about 1% of the mailing list becoming customers.

BB&CC's experience of a 6% response rate means that within the million names are 60,000 people who will respond to the solicitation. Unless BB&CC changes the nature of the solicitation – using different mailing lists, reaching customers in different ways, altering the terms of the

offer – they are not going to get more than 60,000 responses. And of those 60,000 responses, only 10,000 will be good enough risks to become customers. The challenge BB&CC faces is getting to those 10,000 people most efficiently.

The cost of mailing the solicitations is about $1.00 per piece for a total cost of $1,000,000. Over the next couple of years, these customers will generate about $1,250,000 in profit for the bank (or about $125 each) for a net return from the mailing of $250,000.

Data mining can improve this return. Although it won't precisely identify the 10,0000 eventual credit card customers, it will help focus marketing efforts much more cost-effectively.

First BB&CC did a test mailing of 50,000 and carefully analyzed the results, building a predictive model of who would respond (using a decision tree) and a credit scoring model (using a neural net). It then combined these two models to find the people who were both good credit risks and most likely to respond to the offer.

The model was applied to the remaining 950,000 people in the mailing list from which 700,000 people were selected for the mailing. The result was that from the 750,000 pieces mailed overall (including the test mailing), 9,000 acceptable applications for credit cards were received. In other words, the response rate had risen from 1% to 1.2%, a 20% increase. While the targeted mailing only reaches 9,000 of the 10,000 prospects – no model is perfect – reaching the remaining 1,000 prospects is not profitable. Had they mailed the other 250,000 people on the mailing list, the cost of $250,000 would have resulted in another $125,000 of gross profit for a net loss of $125,000. The following table summarizes the results.

	Old	**New**	**Difference**
Number of pieces mailed	1,000,000	750,000	(250,000)
Cost of mailing	$1,000,000	$750,000	($250,000)
Number of responses	10,000	9,000	(1,000)
Gross profit per response	$125	$125	$0
Gross profit	$1,250,000	$1,125,000	($125,000)
Net profit	$250,000	$375,000	$125,000
Cost of model	0	40,000	$40,000
Final profit	$250.000	$335,000	$85,000

Notice that the net profit from the mailing increased $125,000. Even when you include the $40,000 cost of the data mining software, computer, and people resources used for this modeling effort the net profit increased $85,000. This translates to a return on investment for modeling of over 200% which far exceeded BB&CC's ROI requirements for a project.

Increasing the value of your existing customers: cross-selling via data mining

Guns and Roses (G&R) is a company that specializes in selling antique mortars and cannons as outdoor flower pots. They also offer a line of indoor flower pots made from large caliber antique pistols and a collection of muskets that have been converted to unique holders of long stemmed flowers. Their catalog is sent to about 12 million homes.

When a customer calls in to place an order, G&R identifies the caller using caller ID when possible; otherwise they ask for a phone number or customer number from the catalog mailing label. Next, they look up the customer in the database and then proceed to take the order.

G&R has an excellent chance of selling the caller something additional – cross-selling. But G&R had found that if the first suggestion fails and they try to suggest a second item, the customer may get irritated and hang up without ordering anything. And there are some customers who resent any attempts at all to cross-sell them on additional products.

Before trying data mining, G&R had been reluctant to cross-sell at all. Without the model, the odds of making the right recommendation were one in three. And because making any recommendation is for some customers unacceptable, they wanted to be exceptionally sure that they never made a recommendation when they should not. In a trial campaign, they had less than a 1% sales rate and had a substantial number of complaints. They were reluctant to continue for such a small gain.

The situation changed dramatically with the use of data mining. Now the data mining model operates on the data. Using the customer information in the database and the new order, it tells the customer service representative what to recommend. They successfully sold 2% of the customers an additional product with virtually no complaints.

Developing this capability involved a process similar to solving the credit card customer acquisition problem. As with that situation, two models were needed.

The first model predicted whether someone would be offended by recommendations. G&R found out how their customers would react by conducting a very short telephone survey. To be conservative, they counted anyone who declined to participate in the survey as someone who would find recommendations intrusive. Later on, to verify this assumption, they made recommendations to a small but statistically significant subset of those who had refused to answer the survey questions. To their surprise, they found that the assumption was not warranted. This enabled them to make more recommendations and further increase profits. The second model predicted which offer would be most acceptable.

In summary, data mining helped G&R better understand their customers' needs. When the data mining models were incorporated in a typical cross-selling CRM campaign, the models helped G&R increase its profitability 2%.

Increasing the value of your existing customers: personalization via data mining

Big Sam's Clothing (motto: "Rugged outdoor gear for city dwellers") has set up a website to supplement their catalog. Whenever you go to their site they greet you with "Howdy Pardner!" but once you have ordered or registered with them they greet you by name. If you have a record of ordering from them, they will also tell you about any new products that might be of particular interest to you. When you look at a particular product, such as a waterproof down parka, they will suggest other things that might supplement such a purchase.

When they first put up the site, there was none of this personalization. It was just an on-line version of their catalog – nicely and efficiently done but not taking advantage of the sales opportunities presented by the Web.

Data mining greatly increased the sales at their website. Catalogs frequently group products by type to simplify the user's task of selecting products. In an on-line store, however, the product groups may be quite different, often based on complementing the item under consideration. In particular, the site can take into account not only the item you're looking at, but what is in your shopping cart as well, thus leading to even more customized recommendations.

First, Big Sam's used clustering to discover which products grouped together naturally. Some of the clusters were obvious, such as shirts and pants. Others were surprising, such as books about desert hiking and snakebite kits. They used these groupings to make recommendations whenever someone looked at a product.

They then built a customer profile to help them identify those customers who would be interested in the new products they were always adding to their catalog. They found that steering people to these selected products not only resulted in significant incremental sales, but also solidified their relationship with the customer. Surveys established that they were viewed as a trusted advisor for clothing and gear.

To extend their reach further, Big Sam's started a program through which customers could elect to receive e-mail about new products that the data mining models predicted would interest them. While the customers viewed this as another example of proactive customer service, Big Sam's found it to be a program of profit improvement.

The effort in personalization paid off for Big Sam's with significant, measurable increases in repeat sales, average number of sales per customer, and average size of a sale.

Retaining good customers via data mining

For almost every company, the cost of acquiring a new customer exceeds the cost of keeping good customers. This was the challenge facing Know Service, an Internet Service Provider (ISP) whose attrition rate was the industry average, 8% per month. Since they have one million customers, this means 80,000 customers left each month. The cost to replace these customers is $200 each, or $16,000,000 – plenty of incentive to start an attrition management program.

The first thing Know Service needed to do was prepare the data for predicting which customers would leave. They needed to select the variables from their customer database and perhaps transform them. The bulk of their users were dial-in clients (as opposed to clients who are always connected through a T1 or DSL line), so they knew how long each user was connected to the Web. They also knew the volume of data transferred to and from a user's computer, the number of e-mail accounts a user had, the number of e-mail messages sent and received, and a customer's service and billing history. In addition, they had demographic data that customers provided at sign-up.

Next they needed to identify who were "good" customers. This is not a data mining question but a business definition (such as profitability or lifetime value) followed by a calculation. Know Service built a model to profile their profitable customers and their unprofitable customers. They used this model not only for customer retention but to identify customers who were not yet profitable but might become so in the future.

Know Service then built a model to predict who among their profitable customers would leave. As in most data mining problems, determining what data to use and how to combine existing data is where much of the challenge lies in model development. For example, they needed to look at time-series data such as the monthly usage. Rather than use the raw time-series data, they smoothed it by taking rolling three-month averages. They also calculated the change in the three-month average and tried that as a predictor. Some of the factors that were good predictors, such as declining usage, were symptoms rather than causes that could be directly addressed. Other predictors, such as the average number of service calls and the change in the average number of service calls, were indicative of customer satisfaction problems worth investigating.

Predicting who would churn, however, wasn't enough. Based on the results of their modeling, they identified some potential programs and offers that they believed would entice people to stay. For example, some churners were exceeding even the largest amount of usage available for a fixed fee and were paying substantial incremental usage fees. They tried offering these users a higher fee service that included more bundled time. Some users were offered as more free disk space for personal web pages. They then built models that would predict which would be the effective offer for a particular user.

To summarize, the churn project made use of all three models. One model identified likely churners, the next model picked out the profitable ones worth keeping, and the third model matched the potential churners with the most appropriate offer. The net result was a reduction in their churn rate from 8% to 7.5%, for a savings in customer acquisition costs of $1,000,000 per month.

Know Service found that their investment in data mining paid off by improving their customer relationships and dramatically increasing their profitability.

Applying Data Mining to CRM

In order to build good models for your CRM system, there are a number of steps you must follow. The Two Crows data mining process model described below is similar to other process models such as the CRISP-DM model, differing mostly in the emphasis it places on the different steps.

Keep in mind that while the steps appear in a list, the data mining process is not linear — you will inevitably need to loop back to previous steps. For example, what you learn in the "explore data" step may require you to add new data to the data mining database. The initial models you build may provide insights that lead you to create new variables.

The basic steps of data mining for effective CRM are:

1. Define business problem
2. Build marketing database
3. Explore data
4. Prepare data for modeling
5. Build model
6. Evaluate model
7. Deploy model and results

Let's go through these steps to better understand the process.

1. **Define the business problem.** Each CRM application will have one or more business objectives for which you will need to build the appropriate model. Depending on your specific goal, such as "increasing the response rate" or "increasing the value of a response," you will build a very different model. An effective statement of the problem will include a way of measuring the results of your CRM project.

2. **Build a marketing database.** Steps two through four constitute the core of the data preparation. Together, they take more time and effort than all the other steps combined. There may be repeated iterations of the data preparation and model building steps as you learn something from the model that suggests you modify the data. *These data preparation steps may take anywhere from 50% to 90% of the time and effort of the entire data mining process!*

   You will need to build a marketing database because your operational databases and corporate data warehouse will often not contain the data you need in the form you need it. Furthermore, your CRM applications may interfere with the speedy and effective execution of these systems.

   When you build your marketing database you will need to clean it up – if you want good models you need to have clean data. The data you need may reside in multiple databases such as the customer database, product database, and transaction databases. This means you will need to integrate and consolidate the data into a single marketing database and reconcile differences in data values from the various sources. Improperly reconciled data is a major source of quality problems. There are often large differences in the way data is defined and used in different databases. Some inconsistencies may be easy to uncover, such as different addresses for the same customer. Making it more difficult to resolve these problems is that they are often subtle. For example, the same customer may have different names or — worse — multiple customer identification numbers.

3. **Explore the data.** Before you can build good predictive models, you must understand your data. Start by gathering a variety of numerical summaries (including descriptive statistics such as averages, standard deviations and so forth) and looking at the distribution of the data. You may want to produce cross tabulations (pivot tables) for multi-dimensional data.

   Graphing and visualization tools are a vital aid in data preparation, and their importance to effective data analysis cannot be overemphasized. Data visualization most often provides the *Aha!* leading to new insights and success. Some of the common and very useful graphical displays of data are histograms or box plots that display distributions of values. You may also want to look at scatter plots in two or three dimensions of different pairs of variables. The ability to add a third, overlay variable greatly increases the usefulness of some types of graphs.

4. **Prepare data for modeling.** This is the final data preparation step before building models and the step where the most "art" comes in. There are four main parts to this step:

   First you want to select the variables on which to build the model. Ideally, you would take all the variables you have, feed them to the data mining tool and let it find those which are the best predictors. In practice, this doesn't work very well. One reason is that the time it takes to build a model increases with the number of variables. Another reason is that blindly including extraneous columns can lead to models with less rather than more predictive power.

The next step is to construct new predictors derived from the raw data. For example, forecasting credit risk using a debt-to-income *ratio* rather than just debt and income as predictor variables may yield more accurate results that are also easier to understand.

Next you may decide to select a subset or sample of your data on which to build models. If you have a lot of data, however, using all your data may take too long or require buying a bigger computer than you would like. Working with a properly selected random sample usually results in no loss of information for most CRM problems. Given a choice of either investigating a few models built on all the data or investigating more models built on a sample, the latter approach will usually help you develop a more accurate and robust model. of the problem.

Last, you will need to transform variables in accordance with the requirements of the algorithm you choose for building your model.

5. **Data mining model building.** The most important thing to remember about model building is that it is an iterative process. You will need to explore alternative models to find the one that is most useful in solving your business problem. What you learn in searching for a good model may lead you to go back and make some changes to the data you are using or even modify your problem statement.

   Most CRM applications are based on a protocol called supervised learning. You start with customer information for which the desired outcome is already known. For example, you may have historical data because you previously mailed to a list very similar to the one you are using. Or you may have to conduct a test mailing to determine how people will respond to an offer. You then split this data into two groups. On the first group you train or estimate your model. You then test it on the remainder of the data. A model is built when the cycle of training and testing is completed.
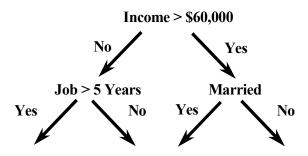
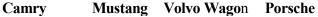6. **Evaluate your results**

   Perhaps the most overrated metric for evaluating your results is accuracy. Suppose you have an offer to which only 1% of the people will respond. A model that predicts "nobody will respond" is 99% accurate and 100% useless. Another measure that is frequently used is lift. Lift measures the improvement achieved by a predictive model. However, lift does not take into account cost and revenue, so it is often preferable to look at profit or ROI. Depending on whether you choose to maximize lift, profit, or ROI, you will choose a different percentage of your mailing list to whom you will send solicitations.

7. **Incorporating data mining in your CRM solution**

   In building a CRM application, data mining is often only a small, albeit critical, part of the final product. For example, predictive patterns through data mining may be combined with the knowledge of domain experts and incorporated in a large application used by many different kinds of people.

   The way data mining is actually built into the application is determined by the nature of the customer interaction. There are two main ways you interact with your customers: they contact you (inbound) or you contact them (outbound). The deployment requirements are quite different.

Income > $60,000

No / Yes

Job > 5 Years   **Married**

Yes / No   Yes / No

Camry   **Mustang**   **Volvo Wago**n   **Porsche**

Outbound interactions are characterized by your company originating the contact such as in a direct mail campaign. Thus you will be selecting the people to whom you mail by applying the model to your customer database. Another type of outbound campaign is an advertising campaign. In this case you would match the profiles of good prospects shown by your model to the profile of the people your advertisement would reach.

For inbound transactions, such as a telephone order, an Internet order, or a customer service call, the application must respond in real time. Therefore the data mining model is embedded in the application and actively recommends an action.

In either case, one of the key issues you must deal with in applying a model to new data is the transformations you used in building the model. Thus if the input data (whether from a transaction or a database) contains age, income, and gender fields, but the model requires the age-to-income ratio and gender has been changed into two binary variables, you must transform your input data accordingly. The ease with which you can embed these transformations becomes one of the most important productivity factors when you want to rapidly deploy many models.

Conclusion

Customer relationship management is essential to compete effectively in today's marketplace. The more effectively you can use the information about your customers to meet their needs the more profitable you will be. But operational CRM needs analytical CRM with predictive data mining models at its core. The route to a successful business requires that you understand your customers and their requirements, and data mining is the essential guide.

Appendix: Data mining technology

Decision trees

Decision trees are a way of representing a series of rules that lead to a class or value. For example, you may wish to offer a prospective customer a particular product. The figure shows a simple decision tree that solves this problem while illustrating all the basic components of a decision tree: the decision node, branches and leaves.

A simple classification tree.

The first component is the top decision node, or root node, which specifies a test to be carried out.
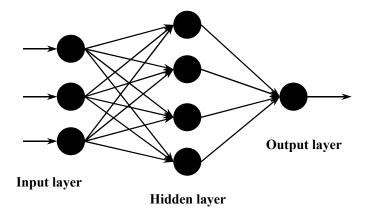
Each branch will lead either to another decision node or to the bottom of the tree, called a leaf node. By navigating the decision tree you can assign a value or class to a case by deciding which branch to take, starting at the root node and moving to each subsequent node until a leaf node is reached. Each node uses the data from the case to choose the appropriate branch.

Decision trees models are commonly used in data mining to examine the data and induce a tree and its rules that will be used to make predictions. A number of different algorithms may be used for building decision trees including CHAID (Chi-squared Automatic Interaction Detection), CART (Classification And Regression Trees), Quest, and C5.0.

Neural networks

Neural networks are of particular interest because they offer a means of efficiently modeling large and complex problems in which there may be hundreds of predictor variables that have many interactions. (Actual biological neural networks are incomparably more complex.) Neural nets are most commonly used for regressions but may also be used in classification problems.

A neural network (see figure) starts with an *input layer*, where each node corresponds to a predictor variable. These input nodes are connected to a number of nodes in a *hidden layer*. Each input node is connected to every node in the hidden layer. The nodes in the hidden layer may be connected to nodes in another hidden layer, or to an *output layer*. The output layer consists of one or more response variables.
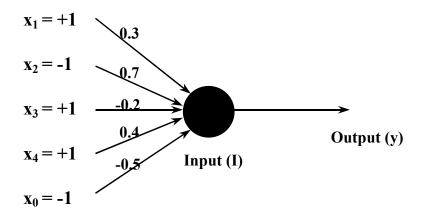
A neural network with one hidden layer.

 After the input layer, each node takes in a set of inputs, multiplies them by a connection weight

adds them together, applies a function (called the activation or squashing function) to them, and passes the output to the node(s) in the next layer. For example, the node above has five inputs (x_0 through x_4) each of which is multiplied by a weight and then added together resulting in a sum I:

$I = .3X_1 + .7X_2 - .2X_3 + .4X_4 - .5X_5 = .3 - .7 - .2 + .4 + .5 = .3$

This output y is then the sum that has been transformed by the non-linear activation function, in this case to a value of .57.

The goal of training the neural net is to estimate the connection weights so that the output of the neural net accurately predicts the test value for a given input set of values. The most common training method is backpropagation. Each training method has a set of parameters that control various aspects of training such as avoiding local optima or adjusting the speed of conversion.

Neural networks differ in philosophy from many statistical methods in several ways. First, a neural network usually has more parameters than does a typical statistical model. For example, a neural network with 100 inputs and 50 hidden nodes will have over 5,000 parameters. Because they are so numerous, and because so many combinations of parameters result in similar predictions, the parameters become uninterpretable and the network serves as a "black box" predictor. However, this is acceptable in CRM applications. A bank may assign the probability of bankruptcy to an account and may not care what are the causes.

One myth of neural networks is that data of any quality can be used to provide reasonable predictions and they will sift through it to find the truth. However, neural networks require as much data preparation as any other method, which is to say they require a lot of data preparation. The most successful implementations of neural networks (or decision trees, or logistic regression, or any other method) involve very careful data cleansing, selection, preparation and pre-processing. For instance, neural nets require that all variables be numeric. Therefore categorical data such as "state" is usually broken up into multiple dichotomous variables (e.g., "California," "New York") , each with a "1" (yes) or "0" (no) value. The resulting increase in variables is called the *categorical explosion.*

Clustering

Clustering divides a database into different groups. The goal of clustering is to find groups that are very different from each other, and whose members are very similar to each other. Unlike classification, you don't know what the clusters will be when you start, or by which attributes the data will be clustered. Consequently, someone who is knowledgeable in the business must interpret the clusters. After you have found clusters that reasonably segment your database, these clusters may then be used to classify new data. Some of the common algorithms used to perform clustering include Kohonen feature maps and K-means.

Don't confuse clustering with segmentation. Segmentation refers to the general problem of identifying groups that have common characteristics. Clustering is a way to segment data into groups that are not previously defined, whereas classification is a way to segment data by assigning it to groups that are already defined.