

---

# Papers

## Lessons learned: A case study using data mining in the newspaper industry

Received (in revised form): 19th July, 2007

### Candace L. Gunnarsson

is owner-operator of S2 Statistical Solutions, Inc. founded in 1992, and performs statistics consulting and training seminars. Dr Gunnarsson has taught statistics courses for more than ten years in the fields of education, psychology, and business at Xavier University, The Union Institute, and University of Cincinnati.

### Mary M. Walker

is Professor of Marketing at Xavier University in Cincinnati, Ohio, and in this role was the recipient of a grant sponsored by the General Electric Company to develop a course on data mining. She has served as Xavier's Acting Vice President for Information Resources, overseeing all information technology during 2004–2005.

### Vern Walatka

is a chemical engineer and chemist with experience as a SAS programmer and data analyst. He currently consults on topics of survey analysis, GIS, and mapping software.

### Kenneth Swann

is Senior Data Analytical Manager with a background in software management and market research. He has organised three successful annual SAS one-day conferences for statisticians and researchers.

**Keywords** *data mining, data-driven decision making, customer relationship management (CRM), decision tree application*

**Abstract** Many organisations across a variety of industries are engaging in the process of data mining as part of an overall strategy for business intelligence, customer relationship management (CRM), including churn prevention. This paper provides an overview of the data mining process and illustrates a case study in which data mining is utilised as a churn prevention tool for a major Midwest USA newspaper. For this case study, a decision tree, a common modelling technique, was the analytical tool of choice. Lessons learned throughout the data mining process are provided to offer insight and to promote the sharing of information. Strategies for getting started in the data mining process are presented to encourage organisations to embrace a data-driven strategy for business intelligence, CRM and churn prevention.

*Journal of Database Marketing & Customer Strategy Management* (2007) **14**, 271–280.

doi:10.1057/palgrave.dbm.3250058

### INTRODUCTION

Many organisations across a variety of industries are engaging in the process of data mining as part of an overall strategy to improve business intelligence, customer relationship management (CRM) and churn

prevention. Data mining is the computer-assisted process of analysing data from different perspectives and distilling it into actionable information. More than ever, companies are using this technique as they recognise their raw transactional data as a

**Mary M. Walker**  
Williams College of Business,  
Xavier University,  
3800 Victory Parkway,  
Cincinnati, OH 45207, US  
Tel: +1 513 745 2980;  
e-mail: walkerm@xavier.edu

valuable source of unique information, which, once analysed via data mining, can create knowledge that advances organisational objectives.

The newspaper industry is one of many sectors that could benefit from the data mining process. Faced with declining circulation and advertisers who are considering other direct channels to reach their customers, editors and executives need to acknowledge these trends and adopt advertising solutions meant to maintain and extend circulation. Due to the downturn in revenues, however, investments in advancing the data warehouse that fuels data mining initiatives have been limited. To date, data mining remains a concept that has yet to serve the research needs of publishers because, despite its potential, data mining is viewed as too expensive and unproven.<sup>1</sup>

This is due to several factors. First, data mining is a relatively new phenomenon that pulls from a variety of different disciplines, thus making it difficult to adopt strong standards and industry-wide operating procedures. Secondly, the integration of multidisciplinary sciences is further complicated by increasing privacy concerns. Companies often have to design customised techniques to meet these needs instead of following standardised operating procedures (SOPs) set forth by their particular industry. Technology for data mining designed in this way is limited by the functional silos within the organisation that contribute to the process.

One way that industries can begin to establish SOPs is through information exchange. Companies within each industry need to share their stories as they foray into the process of data mining. Sharing particular applications and lessons learned during the establishment of the data mining process will aid in reducing this silo effect, allowing companies to build on the experiences of one another.

This paper presents a case study of the data mining process as a churn prevention tool for a major Midwest USA newspaper.

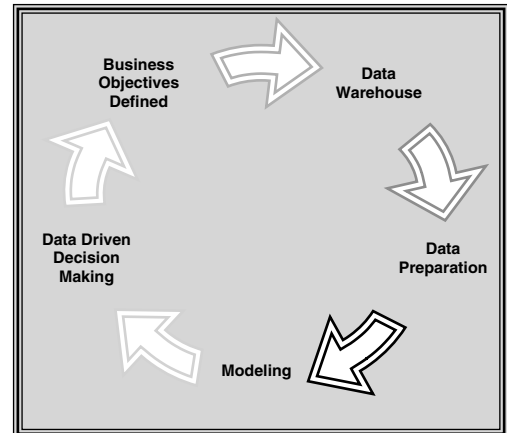


Figure 1: Data mining process

The goal of this paper is to promote industry exchange regarding the lessons learned during the data mining process. First, the important steps in the data mining process are explained in detail. Then each step is applied to the case study. Within each section, particular attention is paid to the lessons learned along the way.

## BACKGROUND

### Data mining is a process of knowledge discovery

Figure 1 illustrates the process flow chart developed for the data mining process. Our process is similar to a well-established data mining methodology, which is widely in use in Europe, called CRISP-DM.<sup>2</sup> This section discusses each of the activities depicted in the chart.

#### *Business objectives defined*

The first step in the data mining process is to clearly define initial business objectives. This is a multidimensional endeavour involving all business team members both directly and indirectly connected to the data mining initiative. Typically, many different domain experts are involved in a data mining process such as salespeople, marketers, statisticians, information systems personnel and computer programmers. Often, it is the marketing department that

coordinates the inputs that establish business objectives. It is critical that all domain experts be part of the data mining activity so that the data mining effort put forth is better able to deliver the defined objectives.

The business objectives best suited to the data mining process are those involving prediction or seeking an explanation of behaviour. For example, CRM objectives include determining best customer profiles and predicting customers who are most at risk to churn. Once initial objectives have been defined and a decision to move forward with the data mining process is made, the knowledge acquired from the process will lead to more business objectives. Companies should not embark on data mining 'projects'. Data mining is an iterative process that is employed as part of an overall strategy for business intelligence and problem solving by employing statistical modelling to large amounts of transformed, transactional and historical data.

Once the business objectives have been delineated, the next step is to determine what data are needed to address the objectives and how the data will be obtained.

#### *The data warehouse and data preparation*

Steps two and three in the data mining process typically go hand in hand, and are crucial elements of the process. While newer innovations create opportunities to make this process more efficient, data warehousing and data preparation have historically represented 85 per cent of the data mining effort,<sup>3</sup> and continue to be a mainstay of successful data mining initiatives.

All data warehouse sources, both internal and external, need to be considered for data mining. Most companies have many different and incompatible internal data sources available, making warehousing inefficient. Businesses collect enormous amounts of transactional and historical data. The task of deciding the appropriate data fields that need to be extracted from these

various data sources, as well as how the data should be organised for modelling, can be arduous.

The data preparation step is generally a large and challenging undertaking. This is because the data utilised in the data mining process is typically massive due to it being transactional in nature. It is typically not collected with analysis in mind. Instead, the data are collected to monitor process control as the result of transactions that took place within the organisation or between an organisation and a constituent.

#### *Modelling*

Once the data are prepared, three modelling techniques are typically used: decision trees, regression analysis and neural networks. These techniques are neither new modelling procedures nor are they exclusive techniques to the data mining process.

A distinct difference exists between data mining versus statistical modelling. Data mining focusses on computer-generated models; whereas traditional statistical modelling stresses theory-driven hypothesis testing<sup>4</sup>: Hypotheses need to be generated *a priori* and assumptions such as linearity need to be tested. By comparison, data mining does not impose the assumptions and limitations of traditional theoretical modelling.

Typically the data mining process uses statistical modelling and methods, and therefore requires target data. In other words, the modelling process attempts to explain or predict a target variable. A target variable is a value that is either known or created from other values in some currently available data, but will be unknown in some future or fresh data. It is the dependent variable when applying the three statistical modelling techniques. A target variable can be dichotomous, nominal or interval. In order to predict behaviour one must be able to assign either a probability of response or a range of response based on a list of attributes associated with the target variable.

SAS Enterprise Miner 5.2. uses the three main modelling techniques: Decision Trees, Regression and Neural Networks. Decision Trees are fitted using recursive partitioning by segmenting the data into sub-groups that are as homogeneous as possible with respect to the target. The method is recursive because each sub-group results from splitting a sub-group from a previous split. The number of possible splits is usually enormous. Decision trees use algorithms to decide splitting rules. Regression analysis is utilised when the target is an interval variable. Here the model predicts the mean of the target variable at the given values of the input variable. When the target variable is dichotomous, logistic regression can be used and the model predicts the probability of the target variable at the given values of the input variables. Regression analyses require a plan for imputation of missing data. Neural networks are a form of artificial intelligence whereby a class of flexible nonlinear models are utilised for prediction.

*Data-driven decision making*

Once the results of the modelling techniques have been communicated to all internal company stakeholders, data-driven decisions on marketing and customer retention programmes can be developed and implemented. In this step of the process, information obtained from the data is put into action. The result of these actions will then need to be tested and validated to measure return on investment. This closes the loop on the process and leads to analysis of the next objective. Customers will change and models will continually need to be fine tuned and updated. New questions and goals will beckon, making data mining a truly iterative data-driven process.

The following case study illustrates the process just outlined for a newspaper company in the Midwest USA with a circulation of about 200,000, covering several municipalities. After each section in the process, the lessons learned are shared.

**CASE STUDY**

**Step One: Defining the business objective**

When first meeting with the newspaper executives and their marketing and database departments, the following business objectives were identified:

- 1 Addressing Churn:
  - a. What is the best time to communicate with former customers after the subscription stops?
  - b. At what point in the prospecting cycle should a former customer be treated like they have never subscribed?
  - c. How often should former subscribers be solicited?
  
- 2 Upgrades and cross-selling opportunities:
  - a. Identify which current customers who are billed periodically for their subscription renewal would switch to a perpetual payment method.
  - b. Identify which customers are best to target for upgrade programmes. Upgrades are business decision campaigns that offer customers of lower frequencies of delivery (FOD) the opportunity to upgrade to the highest frequency for a limited period of time (six or 12 months) at no additional cost.
  - c. Are there seasonal trends that would enhance the upgrade selling timeline?
  - d. For upgrades, are there pricing schemes or price points that would maximise response rate?
  
- 3 Developing new prospects:
  - a. Develop a segmented targeting model to define and/or predict best prospects and contact frequency.
  - b. What do those who have never subscribed look like?
  - c. Are there differences in behaviour between Former and Never subscribers?

- d. For both former subscribers and those who have never subscribed: Are there risk/reward timings between possible sub-segments of potential households, based on demographics or former subscription history?

All of the above questions were tied to the company’s business objectives and were well-suited questions for the data mining process. After careful examination of the data warehouse, however, the data available did not support all objectives and research questions at hand.

*Lesson 1:* Executives have a broad array of business objectives and research questions; but without the appropriate data, the data mining process cannot support knowledge discovery, and a substantial disconnect will exist between upper management and the database marketers they employ.

Once these disconnects are revealed, management must be willing to invest in the data gathering and warehousing resources necessary to meet its business objectives. Companies need to view their data as a corporate asset. Just as data mining is an iterative process, so too is data warehousing. If the data capture is not sufficient for the data mining process,

organisations must commit to putting the scaffolding in place that will support the necessary data infrastructure. Constructing the data infrastructure is typically time consuming and costly, but if done correctly, is well worth the investment.

*Lesson 2:* Identifying and aligning business objectives with the data available to answer these objectives is a crucial step in the data mining process. Domain experts must be willing to commit the necessary time and resources to this task.

Based on this experience, we recognised the scope of this task and developed a three-step process (see Figure 2: Domain Expert Questionnaire) to be utilised with all domain experts to aid in defining goals and business objectives. Domain experts are individuals who will be participating in the data mining initiative and who have at least some degree of decision-making authority.

**Step Two: The data warehouse**

At the newspaper, both internal and external data were stored in different functional databases specific to particular business processes. To maintain customer privacy, we were given a random sample of 10 per cent of the households in the master

<p><b>Step 1: Write down the answers to the following questions:</b></p>
<ol style="list-style-type: none"> <li>1. Which person(s) at your company are responsible for initiating this data mining effort?</li> <li>2. List the names, positions and domain expertise of all the people in your organization who are responsible for initiating this effort.</li> <li>3. List the names, positions and domain expertise of all the people in your organization who will be working on and ultimately responsible for this effort. Note: Keep a tally of the people who overlap, those that are both responsible for initiation, data preparation and analysis, and accountable for the outcomes.</li> </ol>
<p><b>Step 2: Give the following questionnaire to all domain experts involved in both the initiation and organization of the data mining process.</b></p>
<ol style="list-style-type: none"> <li>1. Do you have a particular question you are looking to answer? If so, write it down.</li> <li>2. Do you have a particular problem you are trying to solve? If so, write it down.</li> <li>3. Do you have a business goal? If so, write it down.</li> <li>4. Is this effort part of a bigger organizational policy or practice?</li> <li>5. Are you trying to predict something?</li> <li>6. Are you trying to understand or explain some behavior or phenomenon?</li> </ol>
<p><b>Step 3: Analyze the results from the domain experts. Are expectations the same across the board? Write the business objectives that encompass the statements of the domain experts you surveyed. If the goals and objectives are different, work these out ahead of time. All experts involved in the process should agree on the business objectives.</b></p>

Figure 2: Domain expert questionnaire

database after all identifying information (name, street address, telephone number, etc) was removed from the files. The zip code information was masked and therefore was not useful. Each subscriber was then assigned a unique 'Address Sequence Number'. The newspaper supplied us with three comma delimited data files named Household, Promotion and Transactions.

As the name implies, the Household file (95,000 records) contains one record for each unique address — house, apartment or PO Box — in the database. Each record describes a current subscriber, former subscriber or someone who has never subscribed and contains demographic information about each address. For current and former subscribers, the Household file also contains basic information such as when the subscription to the paper started and stopped, how often the paper is delivered, and other basic subscription information. The Promotion file (135,000 records) lists the date and type of each promotional mailing sent to a household.

Finally, the Transaction file (1,700,000 records) records all interactions with each customer or potential customer. These interactions include all start/stops for vacation and other reasons, rate code changes, promo changes, payment type, payment history and many miscellaneous comments such as 'Please give the carrier a \$5 tip and charge it to my credit card' or 'Do not toss the paper on the lawn', etc. The Transaction file presented difficulties because much of the useful information is contained in a free-form text comment field. Since different operators use different wordings for basically the same information, the field is very difficult to use for analysis.

*Lesson 3:* Regarding free form data, standardising and categorising responses in a consistent manner makes the data easier to sort and analyse.

Unfortunately, we were not able to use the information in the Promotion file since that file only told us that a promotion was

mailed to a particular household. We were not able to consistently determine from the Transaction file whether or not the promotion was successful. For this reason, we could not address some of the original business objectives set forth by the domain experts.

*Lesson 4:* Process-centric methods of data storage can result in silos of unrelated data across the organisation that makes it impossible to link together and track customer behaviour.

### Step Three: Data preparation

Data acquisition and extraction tools can help move data from the various systems into one source for analysis. SAS version 9.1<sup>5</sup> was utilised for all data preparation and SAS Enterprise Miner 5.2<sup>6</sup> was utilised for all analyses. Data preparation involves data cleansing and formatting, a complicated process that validates and corrects the data prior to analysis. More often than not, attributes of one database are different from those of another. For this particular data set, we encountered the following problems: (1) duplicate records within files, (2) missing data values, (3) data values used as comment data and (4) mail dates and promotion codes not matching. These problems were determined by de-duping tables and frequency distributions.

Given the limitations of the data, all three files were matched by household, and using SAS version 9.1, a flat data file was created. For data mining, SAS Enterprise Miner requires a flat data file; that is, a data file that contains only one record for each topic of interest. In this case, the topic of interest is the Household or Customer. The Household file already was flat, but the Promotion and Transaction files contained multiple records for each household. As a result, both the Promotion and Transaction files needed to be aggregated up to the Household level. This was accomplished by taking the data that were contained in multiple records with the same Address\_Sequence\_Number and

combining it into a single record. During this process, important data were retained and transformed in the form of newly defined variables.

The three files now contained only one record per household and were merged into one large file by matching records with the same Address\_Sequence\_Number. For the final flat data file, only those records that appeared in all three files were retained, guaranteeing that all households in the final data file contained valid transactions and received at least one promotion during the 30-month period of the data.

*Lesson 5:* During data preparation, many decisions will be made concerning which data to include. It is important to remember that data mining is a process. Therefore, some of these initial decisions regarding data aggregation, inclusion and exclusion have the potential to change based on the learning from the actual models.

#### Step Four: Modelling

After data preparation, the next step was to recode and derive new variables for modelling. New variables were computed, some of which included tenure and recency of last transaction. For this study, our target or dependent variable was a dichotomous variable, which was named Flag\_active. This target variable, Flag\_active, had a value of either 1 or 0, indicating, respectively, whether a given household was an active subscriber to the newspaper at the end of the 30-month period, from which this data was captured, or a non-subscriber (a household that has churned).

This variable was created because one of the business objectives was to identify those customers most likely to churn. By creating a dependent dichotomous variable comparing active customers to their inactive counterparts, we were able to look for explanatory variables that would be the most significant indicators and predictors of churn. Due to the disconnect between the business objectives and the available data,

the churn model was one of the only objectives we were able to examine.

*Lesson 6:* It is critical that you identify or create the appropriate target variable to meet your business objectives. If this is not done you may have statistically significant results, but you will fall short of having results that are practical and aligned with your business objectives.

For this case study, a decision tree was developed using SAS Enterprise Miner 5.2. The decision tree was the modelling technique of choice over a more classical regression approach because the decision tree did not require *a priori* distribution assumptions and we did not need to impute missing values. Missing values for critical variables of interest were extensive, making imputation not a recommended approach. A decision tree was favourable over the neural network approach because the results were primarily going to be used for explanation purposes making the decision tree easy to understand and interpret for the client. Neural networks can sometimes be viewed as a 'black box' making it difficult to explain the method of modelling.

Before running the decision tree, a Data Partition node was used in Enterprise Miner to randomly split the input SAS dataset into three datasets — Training (40 per cent), Validation (30 per cent) and Test (30 per cent). The training data are used initially to fit the model and the validation data are then used to fine-tune the model. The test data are used to estimate the error of the final model. The chi-square statistic was used to evaluate candidate splits with a maximum acceptable *p* value for each split of 0.2. Missing values were used as a value during the split search.

*Lesson 7:* Choose the correct statistical tool that fits your business objective and your communication needs.

Lesson seven is important because if the goal is an algorithm for prediction purposes to be utilised in a mailing database, a logistic regression model would be a good

choice. In contrast, if the model is to be used for explanation purposes, for presentations and internal decision making (which was the goal in this case study), a decision tree is a better choice due to its ease of use and understanding over the logistic regression coefficients.

**Step Five: Data-driven decision making**

Approximately 100 variables were considered. Twenty-two were utilised as input variables for the decision tree. Seven of the 22 variables listed in Appendix A were statistically significant predictors of a household being an active versus inactive customer. These variables included: Recency, Auto Pay, Tenure, Home Value, Delivery, Home Value and State. Table 1 provides a list of all the categorical variables used for the decision tree. Table 2 provides summary statistics for the quantitative variables used in the decision tree, which is shown in Figure 3.

In this initial analysis, the variable, Recency, shows promise for predicting customer churn. As defined, Recency is simply the time since the most recent contact between the paper and a customer. The contact could have been initiated by either the customer (‘Please hold the paper while I am on vacation’) or by the newspaper (‘We would like to offer you a special discount to switch to automatic payment’). The decision tree has breakpoints for Recency at 1.6, 7.8 and 18 months suggesting that further analysis of Recency is in order. Defining several Recency variables, based on the type of customer contact, might be useful. The decision tree shows that the percentage of households that are active customers decreases as Recency increases. Part of this decrease is inherent in the definition of Recency — if a household and the newspaper have not had any interaction for a long time, then the household is, almost by definition, inactive. Part of the higher percentage of

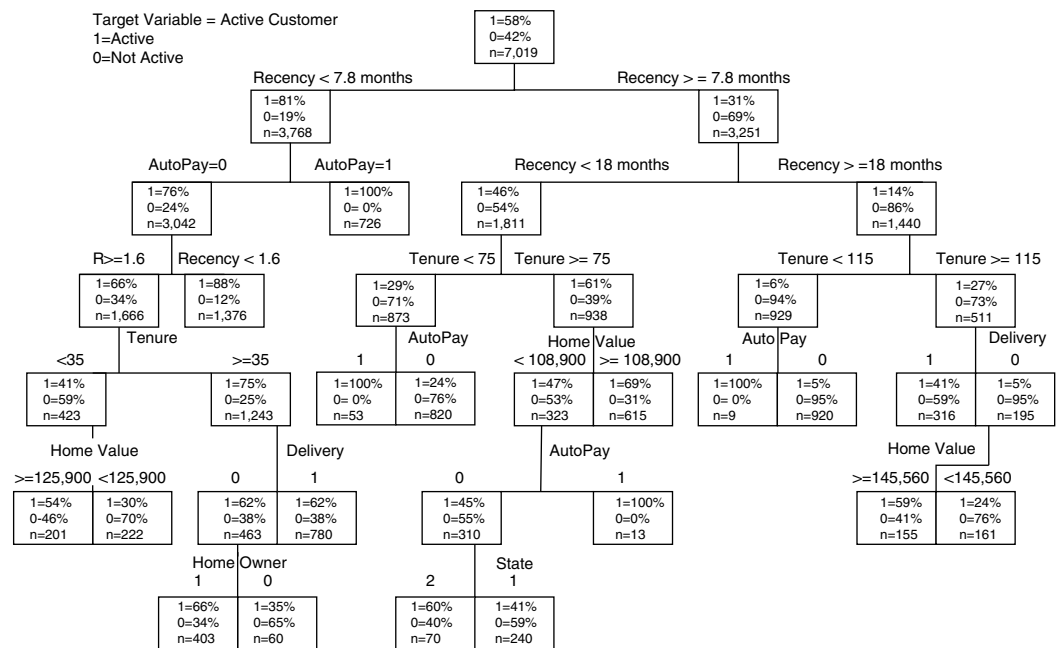
**Table 1:** Frequency distribution of categorical variables used for modelling

Variable name	Frequency	%
<i>Target variable: Active</i>		
Yes	10,141	58
No	7,411	42
<i>Children in the household</i>		
Yes	8,521	49
No	9,031	51
<i>Paid with American Express</i>		
Yes	437	2
No	17,115	98
<i>Paid more than twice w/credit card</i>		
Yes	852	5
No	16,700	95
<i>Auto pay customer</i>		
Yes	2,072	12
No	15,480	88
<i>Gender</i>		
Male	9,175	52
Female	8,377	48
<i>Home owner</i>		
Yes	13,945	79
No	3,607	21
<i>Mail responder</i>		
Yes	16,044	91
No	1,508	9
<i>Working professional</i>		
Yes	3,301	19
No	14,251	81
<i>Mail preference</i>		
Yes	1,112	6
No	16,440	94
<i>Delivery 7 days a week</i>		
Yes	9,475	54
No	8,077	46
<i>Delivery value</i>		
7 days a week	9,475	54
Monday, Friday and Sunday	33	<0.1
Monday to Friday	33	<0.1
Friday to Sunday	2,382	14
Else	5,629	32
<i>Deliver to a dwelling</i>		
Yes	17,537	99.9
No	15	<0.1
<i>Married</i>		
Yes	11,096	63
No	6,456	37
<i>State</i>		
State A	13,480	77
State B	4,072	23
<i>Income</i>		
Less than \$50,000	7,620	43
Great than or equal to \$50,000	9,932	57



**Table 2:** Summary statistics for quantitative variables used for modelling

Variable	N	Mean	s.d.	Minimum	Maximum
Recency	17,552	9.54	8.73	0	29.74
Estimated home value	16,044	171,513.77	165,411.79	0	7,122,320
Age	16,296	52.11	15.42	18	99
Length of residence	17,486	11.06	10.19	0	56
Tenure	17,552	95.2	72.8	-0.3	3,17.3
Number of times customer/newspaper contact occurred	17,552	12.29	13.68	1	203



**Figure 3:** Decision tree results

active customers at low recency (frequent contact) is, however, likely due to contacts initiated by the newspaper to increase customer satisfaction and profitability. A more focused definition of Recency should provide guidance to the type and frequency of customer contact by the paper.

The variable, Tenure, refers to the number of months that a household has subscribed to the paper and has breakpoints at 35, 75 and 115 months. The percentage of active households increases with tenure, confirming the loyalty of long-term customers. The split based on Tenure

at 35 months shows 75 per cent active customers at the longer tenure versus only 41 per cent active at the shorter tenure. This is an example of how increased knowledge could lead to the development of an action plan to support the business objective of churn prevention. This split suggests that a promotional offer for customers reaching tenure of around 30 months could be tried to increase retention. Once implemented, the impact of this promotion could be measured. After evaluating the effectiveness of this action, the iterative process would continue.

The variable, Home Value, had splits at \$108,900, \$125,900 and \$145,560 in three separate branches of the tree. In each case, the percentage of active customers is higher at the higher home value. Active customers were much more likely to be homeowners than renters.

As stated earlier, the data mining process is repetitive with each iteration giving more reliable predictions. As expected, this first analysis shows promising avenues for the next iteration, but also demonstrates that the database must be refined and more clearly focussed. The broad variable, Recency, needs a more narrow definition. Finally, a more effective measurement of the results of individual promotional mailings is needed. With reliable data on individual promotions, effective promotions can be identified resulting in increased customer satisfaction at lower cost. With each improved iteration, the newspaper will be better able to target at-risk subscribers with various forms of marketing communication to keep them as active subscribers. These efforts should be designed to keep the relationship active and might include phone calls to assess user satisfaction and impact of various promotional efforts.

### CONCLUSIONS

Data mining is an iterative process of knowledge discovery that promotes data-driven business intelligence and decision making. Data storage is growing at unprecedented rates, which drives higher demand for tools that can reduce data into information.<sup>7</sup> Setting appropriate business objectives with all domain experts involved in the data mining process is a critical first step. Data capture and retention systems have to be designed to provide the data that assists in the evaluation of these objectives. Data that may come from variant sources must be stored in a way that allows for analyses to be performed without regard to the source of the original domain area and in accordance with industry-wide SOPs. Determining the relevant variables to analyse for relationships and improved business

intelligence is a process. Part of that process entails remaining open to determining which data should be utilised and then adapting methodologies as unexpected patterns are identified.

Consider sharing your process within your industry. Only with industry exchange can we encourage standardisation towards data-driven strategies for business intelligence, CRM and churn prevention.

### References

- 1 Groves, Miles E. (2006) 'Too late to the party?', M G Strategic Research, LTC, 10th January, 2006, [http://www.astech-intermedia.com/resources/articles/MG\\_Economic\\_Notes\\_Vol\\_6\\_Issue\\_III\\_Extract.pdf](http://www.astech-intermedia.com/resources/articles/MG_Economic_Notes_Vol_6_Issue_III_Extract.pdf), accessed 16th April 2007.
- 2 Cross industry standard process for data mining <http://www.crisp-dm.org/Process/index.htm>.
- 3 Thelen, S., Motter, S. and Berman, B. (2004) 'Data mining: On the trail to marketing gold', *Business Horizons*, Vol. 46(November–December), pp. 25–32.
- 4 Magnini, V. P., Honeycutt Jr., E. D. and Hodge, S. K. (2003) 'Data mining for hotel firms: Uses and limitations', *Cornell Hotel and Restaurant Quarterly*, Vol. 44, pp. 94–105.
- 5 SAS 9.1, <http://www.sas.com/> accessed 16th April, 2007.
- 6 SAS enterprise miner documentation: <http://support.sas.com/documentation/onlinedoc/miner>, accessed 16th April, 2007.
- 7 Babcock, C. (2006) 'Data, data, everywhere', *Information Week*, 9th January, Vol. 1071, pp. 49–52.

### Appendix 'A' Variable List

1	Estimated home value
2	Age
3	Length of residence
4	Prizm code
5	Presence of children in the household
6	Credit card user
7	Credit account
8	Delivery
9	Delivery value
10	Type of dwelling
11	Auto pay customer
12	Gender
13	Home owner
14	Income
15	Mail preference
16	Mail responder
17	Marital status
18	Recency
19	State
20	Tenure
21	Tresp/dedupe
22	Working professional