Predictive Models in Churn Data Mining: A Review

David L. García, Alfredo Vellido, and Àngela Nebot

Llenguatges i Sistemes Informàtics. Universitat Politècnica de Catalunya

Abstract. The development of predictive models of customer abandonment plays a central role in any churn management strategy. These models can be developed using either qualitative approaches or can take a data-centred point of view. In the latter case, the use of Data Mining procedures and techniques can provide useful and actionable insights into the processes leading to abandonment. In this report, we provide a brief and structured review of some of the Data Mining approaches that have been put forward in recent academic literature for customer abandonment prediction.

1. Introduction

Anticipating a customer's intention to abandon their current provider company should be considered a key element of any *therapeutic* strategy in churn management. Continuing with the medical analogy, there is no use in predicting an illness unless it is done in time to administer the appropriate treatment. Early diagnosis of customer abandonment should, at the very least, reduce the aggressiveness of the required therapy, increasing as a result the possibilities of customer recuperation.

However, not all cases of churn -abandonment of the commercial relationship between the company and its customer- are equally important, nor are they all predictable. According to the reasons behind their abandonment, customers can be classified in different typologies:

- ***Involuntary cancellation***: Referred to customers from which the actual company withdraws their service (fraud, arrears, …). Generally, companies do not even consider these cancellations as abandonment for their records.
- ***Voluntary cancellation***: It corresponds to customers who consciously decide to change provider. Two variants can be considered:
  - ✓ ***Circumstantial***: Due to changes in the customer's circumstances which do not allow them to continue (change of address, inclusion in the company's social benefit plans, change of marital status, children, …). This cancellation is intrinsically unpredictable.
  - ✓ ***Deliberate***: This occurs when the customer voluntarily decides to abandon their current provider for a competitor.

In the following, we will restrict our review to the *deliberate* scenario. The rest of the report is organized as follows: in section 2, the different stages of a standard process of design and development of a predictive model of abandonment will be described, together with the associated literature. This literature will be summarily compiled in section 3. The report concludes with a discussion section.

2. Building predictive models of abandonment

The process of design and development of abandonment predictive models can be divided into four stages (Datta et al., 2000), as seen in Figure 1. The last three stages of this process form a cycle that ends when adequate prediction results are achieved. We will now take a closer look at each stage in turn.

2.1 Stage 1: Identifying and obtaining the best data

This might arguably be the most relevant stage in the construction of a predictive abandonment model. Experience shows that the quality and suitability of the available data determines the accuracy and predictive power of the resulting model. Different data combinations may be better or worse indicators for different problems and for different sectors. Ultimately, it is a question of identifying the data which best fit the type of analysis being carried out; only in this way will we be able to extract, in subsequent stages, knowledge and actions which are useful and actionable in business terms.

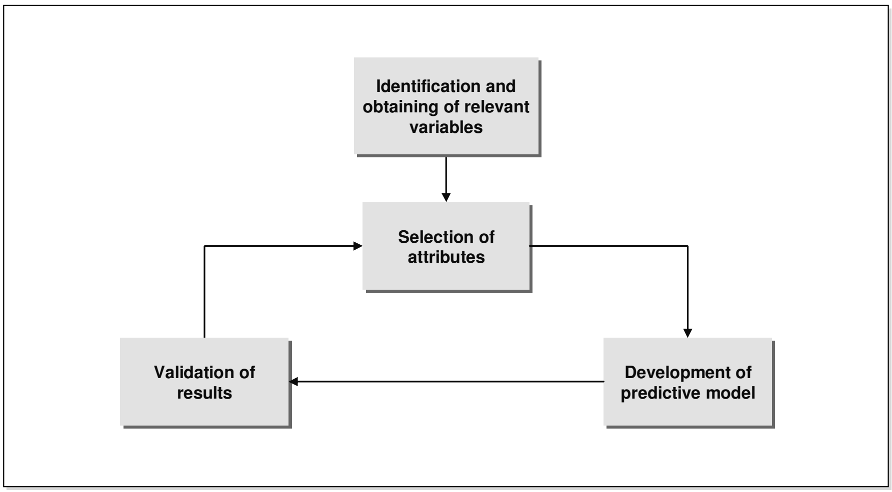

Figure 1: Stages of the predictive model building process

The recent literature presents a selection of different data requirements for the analysis of abandonment. The most relevant are detailed below:

- A large group of studies in this field consider the variables of customer *use/consumption* as key elements in identifying abandonment:
    - ✓ Madden et al. (1999), in their customer retention model for the Australian ISP (Internet Service Provider) industry, classified and used four categories of variables: *economic*, *use*, *ISP choice* and *demographics*. Ng and Liu (2000), suggested their use for identifying churn in the telecommunications market.

- In their study, Verhoef and Donkers (2001) concluded that the purchase of products and services can be better predicted using historic purchasing data.
- This last view is backed, more recently, by Hsieh (2004), who proposed that the analysis of transaction data, through historic account and customer data, can provide us with clues to identify the best incentives for a bank to offer its customers and improve the marketing strategy.
- Data on customer use have also been used to identify the behaviour of website-using customers (Jenamani et al., 2003) and to predict repeat purchasing by mail (Van den Poel, 2003).

- A series of other authors (Pfeifer and Carraway, 2000; Ho Ha et al., 2002; Van den Poel, 2003; Verhoef et al., 2002; Hsieh, 2004; Jonker et al., 2004 and Liu and Shih, 2005a,b) coincide in suggesting the use of three groups of variables[1] as a source for predicting the abandonment probability of a particular customer:
    - Length of time since last purchase.
    - Frequency of use.
    - Economic expense effected over a certain time period.
- Finally, Hung et al. (2006), considered in their study that the most significant variables for churn prediction in the mobile telephone industry are: *demographic data* (age, penetration rate, and gender), *payment and account data* (monthly quota, billing amount, arrears account), *call details* (call duration, cal type -PSTN, IDD, to mobile, etc.-) and *customer service data* (number of PIN number changes, number of blocks and suspensions).

This stage in the building of predictive models of abandonment would fit, from a Data Mining process point of view, the phases of *problem* and *data understanding* and the subsequent one of *data pre-processing*. Bearing this in mind, and from a practical point of view, it is important to note that the predictive model should ideally be constructed on the basis of the available data gathered routinely by the company from the whole customer base, although this can be an extremely costly process. Consequently, those data bearing most of the predictive power may not always be available. We are faced, as a result, with the problem of identifying the best data from what is available.

The process of understanding and interpreting the data often presents difficulties. Even though the data in each field of a database may seem self-explanatory and unambiguous at times, interpretation can become difficult because of the use of specific and *ad hoc* company terms, different numerical formats, or simply because their meaning is different from the apparently obvious. Given the usual lack of standards to facilitate this process at the company level, its success is largely based on the existence of good communication between database managers and the data analysts. In fact, these Data Mining stages have not been duly documented in the majority of investigations carried out in recent years (Hadden et al., in press).

[1] Group of variables globally known as RFM (***R***ecency, ***F***recuency and ***M***onetary).

2.2 Stage 2: Selection of attributes

This stage consists of the selection of the most appropriate fields for prediction, that is to say, those which minimise the classification or prediction error. This process is important as it helps to reduce the dimensionality of the data so that only the important attributes are included, whereas the redundant, noisy and or irrelevant ones are excluded (Yan et al., 2004). Two phases can be distinguished in the selection of attributes:

- **Search phase.** Three search categories are considered: optimal, heuristic and random.
    - ✓ ***Optimal***: The simplest method of optimal search consists of an exhaustive search. In this type of search, the quantity of subgroups of positive attributes grows rapidly becoming unmanageable even for moderately sized groups of attributes. However, there are some optimal search methods, such as the "branch and bound" algorithm (Sun et al., 2004), which avoid this exhaustiveness and its drawbacks.
    - ✓ ***Heuristic***: There are two co-existing methods, both well-known in the literature: *sequential forward selection* (SFS) and *sequential backward selection* (SBS). SFS is performed by starting with an empty group of attributes and gradually adding the best individual attributes. Meanwhile, in SBS the method starts with a complete group and gradually removes the attributes that are least significant. Combinations of both have also been put forward in different variants.
    - ✓ **Random**: In random searches, probabilistic or sampling techniques are used. One of the most used methods, the basis for several variations, is known as *relief* algorithm and consists of assigning weightings to the attributes based on the estimated effectiveness of the classifier, and by selecting those whose weighting exceeds a given limit -defined by the user- the classifier becomes trained.
- **Evaluation phase**. Once the attribute search phase has been carried out we turn to its evaluation. The data used for the selection phase should be independent of those used in the validation phase in order to avoid the risk of data overfitting. Two alternative methods are considered in the evaluation phase: *filter* and *wrapper*.
    - ✓ **Filter**: Evaluates the attribute subgroups independently from the design of the classifier. The filter method is computationally more efficient since it evaluates the usefulness of the selected attributes based on criteria that can be tested rapidly -for example, the reduction of the correlation or the mutual information between attributes-. Nevertheless, the filter method can lead to non optimal attributes, especially if there is dependence between the attributes and the classifier; and, as a consequence, the result of the classifier may be poor.
    - ✓ **Wrapper:** The evaluation of the attribute subgroup is carried out using the same learning algorithm that is employed in the design of the classifier (training the classifier with the selected attributes and using a group of additional data for its validation estimating the classification error). Although the wrapper method is slower, the selected attributes are generally better for the classifier used.

2.3 Stage 3: Development of a predictive model

Once the data available for analysis have been selected, the next stage entails the selection of the most suitable methods and techniques for building the predictive model. In a simple manner, a predictive model can be defined as one that extracts patterns from the available data in order to infer future situations (Rygielski et al., 2002). In the area of abandonment prediction, the most important modelling techniques include decision trees and neural networks (Crespo and Weber, 2005). The following subsections provide an overview of both traditional and soft computing techniques used for predictive churn modelling.

Standard methods

- **Decision trees (DT):** The most popular type of predictive model is the decision tree. DT have become an important knowledge extraction structure, used for the classification of future events (Muata and Bryson, 2004). There are two distinct phases in their design: building and pruning.

  - ✓ **Tree building:** Consists of recursively partitioning the training sets according to the values of the attributes. The partitioning process continues until all, or most of the records in each of the partitions contain identical values.

  - ✓ **Tree pruning:** Involves selecting and removing the branches that contain the largest estimated error rate. Tree pruning is known to enhance the predictive accuracy of the decision tree, while reducing the complexity (Au et al., 2003).

  Two widely used DT models are C5.0 and CART. The **C5.0 classification tree** assembles classification trees by recursively splitting the instance space into smaller subgroups until only instances from the same class remain known as a pure node, or a sub-group containing occurrences from different classes known as impure nodes. The tree is allowed to grow to its full potential before it is pruned back in order to increase its power of generalisation on unseen data.

  On the other hand, a **classification and regression tree (CART)** is constructed by recursively splitting the instance space into smaller sub-groups until a specified criterion has been met. The decrease in impurity of the parent node against the child nodes defines the goodness of the split. The tree is only allowed to grow until the decrease in impurity falls below a user-defined threshold. At this time the node becomes a terminal, or leaf node (Bloemer et al., 2003).

  The recent literature contains some examples of DT being used for the construction of models for abandonment prediction:

  - ✓ Datta et al. (2000) carried out research in the area of churn prediction and developed a model that they called Churn Analysis Modelling and Prediction (CHAMP). CHAMP also uses decision trees to predict customer churn in the telecommunications industry.

  - ✓ Ng and Liu (2000), for the purpose of identifying potential defectors, chose C4.5, a decision tree induction method for classification, which automatically generates classification rules.

  - ✓ Experiments performed by Hwang et al. (2004), using data of the wireless telecommunications market, involved DT, neural networks and logistic regression. The DT showed slightly better accuracy over the other methods; however, the authors state that these results do not prove DT to be the best choice in all cases. This is supported by results reported in Mozer et al. (2000).

- **Regression analysis:** It is a standard and popular technique used by researchers dealing with the prediction of customer satisfaction. Mihelis et al. (2001) developed a method to determine customer satisfaction using an ordinal regression based approach. Another model for assessing the value of customer satisfaction was developed by Rust and Zahorik (1993). They used logistic regression to link satisfaction with attributes of customer retention. They claim that the logistic function can be interpreted as providing the retention probability.

Kim and Yoon (2004) used a *logit* model to determine subscriber churn in the telecommunications industry, based on discrete choice theory (study of behaviour in situations where decision makers must select from a finite set of alternatives).

Hwang et al. (2004) found out that logistic regression performed better than neural networks and DT for predicting customer churn.

The main model used to develop the churn prediction platform presented by Datta et al. (2000) was neural networks; however, they also experimented with KNN and DT. Their research could not establish a best method, and they included the explanation of customer behaviour as a future research goal because, although their model could predict customer churn, it was unable to provide an explanation as to why customers might churn.

Soft Computing methods

Soft Computing methods (such as fuzzy logic, neural networks, and genetic algorithms) provide, in one form or another, flexible information processing capabilities for handling real life problems. Exploiting the tolerance for imprecision, uncertainty, approximate reasoning and partial truth in order to achieve tractability, robustness, low solution cost, and close resemblance with human-like decision making, is the aim of Soft Computing methods (Pal and Ghosh, 2004). Techniques that fall into the category of Soft Computing are evolutionary computation (EC), artificial neural networks (ANN), fuzzy logic (FL), probabilistic computing and their combinations, such as neuro-fuzzy systems (Hadden et al., in press). We shall briefly review these from the perspective of predictive models of customer abandonment:

- **ANN:** Artificial neural networks have been successfully used to estimate complex non-linear functions. A neural network is a Machine Learning model, loosely based on the biological brain (a natural neural network). It has successfully been applied to many types of problems, such as classification, control and prediction (Behara, 2002). One of the features that make ANN different from DT and other classification techniques is that their prediction can be interpreted as a probability. An important factor when considering the practical use of ANN is that they do not necessarily uncover patterns in an easily understandable form (Au et al., 2003). Datta et al. (2000) stated that ANN were still scarcely being used by companies in their day-to-day operations, and considered that a possible reason for this could be lack of clear interpretability of their output. In spite of that, many authors have used ANN in an entrepreneurial setting (Lisboa et al., 2000).
- **DMEL (Data Mining by Evolutionary Learning):** An algorithm proposed by Au et al. (2003), aimed to overcome the limitations of interpretation and understanding of the results obtained through Soft Computing techniques -in contrast with the clarity of the if-then-rules obtained through DT, for example-. DMEL uses non-random initial population based on first order rules. Higher order rules are then obtained iteratively using a GA type process. The fitness value of a chromosome uses a function that defines the probability that the attribute value is correctly determined using the rules it encodes. The likelihood of prediction is estimated and the algorithm handles missing values.

DMEL was used to predict churn in the telecommunications industry by Au et al. (2003).

- **Bayesian networks:** Baesens et al. (2004) reported an attempt to estimate whether a new customer will increase or decrease future spending. A Bayesian network was defined in this work as probabilistic "white box", representing a joint probability distribution over a set of discrete stochastic variables.

Other alternative methods

- **Semi-Markov processes**: Used by Jenamani et al. (2003) to propose a model that considers e-customer behaviour. The discrete-time semi-Markov process was designed as a probabilistic model, for use in analysing complex dynamic systems.
- **Mixture transition distribution (MTD):** Prinzie and Van den Poel (2006) introduced a mixture transition distribution (MTD) to investigate purchase-sequence patterns. The MTD was designed to allow estimations of high order Markov chains, providing a smaller transition matrix facilitating managerial interpretation.
- **Goal-oriented sequential pattern:** Chiang et al. (2003) introduced their own algorithm for identifying potential churners using association rules, which are defined as a technique that identifies relationships amongst variables. The authors defined two steps for finding out association rules: The first step entails the detection of the large item set (attribute-value pairs), requiring compliance with certain minimum conditions of *support* and minimum *confidence* defined by the researcher. In the second step, an *A Priori* algorithm is used to explore the rules of association.

2.4 Stage 4: Validation of results

The most used methods for model validation are:

- **Cross-fold validation:** Most suitable in those cases in which there is a scarcity of data. Hwang et al. (2004), performed validation by creating a 70/30 divide of the data. The 70% divide created the training set, and the 30% divide created the validation set. Cross-fold validation is based on the principle of using the available data for both training and validation. Several cross-validation methods have been proposed in the literature (Hadden et al., in press), including:
    - ✓ **V-fold cross validation:** The learning set is randomly partitioned into limited datasets of equal size. Each set is then used as a validation set.
    - ✓ **Monte Carlo cross validation:** The learning set is repeatedly divided into two random sets, one of which is used for training and the other for validation.
- **Separate validation dataset:** Several authors (Datta et al., 2000; Bloemer et al., 2003; Prinzie and Van den Poel, 2006) have successfully used validation sets separated from the training sets in the validation of their predictive models of abandonment. This method of validation is more suitable in those cases in which data availability is not an issue.

3. A summarized review of the literature

The following summary Table 1 lists the main references that have appeared in the recent literature relating to the building of predictive models of customer abandonment. Following the same scheme proposed as the guiding index for this chapter, the table shows the references to the articles, the type of data used in the analysis, the source from which it has been obtained; attribute selection technique employed; if the model uses temporary data series in its definition and, finally, the technique used to develop the predictive model.

Table 1: List of references on abandonment prediction modelling

Authors	Data type	Data gathering	Attribute selection techniques	Time periods	Modelling techniques
Au et al. (2003)	Customer localisation, customer type, payment method, service plan, monthly use, number of calls made and number of calls abnormally ended (251 variables)	Database	Interviews with experts.	No (average of 2 months for training and 1 month of prediction)	Algorithm DMEL (based on GA) comparing with C4.5 and ANN (using lift factor)
Baesens et al. (2004)	Purchasing behaviour: volume of purchases during first 6 months as customer; "broadness" of purchases; "bargaining tendency" and "price sensitivity"; evolution averages during first 6 months	Database	Not specified	Yes (8 weekly periods)	Algorithms of classification based on Bayesian Networks
Chiang et al. (2003)	Transaction data (frequency of transactions of banking customers are analysed)	Database	The algorithms used are based on selecting the variables involved in the rules.	No (last operative month in the last 6 for each customer)	Own algorithm: "*Goal-oriented sequential pattern*" (based on association rules and compared to algorithm a priori. Adaptation of the "Goal-orirnted pattern")
Ho Ha et al. (2002)	RFM variables	Database	Not specified	Yes (18 months; without specifying how many periods they are divided into)	SOM networks
Hung et al. (2006)	Transaction and contract data: demographic, payment, call details and customer service data.	Database	Initial selection of variables via interviews. Final selection via z-test.	No (6 month: averages and totals. 1 month prediction)	Decision Trees (C5.0) Neural networks: ANN with BPN.

Hwang et al., (2004)	Socio-demographic and usage variables	Database	R^2 Method	No (6 months data)	Neural networks, Decision Trees and logistic regression
Kim and Yoon (2004)	Company service, demographic and service use characteristics.	Survey	Not specified	No	regresión (Logia Modelo)
Madden et al. (1999)	4 categories: use, economical, choice of provider and demographical.	Web Survey	Not specified	No	regresión (Probit Binomial model)
Mozer et al. (2000)	Call details, quality (interferences and signal coverage), financial and service application (contract details, rate plan, handset type and credit report) and demographic information.	Database	Not specified	No (average of 3 months to predict churn of the following 2)	Regression, Neural networks and Decision Trees (C 5.0)
Ng and Liu (2000)	Usage data	Database	Induction algorithm: Using decision trees	Not specified	Decision Trees (C4.5)
Shin and Sohn (2004)	Transactions carried out	Database	Not specified	No (3 month: averages and totals)	K-means, SOM map and Fuzzy K-means
Van den Poel (2003)	RFM, behaviourals and specifics of the company, non-behaviourals and specifics of the company (satisfaction), behaviourals and non-specifics of the company (prediction of whether purchase by post will be repeated or not)	Survey and Database	Sequential Search Algorithm.	No (4 year data. One Survey only. 6 months of prediction)	Regression (Logit model)
Van den Poel and Larivière (2004)	Customer behaviour information, socio-demographic information, merger and prosperity index	Database	Not specified	Yes (77 year database. Length of sub-periods are not specified)	Survival análisis
Wei and Chiu (2002)	Variables relative to the contract: length of service, payment type, contract type Consumption variables: minutes of use, frequency and sphere of influence	Database	Interviews with experts	Yes (3 periods: Observation, Retention and prediction.)	Decision Trees (C4.5)

4. Discussion

In this report, we have carried out a review of the models and techniques used in the field of prediction of customer abandonment -churn- that have been proposed in recent literature, aiming to gain a broad vision of the subject in order to lay the foundations of the research concerning, either directly or indirectly, the thesis proposal of the first signing author. By way of corollary, we can extract some practical conclusions:

No much research has been carried out on models that take into account customers' *memory of satisfaction in previous interactions with the company* as a dimension that could explain their level of bonding. Although sporadic references have in the past appeared in the literature, the difficulty of obtaining recurring data on service evaluation makes it difficult to build an empirical model. This "absence of memory" can be extended to the development of models of abandonment prediction found in the literature. In general, the reviewed models do not work with time series of customer data and, as a result, are created based on static snapshots -corresponding to periods that may be longer or shorter-. This makes it much more difficult to construct an effective system of preventive churn warning.

With few exceptions, the existing prediction models work with anticipatory scenarios that are too short-sighted to be used in real commercial actions addressed to customers who show a tendency to abandon. Identifying the precise, clear patterns and trends of abandonment serves no purpose unless the can be rapidly converted into retention actions. Early prediction of the problem should significantly reduce the aggressiveness of the required retention actions, with the consequent increase in the chances of winning back the customers even before they take the decision to leave.

As a conclusion, there is clear research interest in the idea of developing abandonment prediction models based on time series of customers' data to generate customer churn early warnings.

Bibliography

Au, W.-H., Chan, K.C.C. and Yao, X. 2003. A novel evolutionary data mining algorithm with applications to churn prediction. *IEEE Transactions on Evolutionary Computation*, 7:532–545.

Baesens, B., Verstraeten, G., Van den Poel, D., Egmont-Peterson, M., Van Kenhove, P. and Vanthienen, J. 2004. Bayesian network classifiers for identifying the slope of the customer lifecycle of long-life customers. *European Journal of Operational Research*, 156:508–523.

Behara, R.S., Fisher, W.W. and Lemmink, J.G.A.M. 2002. Modelling and evaluating service quality measurement using neural networks. *International Journal of Operations and Production Management*, 22:1162–1185.

Bloemer, J.M.M., Brijis, T., Vanhoof, K. and Swinnen, G. 2003. Comparing complete and partial classification for identifying customers at risk. *International Journal of Research in Marketing*, 20:117–131.

Chiang, D., Wang, Y., Lee, S. and Lin, C. 2003. Goal-oriented sequential pattern for network banking and churn analysis. *Expert Systems with Applications*, 25:293–302.

Crespo, F. and Weber, R. 2005. A methodology for dynamic data mining based on fuzzy clustering. *Fuzzy Sets and Systems*, 150:267–284.

Datta, P., Masand, B., Mani, D.R. and Li, B. 2000. Automated cellular modelling and prediction on a large scale. *Artificial Intelligence Review*, 14(6): 485–502.

Hadden, J., Tiwari, A., Roy, R. and Ruta, D. Computer assisted customer churn management: State-of-the-art and future trends. *Computers & Operations Research,* In press.

Ho Ha, S. Min Bae, S. and Chan Park, S. 2002. Customer's time-variant purchase behaviour and corresponding marketing strategies: an online retailer's case. *Computers & Industrial Engineering*, 43:801–820.

Hsieh, N. 2004. An integrated data mining and behavioural scoring model for analysing bank customers. *Expert Systems with Applications*, 27:623–633.

Hung, S.Y., Yen, D.C. and Wang, H.Y. 2006. Applying data mining to telecom churn management. *Expert Systems with Applications* 31(3): 512-524.

Hwang, H., Jung, T. and Suh, E. 2004. An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry. *Expert Systems with Applications*, 26:181-188.

Jenamani, M., Mohapatra, P.K.J. and Ghose, S. 2003. A stochastic model of e-customer behaviour. *Electronic Commerce Research and Applications*, 2:81–94.

Jonker, J., Piersma, N. and Van den Poel, D. 2004. Joint optimization of customer segmentation and marketing policy to maximize long-term profitability. *Expert Systems with Applications*, 27:159–168.

Kim, H. and Yoon, C. 2004. Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market. *Telecommunications Policy*, 28:751–765.

Lisboa, P.J.G, Edisbury, B., and Vellido, A. (Eds.) *Business Applications of Neural Networks.* Singapore: World Scientific, 2000.

Liu, D. and Shih, Y. 2005a. Integrating AHP and data mining for product recommendation based on customer lifetime value. *Information & Management,* 42:387–400.

Liu, D. and Shih, Y. 2005b. Hybrid approaches to product recommendation based on customer lifetime value and purchase preferences. *The Journal of Systems & Software*, 77:181–191.

Madden, G., Savage, S.J. and Coble-Neal, G. 1999. Subscriber churn in the Australian ISP market. *Information Economics and Policy*, 11:195–207.

Mihelis, G., Grigoroudis, E., Siskos, Y., Politis, Y. and Malandrakis, Y. 2001. Customer satisfaction measurement in the private bank sector. *European Journal of Operational Research*, 130:347–360.

Mozer, M.C., Wolniewicz, R., Grimes, D.B., Johnson, E. and Kaushansky, H. 2000. Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry. *IEEE Transactions on Neural Networks,* 11:690-6.

Muata, K. and Bryson, O. 2004. Evaluation of decision trees: a multi criteria approach. *Computers & Operational Research,* 31:1933–45.

Ng, K. and Liu, H. 2000. Customer retention via data mining. *Artificial Intelligence Review*, 16(4): 569-590.

Pal, S.K. and Ghosh, A. 2004. Soft computing data mining. *Information Sciences,* 163(1–3): 5-12.

Pfeifer, P.E. and Carraway, R.L. 2000. Modeling customer relationships as Markov chains. *Journal of Interactive Marketing*, 14(2): 43 - 55

Prinzie, A. and Van den Poel, D. 2006. Investigating purchasing-sequence patterns for financial services using Markov, MTD and MTDg models. *European Journal of Operational Research*, 170:710-34.

Rust, R.T. and Zahorik, A.J. 1993. Customer satisfaction, customer retention, and market share. *Journal of Retailing*, 69:193–215.

Rygielski, J., Wang, J. and Yen, D.C. 2002. Data mining techniques for customer relationship management. *Technology in Society*, 24:483–502.

Shin, H.W. and Sohn, S.Y. 2004. Segmentation of stock trading customers according to potential value. *Expert Systems with Applications*, 27:27–33.

Sun, Z., Bebis, G. and Miller, R. 2004. Object detection using feature subset selection. *Pattern Recognition*, 37:2165–2176.

Van den Poel, D. 2003. Predicting mail-order repeat buying: which variables matter?. *Tijdschrift voor Economie and Management*, 48(3):371–403, 2003.

Van den Poel, D. and Larivière, B. 2004. Customer attrition analysis for financial services using proportional hazard models. *European Journal of Operational Research*, 157:196–217.

Verhoef, P.C. and Donkers, B. 2001. Predicting customer potential value an application in the insurance industry. *Decision Support Systems*, 32:189–99.

Verhoef, P.C., Spring, P.N., Hoekstra, J.C. and Leeflang, P.S.H. 2002. The commercial use of segmentation and predictive modelling techniques for database marketing in The Netherlands. *Decision Support Systems*, 34:471–81.

Wei, C.P. and Chiu, I.T. 2002. Turning telecommunications call details to churn prediction: a data mining approach. *Expert Systems with Applications*, 23(2):103-12.

Yan, L., Wolniewicz, R.H. and Dodier, R. 2004. Predicting customer behaviour in telecommunications. *IEEE Intelligent Systems*, 19:50–58.