CRANFIELD UNIVERSITY

JOHN HADDEN

A Customer Profiling Methodology for Churn Prediction

SCHOOL OF APPLIED SCIENCES

PhD

CRANFIELD UNIVERSITY

SCHOOL OF APPLIED SCIENCES
MANUFACTURING DEPARTMENT

PHD
Academic Year 2007-2008

JOHN HADDEN

A Customer Profiling Methodology for Churn Prediction

Supervisors: Dr. Ashutosh Tiwari and Prof. Rajkumar Roy
April 2008

This thesis is submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy

Abstract

As markets have become increasingly saturated, companies have acknowledged that their business strategies need to focus on identifying those customers who are most likely to churn. To address this, a method is required that can identify these customers, so that proactive retention campaigns can be deployed in a bid to retain them. To further complicate this, retention campaigns can be costly. To reduce cost and maximise effectiveness, churn prediction has to be as accurate as possible to ensure that only the customers who are planning to switch their service providers are being targeted for retention.

Current techniques and research as identified by literature focus primarily on the instantaneous prediction of customer churn. Much work has been invested in this method of churn prediction and significant advancement has been made. However one of the major drawbacks of current research is that the methods available do not provide adequate time for companies to identify and retain the predicted churners. There is a lack of time element in churn prediction. Current research also fails to acknowledge the expensive problem of misclassifying non-churners as churners. In addition, most research efforts base their analysis on customer demographic and usage data that can breach governing regulations. It is proposed in this research that customer complaints and repairs data could prove a suitable alternative.

The doctoral research presented in this thesis aims to develop a customer profiling methodology for predicting churn in advance, while keeping the misclassification levels to a minimum. The proposed methodology incorporates time element in the prediction of customer churn for maximising future churn capture by identifying a potential loss of customer at the earliest possible point. Three case studies are identified and carried out for validating the proposed methodology using repairs and complaints data. Finally, the results from the proposed methodology are compared against popular churn prediction techniques reported in literature. The research demonstrates that customers can be placed into one of several profiles clusters according to their interactions with the service provider. Based on this, an estimate is possible regarding when the customer can be expected to terminate his/her service with the company. The proposed methodology produces better results compared to the current state-of-the-art techniques.

Acknowledgements

I would like to take this opportunity to thank my principal supervisor, Dr. Ashutosh Tiwari, for all his dedicated support and guidance throughout this research. His guidance and inspiration are the key factors that enabled successful completion of this research. Thanks also to my second supervisor, Professor Rajkumar Roy, for all his help and valuable guidance in this research.

I would also like to thank the funding body of this research: Engineering and Physical Sciences Research Council (EPSRC), and its industrial collaborators: British Telecom (BT), particularly my industrial supervisor at BT Dr. Dymitr Ruta for all his help and support.

I would like to thank all the members of the Decision Engineering Centre (DEC) for the useful discussions we had throughout the past 4 years. Thanks also to all the staff of the School of Applied Sciences for their support and encouragement.

List of Publications

Hadden, J., Tiwari, A., Roy, R. and Ruta, D. (2007). Computer assisted customer churn management: State-of-the-art and future trends. *Computers and Operations Research,* Elsevier, vol. 34, no. 10, 2902-2917, ISSN: 0305-0548.

Hadden, J., Tiwari, A., Roy, R. and Ruta, D. (2006). Churn prediction: Does technology matter? *International Journal of Intelligent Technology,* World Enformatika Society, vol. 1, no. 2, 104-110, ISSN: 1305-6417.

Hadden, J., Tiwari, A., Roy, R. and Ruta, D. (2006). Churn prediction using complaints data. *Transactions on Engineering, Computing and Technology,* Enformatika, vol. 13, 158-163, ISSN: 1305-5313.

Roy, R., Tiwari, A., Shah, S. and Hadden, J. (2006). Soft computing in the service industry. In: *6th International Conference on Recent Advances in Soft Computing (RASC-2006)*, ISBN: 1-902671-43-0, 10-12 July, Canterbury, UK, 1-9.

Hadden, J., Tiwari, A., Roy, R., Ruta, D. (2008). Computer aided churn prediction using customer profile analysis, *Computers in Industry*, Elsevier, 0166-3615: (Submitted)

Hadden, J., Tiwari, A., Roy, R., Ruta, D. (2008). A novel churn prediction methodology through neural network based customer profile analysis, *Expert Systems with Applications,* Pergamon, 0957-4174: (Submitted)

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AID	Automatic Interaction Detection
AUC	Area Under the Curve
BT	British Telecom
CART	Classification and Regression Trees
CASE	Computer Aided Software Engineering
CHAMP	Churn Analysis Modelling and Prediction
CLV	Customer Lifetime Value
CRM	Customer Relationship Management
DB	Database
DEC	Decision Engineering Centre
DMEL	Data Mining by Evolutionary Learning
DT	Decision Tree
EC	Evolutionary Computing
EPSRC	Engineering and Physical Sciences Research Council
EQ	Estimating Quality
ER	Entity Relationship Model
FFS	Forward Feature Selection
FL	Fuzzy Logic
GA	Genetic Algorithm
GP	Genetic Programming
GUI	Graphical User Interface
ISP	Internet Service Provider
ISP	Internet Service Provider
JDBC	Java Database Connectivity
KISS	Keep it Simple, Stupid
KNN	k-Nearest Neighbour
LOY	Loyalty Index Score
LRM	Logistic Regression Model
LW	Layer Weight
MS	Microsoft

MSE	Mean of Squared Error
MTD	Mixture Transition Distribution
NN	Artificial Neural Network
NRE	Net Revenue Equation
RAD	Rapid Application Development
RFM	Recency, Frequency, Monetary
RM	Relational Model
ROC	Receiver Operating Characteristic
RQI	Relative Quality Importance
SAT	Customer Satisfaction Index Score
SBFS	Sequential Backward Floating Search
SBS	Sequential Backward Selection
SDLC	Software Development Lifecycle
SFFS	Sequential Forward Floating Search
SFS	Sequential Froward Selection
SOM	Self Organising Map
SPSS	Statistical Package for the Social Sciences
SQL	Standard Query Language
SSE	Sum of Squared Error
SVM	Support Vector Machine
UI	User Interface
URL	Uniform Resource Locators
USC	Usage Segment Code
USM	Unifying Semantic Model
XP	Extreme Programming
YAGNI	You Aren't Going to Need It
YEAR	Year of Data

This Page is Intentionally
Left Blank

1. Introduction

With new companies continuously emerging in the telecommunications industry, and new customers to those companies sparse, markets have matured to a point where they have saturated, so possible new customer acquisition is scarce. Because of this, companies are beginning to recognise that their most valuable assets are their existing customers. Holding onto existing customers is not an undemanding procedure, with further complications arising from governing bodies such as Ofcom. These authorities have emerged to act in the customers' best interests, and the promotion of healthy competition amongst service providers to ensure that the customer has a choice, and services are not monopolised. A further challenge that is commonly being recognised regarding the acquisition of new customers is lack of data, making targeted sales campaigns difficult to deploy.

In today's business world companies are recognising that customer value and increased revenue is more likely to come from their existing customer base than from new customer acquisition. The reasoning behind this is that companies know their existing customers, already have a relationship with them, and amplitude of data on them. In common recognition of this problem, industry has seen an emergence of customer relationship management (CRM) products. Software companies have realised that CRM has become a receptive topic within industry and therefore an area of potential wealth and opportunity for the development and promotion of products that promise to not only boost customer retention, but also for increasing selling opportunities.

There are three core components that are typically covered by CRM: cross-selling, up-selling and customer retention. Each of these areas of CRM is described below. The focus of this research is on customer retention.

1.1. CRM

Cross-selling is attempting to promote products to the customer that the customer does not usually purchase. For example a company might recognise that a customer has a history of purchasing books from a specific author. Cross-selling could take the genre that the author typically covers in his/her writing and use that as a basis of recommending books to the customer from alternative authors who write in the same genre, attempting to persuade the customer to try something different.

Up-selling attempts to promote repeat purchases of the same products that the customer has purchased in the past. Up-selling also involves the promotion of more expensive versions of the same service as a way of increasing revenue.

The focus of this research is on customer retention, which is the process of retaining existing customers. There are several reasons why retaining existing customers is important. The first reason is that markets have become saturated to a point where new custom to the industry is scarce. The second reason is cost. New customer acquisition can be costly to a business for numerous reasons:

> It has been reported that the acquisition of new customers can be over ten times more costly to a business than retaining existing customers. In the US new customer acquisition cost around $300 per customer in 2004, compared to retention costs of around $25 per customer in the same year. This is largely because in saturated markets, the acquisition of new customers often involves enticing customers away from competitors through offers of expensive special deals (Seo et al., 2008).
>
> Cross-selling and up-selling can be difficult with new customers because the company has had no relationship with them, and hence holds no data on them. Increasing the revenue value of a new customer can take a long time (Seo et al., 2008).

Customer retention addresses the issue of customer churn, where churn describes the turnover of customers, and churn management describes the efforts a company makes to identify and control the problem of customer churn (Hung et al., 2006).

1.2. Introduction to the Techniques for Modelling Customer Churn

In order to manage customer churn within a company it is important to build an effective and accurate customer churn model. There are several modelling techniques available that can aid in the prediction of customer churn. The most common techniques have been identified from literature as:

Classification and regression trees (CART)
Logistic regression models (LRM)
Artificial neural networks (NN)

The list provided above is not exhaustive; these are the techniques that most commonly appear throughout customer churn research. Other statistical and classification methods may also apply to the problem domain, however they have either not been widely explored or they have failed to provide satisfactory results (Coussement and Van Den Poel, 2008). This section will introduce the above commonly used methods.

1.2.1. CART

CART (also known as recursive partitioning regression) is a popular classification technique used for predicting events. CART works by recursively dividing the response variables into increasingly homogenous subsets based on significant thresholds of the predictor variables (Lozano et al., 2008). CART development usually consists of two phases: tree building and tree pruning. The tree building phase consists of recursively partitioning the training set 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. Certain branches may need to be removed because it is possible that they may contain noisy data. The trees are allowed to grow as large as possible to avoid early stopping where important rules could be missed. The trees are then pruned backwards to avoid over fitting (Gray and Fan, 2008). The pruning phase 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 complexity (Au et al., 2003). Pruning should be performed with care because it is possible that over pruning the tree could decrease the accuracy of the output rather than enhance it. The first CART tree was developed as an automatic interaction detection (AID) model in the early 1960's. CART is commonly used in industry because of the speed at which it can generate results. Once rules have been constructed they can be applied directly to a dataset using the database standard query language (SQL).

1.2.2. Logistic Regression

Logistic regression model (LRM) is an extension of multiple regressions. It provides an output that is in the form of a probability between the values 0 and 1 (Nefeslioglu et al., 2008). A description of a LRM is provided by Nannings et al. (2008) who state "One reason for the popularity of the LRM is the interpretation that is given to a covariate coefficient $_i$ in terms of an *odds ratio*. For an event with probability p its odds are $p/(1-p)$.

Regression is the study of dependence and is the central part of many research projects. It can answer questions about the dependence of a response variable on one or more predictor variables, including predictions of future values (Weisberg, 2005).

1.2.3. Artificial Neural Networks

Artificial neural networks consist of basic elements known as 'neurons'. These neurons consist of three main components: (i) weight, (ii) bias and (iii) activation function. Each neuron receives an input on which it applies a weight value. This weight holds the key to the neural network's overall performance because it provides the strength of the connection to the specific input. Each input is assigned a unique weighted value per neuron connection and all inputs are connected to all neurons in the first hidden layer. Each input is multiplied by its corresponding weight. These values are summed and a bias value that is a constant non-zero value is added. This summation is transformed using scalar-scalar functions known as the activation or transfer functions. These functions help to establish non-linearity into the NN, contributing to the immense power of this technique. A myriad of NN architectures have emerged over time, each identifiable by their topology (Cevik and Guzelbey, 2008).

1.2.4. Building a Model for Predicting Customer Churn

Experiments using each of the above mentioned predictive techniques have shown that predictive models alone do not provide a strong enough churn accuracy to make them directly applicable for the capture of future customer churn. These experiments are discussed in detail in chapter 4. The knowledge gained from these experiments is that directly applying a predictive model to the problem of capturing future customer churn result in low churn accuracy and high misclassification rates. The predictive models are good at identifying and capturing churn if the response variables are from the same month as the predictor data. Applying a predictive model directly to a dataset only provides a short timescale and response.

To perform churn analysis the predictor variables have to be of sufficient quality. If the predictive model is unable to converge with the data the output predictions will be inaccurate. It has been identified from literature that a good source of predictor data is from a customer's demographics and usage information. The problem with these data types is they could result in a model that does not conform to industrial data analysis regulations. This is discussed in more detail in chapter 4.

Recognising the importance to establish a maximum timeframe between initial churn detection and actual churn event, it would be logical to develop a methodology in such a way that it could be implemented into industry as a real-time monitoring system. Most available churn prediction models have to be applied in a batch processing scenario. This is discussed in greater detail in chapter 2. If the ability was provided to implement the model as a real-time monitoring system, this would maximise the time between churn detection and churn event. It could be possible to implement a predictive model that adjusts its prediction probability output as new data about a customer arrives.

This research takes the best identified predictive model and uses it as a foundation on which to build a comprehensive churn prediction methodology for the purpose of customer retention. It is demonstrated that a predictive model alone is not sufficient to address the complex problem of churn prediction, and stages are implemented on top of the predictive model to develop a novel and powerful methodology.

1.3. Problem Statement and Motivation

From the issues imposed through market saturation and cost implications as described in the first section of this chapter, there has been an identification of a need for a computer based churn prediction methodology that is capable of accurately identifying a loss of customer in advance, so that proactive retention strategies can be deployed in a bid to retain the customer. The churn prediction has to be accurate because retention strategies can be costly. A limitation of current research is that other studies have focussed almost exclusively on churn capture, neglecting the issue of misclassification of non-churn as churn. Retention campaigns commonly include making service based offers to customers in a bid to retain them. These offers can be costly, so offering them to customers who do not intend to churn can have a considerable impact on the total cost of a retention strategy. A further limitation of current research is that it is usually based on a single output in the form of 0 for non-churn and 1 for churn. This has been recognised as a limitation because it restricts analysis possibilities.

In order to address the problems mentioned above, a profile based analysis methodology is identified as a possible solution. It is anticipated that profile based analysis will enable future prediction, through the ability to match customers to profile clusters that are identified as most suitable for capturing future churn. It is also anticipated that profile analysis will provide a method for controlling misclassification levels through eliminating the profile clusters that statistically hold the smallest future churn capture accuracy.

Customer churn management can be viewed from two separate angles. One area of customer churn management focuses on churn prevention. This includes investigations such as competitor analysis, pricing and service strategies. The area being addressed by this research is churn prediction. In this area, it is common to use demographic and usage data as predictor variables where customer churn is the response variable. This research identifies an alternative approach to the use of demographic and usage data by showing how repairs and complaints data can be used successfully for the prediction of customer churn.

1.4. Sponsoring Organisations

There are two sponsoring bodies for this research, the Engineering and Physical Science Research Council (EPSCR) and British Telecom (BT). EPSRC is the UK governments leading funding agency for research and training. They invest around £740 million a year in subjects ranging from mathematics to materials science and information technology to structural engineering. Their mission is to support and promote high quality research and related postgraduate training in engineering and physical sciences. They also provide support in advancing knowledge and technology to contribute to the economic competitiveness of the United Kingdom and the quality of life (EPSRC, 2008).

BT is the leading telecommunications service provider in the UK, offering wide ranges of products to both business and domestic customers. The BT brand became established in 1980 as the official name of Post Office Telecommunications, becoming a state owned organisation independent of the post office. At this time BT was the UK's only telecommunications provider. This changed in 1982 when their monopoly was broken by Mercury Telecommunications who was also granted a licence to provide telecommunications services. Presently there are multiple telecommunications services in operation throughout the UK (Wikipedia, 2008a).

The proposed methodology targets three areas of the telecommunications industry, home telephone service, broadband, and mobile telephone service, through the validation case studies.

1.5. Thesis Layout

The layout of this thesis is developed based on the story of the research. This story aids in the identification of specific chapters (see Figure 1). A brief description of these chapters is provided below:

Chapter 1 provides the background to this research and briefly describes the context of this research. The aim and objectives are outlined and the problem statement and motivation for the research is defined.

Chapter 2 provides a survey of literature. This survey begins with defining the industrial problem in greater detail, and continues by discussing the research that has already been developed as an attempt to deal with the problem of customer churn. The literature moves into a discussion regarding customer satisfaction and loyalty. The stages for developing a churn prediction framework are identified. Predictive models are discussed in detail, separating the various available predictive techniques into two sections, traditional approaches and soft computing techniques. Validation methods are presented and finally a discussion on identified gaps in research is presented.

Chapter 3 gives a brief description of this research, outlining its aim, objectives and scope. It also discusses the methodology that is adopted for ensuring that the aim and objectives of this research are attained.

Chapter 4 presents the stages and iterations that aid in the development of a churn prediction methodology. It begins with an identification of the challenges for creating such a methodology and then reports the experiments that are carried out for the identification of the best predictive model. With the most suitable predictive model identified, the story of how the methodology is built and evolved throughout the research process is described. The complete methodology is presented in several stages.

Chapter 5 discusses the development of a churn prediction software prototype based on the research methodology. The chapter begins with a discussion on popular software development methodologies and documents the analysis used for identifying the most suitable methodology for aiding in the development of the churn prediction methodology. Through the application of the best methodology, numerous software releases are identified. The software releases are strategically created until the desired software prototype is complete. Software releases are presented in the chapter in the form of flow charts to provide a clear understanding of what has been developed.

Chapter 6 attempts to validate this research using 3 case studies. Due to the complexity of predicting customer churn quantitative and qualitative validation is applied to the churn prediction methodology. Quantitative Validation is performed in three stages. First the methodology is validated by applying it to each of the case studies to prove that i) the methodology can be successfully applied to varying data sources to show that the methodology is not data specific, and ii) a model can be successfully applied to varying data sources effectively and robustly. Secondly validation will use the generated model from the first stage of validation to determine advanced predictions of customer churn to illustrate the predictive power of the methodology, demonstrating that all identified objectives have been addressed and that the methodology achieves the identified research aim. Finally the methodology is compared against two other popular classification techniques to determine advancements provided by the methodology. Qualitative validation is applied through the deployment of a semi-structured questionnaire to provide expert opinion about the proposed methodology, and ensure that results obtained from the methodology exceed current standards.

Chapter 7 concludes this thesis with a discussion on the generality of this research, contribution to knowledge, and limitations of the research methodology. A discussion is provided on the proposed churn prediction methodology and limitations of the research are identified. Future research directions that could follow from this research are discusses and finally conclusions are presented.

A visualisation of the research and the layout of this thesis are presented in Figure 1 to provide the reader with a broader understanding of how this document is structured.

1.5.1. Thesis Structure Flow Chart

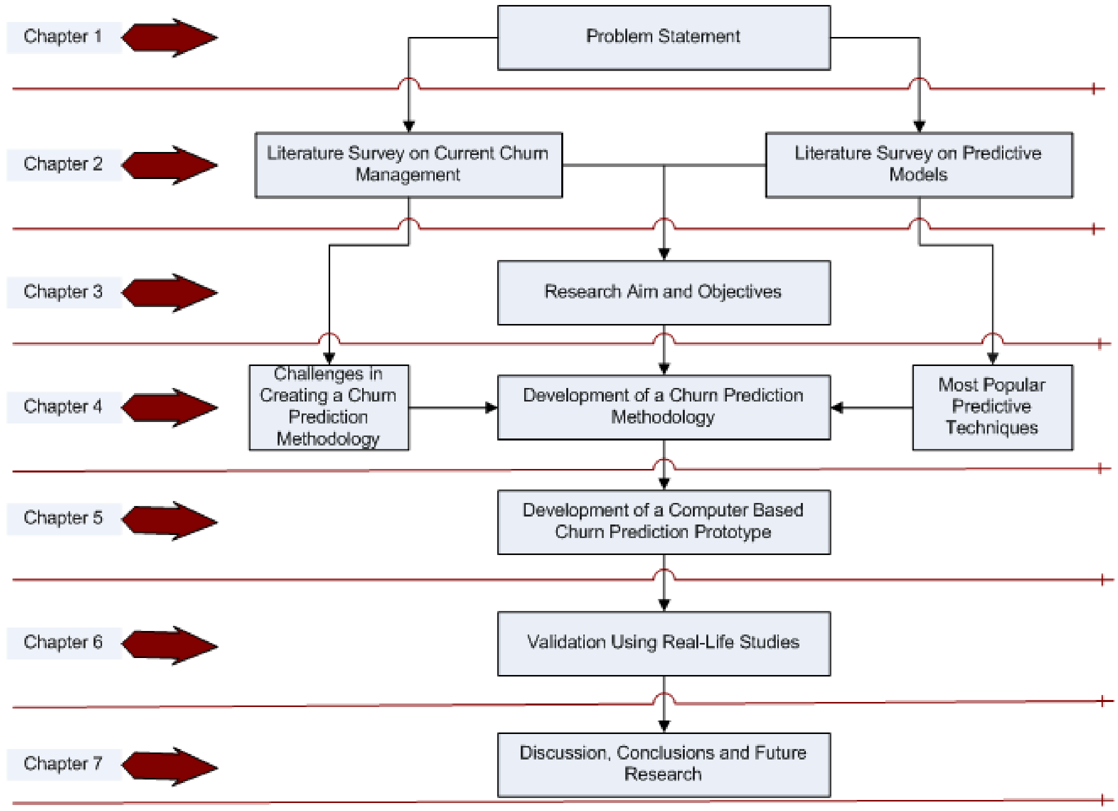


Figure 1: Thesis Structure

2.A Review of Literature

As markets have become increasingly saturated, companies have acknowledged that their business strategies need to focus on identifying those customers who are most likely to churn. In order to maximise business profits a company should focus on minimising customer churn. Typical churn for a service provider is usually around 4% each month. For the wireless telecommunications industry it has been suggested that customer churn cost the service sector around four billion dollars in 2006 (Chu et al., 2007). This statement is reinforced by Sweeney and Swait (2008) who state, "Customer churn is an ever-growing issue in the relational services sector (e.g., retail banking, telecommunications), where business models ultimately depend on long-term relationships with customers as the basis for profitability". This is supported by Eriksson and Vaghult (2000) who state, "customer retention leads to reduced sales and marketing costs compared to selling to new customers". This in turn explains the statement by Hidalgo et al. (2008) who claim, "a 1% improvement in the customer retention rate improves firm value by 5%".

Predictions of customer behaviour, customer value, customer satisfaction and customer loyalty are examples of some of the information that can be extracted from the data that is typically already stored within a company's database. To perform such an analysis it is necessary to either purchase commercial software or implement an 'in-house' solution from one of the many data mining techniques available. This chapter documents an investigation into current research, identifying data technologies and how these technologies have been used for the prediction of customer churn. Research is investigated on the most common techniques for predicting customer churn and an indication of some of the current trends is reported, along with directions for further research.

The majority of the study presented in this chapter has resulted in the publication of the review paper (Hadden et al., 2006c).

2.1. Customer Churn Management

It is becoming common knowledge in business, that retaining existing customers is the best core marketing strategy to survive in industry (Kim et al., 2004, Lariviere and Van Den poel, 2004). Retained customers generate more financial returns than new customers, which is why businesses should make every effort to retain their existing customer base, rather than investing valuable revenue in attempting to capture new subscribers (Buckinx and Van Den Poel, 2004). Seo et al. (2008) expand on this by claiming "in addition to the reduced costs, there is potential and opportunity value of customers which is gained over a long period of time. Because wireless telecommunications is not a onetime sale like commodity products, service providers can offer additional services over the length of a customer's tenure to generate more revenue". This is supported by Eriksson and Vaghult (2000) who state, "customer retention leads to increased sales and reduced marketing costs compared to selling to new customers". Ahn et al. (2006) support this claim by stating "with an increase in customer retention rates of just 5%, the average net present value of a customer increases by 35% for software companies and 95% for advertising agencies. Therefore, in order to be successful in the maturing market, the strategic focus of a company ought to shift from acquiring customers to retaining customers by reducing customer churn". Further to the increase of sales and profits that are generated by loyal customers it has also become apparent that when a customer churns from his/her current service provider cost are imposed on that service provider that are in most cases unrecoverable (Gans, 2000).

Churn is the term that has been adopted to define the movement of customers from one provider to another, and churn management is the process of the operator's efforts to retain those customers. These efforts usually involve the deployment of proactive retention campaigns in a bid to win the customers business before that business is lost (Hung et al., 2006). When the number of customers belonging to a specific service industry reaches its peak, finding and securing new customers becomes increasingly difficult and costly. At this point of the businesses lifecycle it should be higher priority to retain existing customers than trying to win new ones. It is also very difficult for a company to attempt to win new business. This is because the process of winning new business often involves offering attractive introductory packages. These packages are usually not available to existing customers which can cause animosity, resulting in decreased customer satisfaction levels. Further to this, existing customers might move to competitors to themselves take advantage of new customer offers, due to the fact that they cannot qualify for them through their existing service provider (Farquhar and Panther, 2008).

Churning customers can be divided into two main groups, voluntary churners and non-voluntary churners. Non-voluntary churn is the type of churn in which the service is purposely withdrawn by the company. There are several reasons why a company could revoke a customer's service. Reasons such as abuse of service and non-payment of service are usually the main causes. Voluntary churn is more difficult to determine. This type of churn occurs when a customer makes a conscious decision to terminate his/her service with the provider. This type of churn has been a serious and puzzling problem for service providers. The varied behaviour of consumers has baffled researchers and market practitioners alike (Berne et al., 2001). Voluntary churn can be divided into two sub categories, incidental churn and deliberate churn.

Incidental churn happens when changes in circumstances prevent the customer from further requiring the provided service. Examples of incidental churn include changes in the customer's financial circumstances so that the customer can no longer afford the service, or a move to a different geographical location where the company's service is unavailable. Incidental churn usually only explains a small percentage of a company's voluntary churn. This type of churn is also known as 'financial churn' (Burez and Van Den Poel, 2008).

Deliberate churn is the problem that most churn management solutions attempt to identify. This type of churn occurs when a customer decides to move his/her custom to a competing company due to reasons of dissatisfaction. Reasons that could lead to a customer deliberately churning within the telecommunications industry include technology-based reasons, when a customer discovers that a competitor is offering newer technology that their existing supplier does not provide. Economical reasons can also be a cause of deliberate churn such as finding the same product from a competitor at a better price. Examples of other reasons for deliberate churn include quality factors like poor coverage, bad experiences with call centres, and consistent faults with service, (Kim and Yoon, 2004). Continuing to use the telecommunications industry as an example, customers enter into a long term contract with a service provider, usually trusting that the service provider will always provide the promised service quality. If this quality is not maintained the customer is at risk of churning (Gerpott et al., 2000).

Deliberate churn within the telecommunications industry use to be minimised because switching would mean the requirement to change telephone number. In 2003 customers were given the option to switch mobile telephone provider but keep their existing phone number. In the US a law was passed in the November of 2003 bringing number porting into action, and as soon as this law came into force 12 million customers immediately churned from their service providers (Eshghi et al., 2007). This elimination of switching barriers has fuelled the retention battle (Turel and Serenko, 2006). Deliberate churn is also known by the term ‘commercial churn’ (Burez and Van Den Poel, 2008).

A churn management solution should not target the entire customer base because (i) not all customers are worth retaining, and (ii) customer retention costs money; attempting to retain customers that have no intention of churning is a waste of resources. Companies need to understand their customers. Liu and Shih (2004b) reinforce this statement by suggesting that intense competition is forcing organisations to develop novel marketing strategies to capture customer needs in an attempt to improve satisfaction and retention. Canning (1982) states that trying to sell more to everyone is no longer a profitable sales strategy and a market place that continually grows more competitive requires an approach that focuses on the most efficient use of sales resources. It is further reinforced by Knox (1998) who states that “the profit impact of customer retention has become accepted wisdom”.

A number of products exist for customer relationship management (CRM) which aims at analysing a company’s customer base. CRM is not a new concept, beginning in the mid 1990’s with IT based systems being developed to track multiple customer activities (Minami and Dawson, 2008). Today there are many commercial products available for the purpose of CRM with the amount of money being invested in this field exploding over recent years (Ang and Buttle, 2002). With so many products emerging organisations should take great care when deciding to purchase one of these solutions ‘off-the-shelf’. Chen and Popovich (2003) state that “CRM vendors might entice organisations with promises of all powerful applications. To date there is no 100% solution". The author agrees with the quote from Chen and Popovich (2003); however, it is believed that the main reason no 100% solution is available is due to the uncertainty and complexity involved in churn prediction, and the fact that churn can be a result of many varying factors. The main contributors to churn within the telecommunications service sector have been reported by Chu et al. (2007). Figure 2 illustrates the significance of each of these contributors:

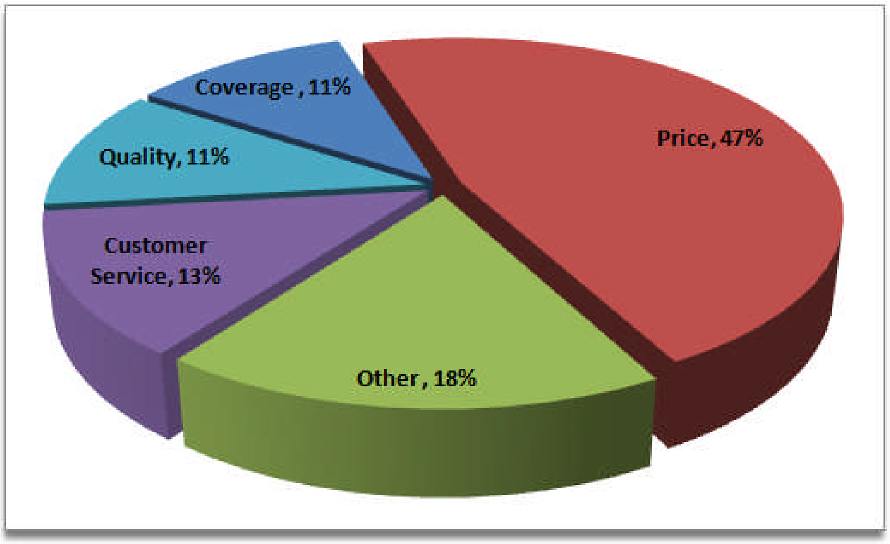


Figure 2: Main Churn Contributors For The Telecommunications Industry (Chu et al., 2007).

The main contributors to churn as illustrated in Figure 2 have a significant implication to researchers attempting to develop a model to capture it. According to these contributors, the main reason for customer churn is pricing issues. Pricing is responsible for 47% of defection. It is anticipated that pricing is the major contributor of what could be regarded as spontaneous churn. The customer has moved competitor through no real fault of the service provider but because he/she has found a similar service at a lower price. This means that any strategy will be at best 53% effective if a methodology targets the churn of the customers who defect due to reasons of dissatisfaction.

A further hurdle in the identification of customer churn comes from Ranaweera and Neeley (2003) who state, "there is evidence to show that in addition to satisfaction, other emotional responses such as inertia and indifference may also have an impact on retention". For researchers looking to develop a model to tackle the significant issue of defection through pricing issues, it has been suggested that loyalty point programmes were originally introduced for this reason (Cortiñas et al., 2008). Loyalty programmes can enhance customer retention, lifetime duration and product spending if applied effectively, although a great deal of research is still required in this area (Meyer-Waarden, 2006). It has been identified that although related to customer satisfaction and customer retention, investigations into loyalty programme schemes target a slightly different direction to the author's research and further investigation is deemed inappropriate. The author's research will focus on the identification of the 53% of customers who churn through reasons of dissatisfaction.

2.2.Customer Satisfaction

Referring again to the pie chart presented in Figure 2 it is shown that the cause of 35% of churn within the telecommunications industry is related to customer service, quality and coverage issues. These three categories can be linked to customer satisfaction, as customer satisfaction is a reflection of the customer's perceived service in relation to the service that he/she expects to receive. Hahm et al. (1997) state, "satisfaction is really a gap measure between performance and expectation". This statement is backed up by Kim et al. (2007) who state "a completely satisfied customer perceives their service to meet or exceed expectation".

Finding a suitable measurement of customer satisfaction is a major problem for organisations and has been a focus of research for quite some time (Yi, 1989). This statement is reinforced by Siskos et al. (1998) who state, "measuring customer satisfaction is a major problem for every firm or organisation, especially within the frame of marketing management practices. Satisfaction of customer needs is the main objective according to the principles of modern marketing science".

The importance of customer satisfaction has been accepted, and has received attention from researchers accross multiple service sectors. Some examples include determining customer satisfaction levels in the context of online retailing (Hsu, 2008, Souitaris and Balabanis, 2007), ecommerce (Liu et al., 2008), construction project management (CPM) which is a technically an oriented service for construction project clients (Yang and Peng, 2008), and customer satisfaction levels within the telecommunications industry (Eshghi et al., 2007).

Companies should be continuously improving their products and services to retain and enhance customer satisfaction; however customer satisfaction can suffer from other areas of business such as customer service, billing problems and faults. Seeking to maintain customer satisfaction through a product based approach alone is insufficient. Improving the range of products and enhancing services does not necessarily improve customer satisfaction levels, however poor product availability and bad service can have a high negative impact on customer satisfaction levels (Conklin et al., 2004). This suggests that although services and technologies need to be up-to-date from a strategic point of view, service errors are more damaging than service upgrades are rewarding, in terms of existing customer perceptions.

Customer loyalty is regarded in industry to be different to customer satisfaction. Loyal customers have been described as customers who show a psychological reaction and conviction to a specific product or service experience. Loyal customers are believed to possess a positive mental attitude towards their service provider, executing a continued repurchase conduct (Mankila, 2004). Customer loyalty is not a new concept in business. In fact it was realised that repeat purchasing from existing customers was crucial to business strategy as early back as 1942 (Sheth and Parvatiyar, 1995). Customer loyalty is more prevalent in service type businesses rather than product type businesses (Hidalgo et al., 2008). Bruhn and Grund (2000) have determined that in the case of the telecommunications industry satisfaction explains nearly 100% of customer loyalty (Eshghi et al., 2007). This is supported by Spiteri and Dion (2004) who state that the concept of customer value is related to, but different from, that of satisfaction The literature identifies two types of satisfactions: transactional and overall satisfaction (or cumulative satisfaction). Transactional satisfaction is defined as post choice evaluative judgment of a specific purchase occasion, whereas cumulative customer satisfaction is an overall evaluation based on the total experience.

The main benefits that are unique and common to loyal customers have been recognized by research performed by Rundle-Thiele (2005) who claim that repeat buying, immunity to competing offers and complaining behaviour are all traits of customer loyalty. The concept of customer loyalty has received multiple definitions and interpretations. Researchers holding a deterministic view, generally regard loyalty from an attitudinal perspective, while researchers holding a stochastic view tend to regard loyalty from a behavioural perspective. For example, word of mouth has been used in some cases as a dimension of loyalty by some researchers and an outcome of loyalty by others. Researchers have since taken the view of combining both loyalty views, to create a multidimensional view of loyalty, including a broad range of loyal states to benefit both the customer and the marketer. Examples of the dimensions of loyalty are as follows: -

Situational Loyalty - It has been regarded that loyalty could be the result of situations faced over time, and as a tendency for a person to exhibit similar behaviour. To elaborate, situational loyalty is the understanding that customers purchase products depending on the situation, such as purchasing a gift for an anniversary etc.

Resistance to competing offers - Occurring when customers are resistant to, or protected from other competing offers. One example of a customer that would be resistant would be a customer who is contracted to a supplier, making them unable to respond to competing offers. The relationship between resistance to competing offers and loyalty is still unclear. Again, some researchers regard resistance to competing offers as a dimension of loyalty, while others regard it as a consequence.

Propensity to be loyal - Measuring loyalty as a personal trait. Propensity to be loyal is regarded as an important measure for marketers, because it is regarded that it can sound the alarm for a decline in other loyal states.

Attitudinal Loyalty - Measures of attitudinal loyalty include preference, intention to repurchase, and commitment. Attitudinal loyalty is usually used for predicting behaviour. Word of mouth has been used by some research as a measure of attitudinal loyalty.

Complaining Behaviour – It may appear strange to include complaining behaviour as an aspect of customer loyalty, however researchers view complaining behaviour as ‘the customer using his/her voice’, and it is regarded that complaints can provide positive feedback to a company as well as negative.

Some researchers view complaining behaviour as a dimension of loyalty, while other researchers view it as a consequence of loyalty.

Athanassopoulos (2000) states, "Customer satisfaction is a multidimensional construct that is a direct result of the multiplicity and divergence of the contextual effects of customer expectations. According to Souitaris and Balabanis (2007), customer loyalty is a key driver of sustainable profitability and growth".

2.3.Churn Management Framework

A five stage model for developing a customer churn management framework has been identified (Datta et al., 2001). These stages are illustrated in Figure 3:

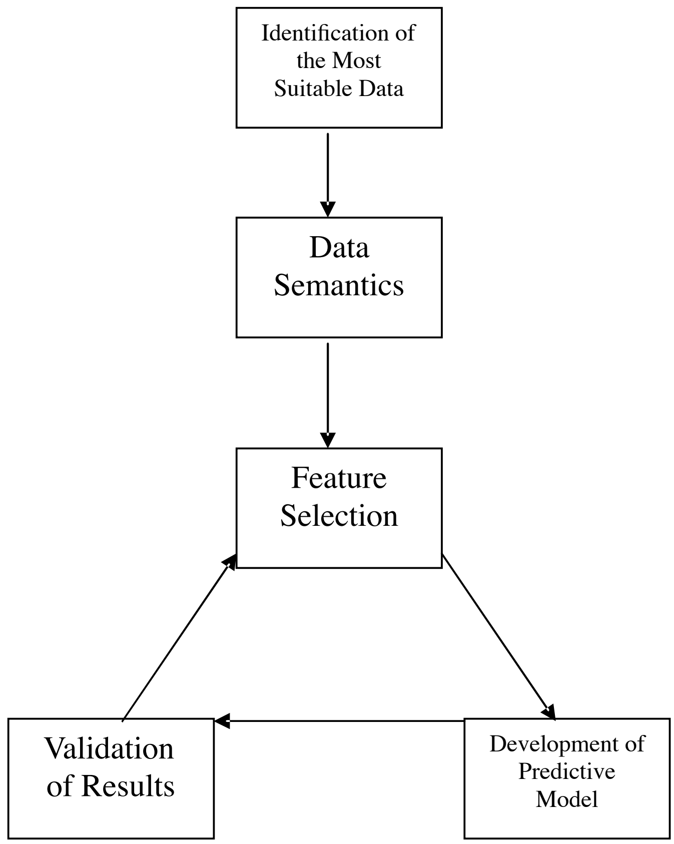


Figure 3: The Stages a Churn Management Framework (Datta et al., 2001).

Before any customer churn analysis can take place all stages of the model presented in Figure 3 should be considered. The quality of the data necessary for prediction is an important factor. Which data held within the data warehouse would provide greatest accuracy for predicting customer churn? How much of the data should be used; i.e. should all historical data be considered or should the input be based on only a few of the most recent months (Datta et al., 2001)?

Stage 2 is the problem of data semantics. Data semantics has been included in the model shown in Figure 3 because it has a direct relationship to stage 1, identification of the most suitable. In order to identify the most suitable there has to be a complete understanding of the data and the information each variable represents. Data quality is an important problem with many issues being directly related to data misinterpretation. Data semantics also covers representational heterogeneity and ontological heterogeneity. Representational heterogeneity understands the representation of variables. Similar variable names can have different values types associated with them. For example size can be stored in units cm, inches, feet, meters, etc and the date 1st April 2008 could be stored as 01/04/08 or 04/01/08 where both formats could be misinterpreted as being 4th January 2008. Ontological heterogeneity is where the meaning of the variable is understood but with slight incorrect variations. I.e. 1 month is represented as 31 days, 28 days, 4 calendar weeks, 4 working weeks or 3 financial weeks + 1 weeks forecast (Madnick and Zhu, 2006).

Stage 3 covers feature selection. A definition for feature selection has been taken from Chen et al. (2008) who state, "Feature selection is about finding useful (relevant) features to describe an application domain. Selecting relevant and enough features to effectively represent and index the given dataset is an important task to solve the classification and clustering problems intelligently".

Stage 4 is the development of a predictive model. Many models exist for determining the prediction of a desired event including statistical, classification and soft computing approaches. These methods are fully documented in section 2.4.

The final stage involves validating the model to ensure that it is achieving an accurate prediction. There are various preferred methods of validations and these are discussed in detail in section 2.7. All stages of the model presented in Figure 3 are discussed in detail in the following sections.

2.3.1. Techniques for Identification of the Most suitable

Figure 3 states that the ‘identification of the most suitable’, is the initial step in developing an effective customer churn management framework. Different combinations of data hold different analytical powers so it is necessary to identify the data that best suits the type of analysis to be performed.

Specific sets of data can provide better indicators for different problems and service sectors. For example, Ng and Liu (2001) suggest that usage data should be mined for identifying customer churn in the ISP (Internet Service Provider) service sector and telecommunications industry. Usage data has also been used for understanding the e-customer behaviour of website users (Jenamani et al., 2003), and predicting mail order repeat buying (Van Den Poel, 2003). Verhoef and Donkers (2001) state that the task of predicting repeat purchases of products and services is best performed using the customer’s historical purchasing data. This claim is supported by Hsieh (2004) who propose that the analysis of transaction data, along with account data and customer data could provide clues about the best incentives to the customers for a better marketing strategy. These examples reinforce the importance of the first stage (identification of the most suitable) of developing a churn management framework. The author understands that care, thought and research should be committed to this initial phase because the quality of the data will determine the power and accuracy of the overall model.

Customer data usually consists of many variables. There is a large amount of research that suggests recency, frequency and monetary (RFM) variables appear to be a good source for predicting customer behaviour (Liu and Shih, 2004b, Verhoef and Donkers, 2001, Hsieh, 2004, Liu and Shih, 2004a, Jonker et al., 2004, Van Den Poel, 2003). This statement is supported by Kitayama et al. (2002b) who state, “What is often given as an example of a measurement of the value of the customers is the RFM analysis approach”. Much can be obtained from an analysis of the customer’s value in terms of the customer’s satisfaction levels with a service provider, predictions through an analysis of spending patterns, the intention to churn if spending suddenly and rapidly decreases. The following definitions of the three types of RFM variables have been taken from Liu and Shih (2004a). Recency variables store information regarding the timeframe between purchases or use of service. A lower value suggests a higher probability of the customer making a repeat purchases. Frequency variables are those connected to how often the service is used. In general it can be assumed that the higher the frequency, the more satisfied the customer is with the service. However, there are other reasons for which a customer could be a heavy user such as personal or business reasons. A monetary variable for a customer would be the total sum of money a customer spends on his/her services over a certain time period. Those customers with high monetary values are the ones an organisation should be most interested in retaining. An example provided by Ryals, (2002) states “20% of a retail banks customers may account for more than 100% of its profits. At First Manhattan, a study showed that 20% of the banks households bring in about 60% of its revenues and more than 100% of its pre-tax income”.

Customer complaints data in conjunction with customer loyalty data has been used for the determination of customer satisfaction levels within the online retailing industry (Hsu, 2008). It has been advised by the sponsoring company that the UK telecommunications is regulated by Ofcom who have imposed monopoly regulations preventing any data being the source of analysis that could hinder fare competition and the successful growth of the telecommunications industry. Billing and demographic data analysis could breach these regulations through the detection and counteraction of fare competitor offers; therefore it has been advised that customer complaints and repairs data would be a more generic foundation to base model development.

2.3.2. Establishment of Data Semantics

Data semantics is the process of understanding the context of the data in a database. A definition provided by Ram (1995) describes a semantic model as “objects, relationships amongst objects, and properties of objects”. The fields and data that are stored in a database can usually be viewed as a collection of words. These words could seem self-explanatory and unambiguous; however, it is common for data to be difficult to interpret. For example, the data stored could be abbreviations to company specific terms, stored in a numerical format, or have a dissimilar meaning to that which could be considered obvious. For example there are more than 500 free bio-informatics databases available over the internet. These databases can be difficult to understand due to inconsistent data definitions. For example, according to one database the word "gene" is defined as "a DNA fragment that can be transcribed and translated into a protein". Another bio-informatics database describes a gene as "DNA region of biological interest with a name that carries a genetic trait or phenotype". Despite each claim being accurate the physical description is inconsistent (Volz et al., 2004).

There are various models available for the capture of the meaning and structure of data in a database (Ram and Khatri, 2005). Ram and Khatri (2005) state, "semantic models were developed to provide a precise and unambiguous representation of an organisation's information". The different types of tools that could be used for communication between the designer of a database and the end-users are as follows:

> *The entity relationship model* (ER)—the ER model represents data using entities, relationships, and attributes (Lee, 1999).
>
> *The relational model* (RM)—RM supports a data sub-language for data definition but can be seen as a complete database model supporting all aspects of data management (Apenyo, 1999).
>
> *The unifying semantic model* (USM)—USM is a formal specification for providing an accurate means of documentation and communication between users (Ram, 1995).

Despite claims by Datta et al. (2001) that data semantics is one of the most difficult phases of dealing with large data warehouses, an analysis of 100 other research papers has discovered that the phase has not been documented in most research. There could be several reasons for this, including data sensitivity resulting in researchers being unable to publish any examples based on the companies' data, or again due to sensitivity, researchers are forced to use what they have been given, and are not given the chance to explore a company's data warehouse for themselves.

Unforeseen circumstance could cause the fields of the data warehouse to change over time. Furthermore, the data warehouse is likely to use a range of regional databases as sources with each region having variances in the semantics of some of the stored fields. Depending on the methods to be used for churn analysis null fields can cause problems.

Data semantics is recognised as a continuous task, requiring close collaboration with the data semantic and model developers (Datta et al., 2001).

2.3.3. Feature Selection

Feature selection is the process of identifying the fields which are most suitable for predicting an event, described by Sun et al. (2004) as a critical process. It is an important stage because it helps with both data cleansing and data reduction by including the important features of a database and excluding the redundant, noisy and less informative ones (Yan et al., 2004). There are two main stages to feature selection. The first stage is a search strategy for the identification of the feature subsets and the second is an evaluation method for testing their integrity, based on some criteria.

2.3.3.1. Search Phase

The search phase of feature selection can be split into three categories, optimal, heuristic and randomised searches. The optimal search method is straight forward; however it is an exhaustive search method. With this process the number of possible subsets grows rapidly, making it unusable for even moderately sized feature sets. There are optimal search methods that avoid the exhaustive approach, using for example the branch and bound algorithm (Sun et al., 2004).

Two well-known heuristic methods are sequential forward selection (SFS) and sequential backward selection (SBS). SFS begins with an empty feature set and adds the best single feature to it. SBS works in the opposite way, by starting with the entire feature set, and at each step, drops the feature that least decreases the performance. SFS and SBS can be combined to create hybrid search methods. Two variations of these hybrid methods are known as sequential forward floating search (SFFS) and sequential backward floating search (SBFS). These methods fall into the general category '$l - r$' feature selection because first feature 'l' is added to enlarge the subset using the SFS method, and then feature 'r' is removed using SBS. The values of l and r are automatically and dynamically update within the method therefore no specific examples of the values can be offered. These methods face limitations when applied to globally optimal solution (Sun et al., 2004).

Randomised search uses probabilistic steps, or sampling techniques. One such method is called the relief algorithm. This method assigns weights to the features. The features with weights exceeding a user-defined threshold are selected to train the classifier. Another randomised search method, which has been attracting more and more popularity, is the genetic algorithm (GA) based approach. The research suggests that a GA can offer higher accuracy at the expense of greater computational effort. The main steps of feature selection using a GA are illustrated in
Figure 4 (Sun et al., 2004).

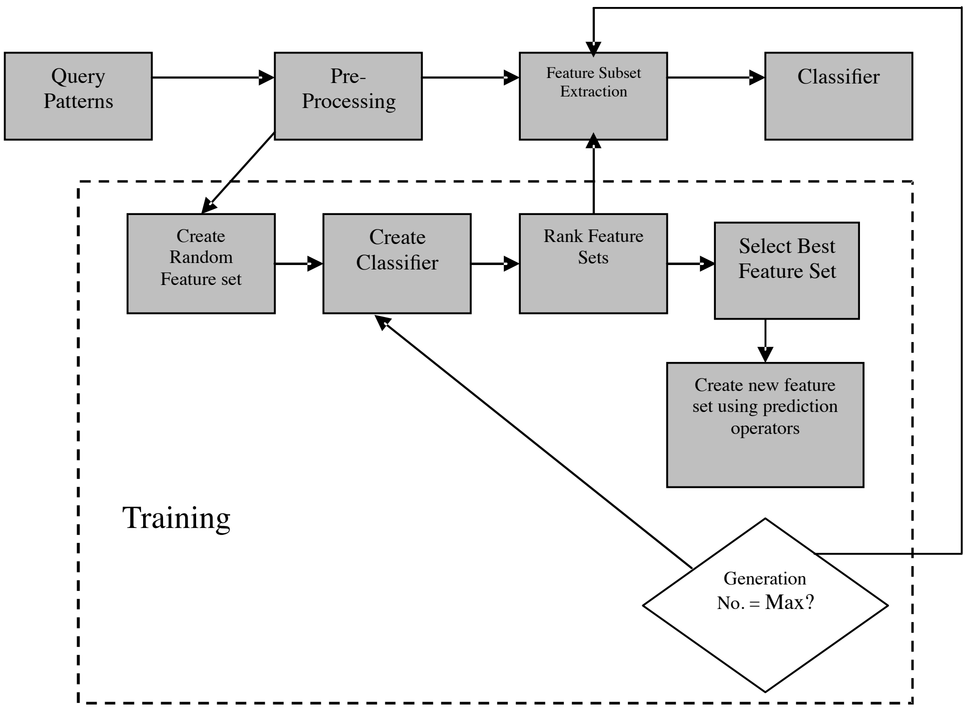


Figure 4: The Process of GA Based Feature Selection, Adapted From Sun et al. (2004).

This paragraph outlines other feature selection methods that are recommended by various researchers. Sun et al. (2004) suggest a GA search approach stating that a GA can provide a simple, general and powerful framework for selecting good subsets of features. Due to the characteristics of a GA, several feature subsets are created and tested. The best feature subset will then be evolved based on a fitness assignment. Yan et al. (2004) suggest a receiver operating characteristic (ROC) curve for feature selection, or more specifically, they state the area under the curve (AUC) can be used for feature selection. The AUC is measured using the Wilcoxon–Mann–Whitney statistic formula shown in Equation 1:

$$u = \frac{\sum_{i=0}^{m-1} \sum_{j=0}^{n-1} I(x_i, y_j)}{mn}$$

where

$$I(x_i, y_j) = \begin{cases} 1 & x_i > y_j \\ 0 & \text{Otherwise} \end{cases}$$

Equation 1: Wilcoxon-Mann-Whitney Statistic Formula.

The values [*x*0, *x*1, . . . , *xm-1*] are identified as the classifier output for positive samples, and the values [*y*0, *y*1, . . . , *y*n-1] are identified as the classifier output for negative samples. The samples are used to perform a pair wise comparison. The AUC as represented by the symbol *u* in equation 1, is then obtained for a linear classifier using only this single feature as the input. The feature is ranked using a larger value between the AUC and an alternative value obtained by replacing *I (xi, yj)* with *I (−xi ,−yj)*.

A method suggested by Datta et al. (2001) involves initially finding a subset of features from a data warehouse by manually selecting those that appear most suitable for the task. It has been suggested by Meyer-Base and Watzel (1998) that neural networks can be used for feature selection. Their research investigates the use of radial basis neural networks (belonging to the class of three-layered, feed-forward network); however, they found it necessary to include an additional layer to the traditional architecture in order to obtain a representation of relevant features.

Ng and Liu (2001) have performed feature selection by running an induction algorithm on the dataset. Datta et al. (2001) have developed a technique to predict churn for the cellular mobile service sector. A two-step process is used for feature selection. The first step uses forward feature selection (FFS). A decision tree sorts the features according to their error rate, which is done by using a single feature to predict churn. The first 30–50 features with the lowest error are then presented to the modelling system. The second step is then added to the existing generation. A GA is used to find groups of features that are more predictive than using single features alone. A GA uses logistic regression to evaluate each possible set of features. Some of the features that tend to be valuable for churn prediction with the mobile telephone service sector include *minutes of use*, *balance brought forward from previous bills* and *the tenure of the customer*. Datta et al. (2001) use a decision tree to sort the features according to their error rate. Although the decision tree method was the one preferred by Datta et al. (2001) they state that they also experimented with K-nearest-neighbour (KNN) and found no differences in the accuracy and performance of the models. Ng and Liu (2001) also recommend the decision tree approach for feature selection suggesting that this method is highly accurate, and has a record for good performance.

Once features have been extracted from a dataset they have to be validated. It is important to note that the data that the feature extraction was based on should be completely independent from the validation data; otherwise there is a risk of over fitting.

2.3.3.2. The Evaluation Phase

Evaluation strategies can be divided into two categories, the first is called filter and the second is called wrapper. A wrapper evaluation method is where the feature subset evaluation is performed using a learning algorithm that is incorporated in the classification design, while the filter approach uses feature subset evaluation external to the classification design. Filter approaches are more efficient than wrapper approaches because they evaluate the fitness of features using criteria that can be tested quickly. However these approaches could lead to non-optimal features, especially when the features are dependent on the classifier, leading to poor classifier performance (Sun et al., 2004, Kavzoglu and Mather, 2001).

2.4.Development of predictive model

A predictive model is defined as one that takes patterns that have been discovered in the database, to predict a future event (Rygielski et al., 2002). According to Crespo and Weber (2004) the most important predictive modelling techniques include decision trees and neural networks. The popularity of these technologies are reinforced by Baesens et al. (2004) who claim, neural networks and decision trees are typically traditional classification technologies. The following subsections provide an overview of both traditional and soft computing techniques used for predictive modelling.

2.4.1. Traditional Methods

This section covers the most common techniques that have been found in literature which have commonly been used for predictive analysis and data mining.

2.4.1.1. Decision Trees

The most popular type of predictive model is the decision tree. Decision trees have become an important knowledge structure, used for the classification of future events (Muata and Bryson, 2004a). Decision tree development usually consists of two phases, tree building and tree pruning. The tree-building phase 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. Certain branches may need to be removed because it is possible that they could consist of noisy data. The pruning phase 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 complexity (Au et al., 2003). Pruning should be regarded as a process of experimentation because it is possible that pruning the tree could decrease the accuracy of the output rather than enhance it.

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. These instances are known as pure nodes. Likewise, sub-groups containing occurrences from different classes are 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 (Au et al., 2003).

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 terminal, or leaf node (Bloemer et al., 2002).

Kitayama et al. (2002a) used a decision tree based approach to propose a model for customer profile analysis. The customer base was first segmented into groups of customers that were labelled as preferred and regular, the preferred customers being those most valuable to the company. The decision tree was then applied to the segments in order to determine the necessary measures to take for both the preferred and regular divisions, aiming to prevent customers from switching to alternative companies. Dividing a population of data and generating nodes using optional explanatory variables creates the decision tree.

Experiments performed by Hwang et al. (2004) involved a decision tree, a neural network and logistic regression. The decision tree showed slightly better accuracy over the other technologies; however, Hwang et al. (2004) state that these results do not prove decision trees to be the best choice in all cases. This is supported by Mozer et al. (2000). Ng and Liu (2001) suggest that for the purpose of customer retention, the feature selection process has to be accurate. For the purpose of identifying potential defectors they chose a C4.5 classification method. This is a decision tree induction method for classification because it has shown a proven good performance and it automatically generates classification rules.

Decision trees have been used for a variety of tasks. Table 1 demonstrates how classification rules can be applied to data to perform decision-making tasks.

Table 1: Example of Classification Rules (Ng and Liu, 2001)

Jobless	Bought	Sex	Age	Savings	Granted
No	Car	Male	38	$150K	Yes
Yes	Jewel	Female	26	$60K	Yes
Yes	Stereo	Male	20	$10K	No

Table 1 shows an example of how classification rules can be used in a decision-making process. Examples of possible rules that could be generated from Table 1 include:
If (Jobless = “No” and Bought = “Car” and Savings > 100k)

Granted = “yes”

If (Jobless = “Yes” and Bought = “Jewel” and Savings > 50k and Sex = “Female”)

Granted = “Yes”

If (Jobless = “Yes” and Bought = “Stereo” and Sex = “male” and Age < 25)

Granted = “No”

Many variations of rules can be generated from a set of attributes. However, in a real case scenario not all attributes will be valuable to the decision-making process. Those attributes that are influential are known as objective indicators. It should be noted that care must be taken when establishing the best classification rules because although a classification rule set can be highly accurate if correctly defined, poorly defined rules will result in a poor, unreliable system.

Classification rules appear to be common practice as a data mining technique. Liao and Chen (2004) use classification rules to perform customer segmentation, establishing the rules from a relational database. Datta et al. (2001) also 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.

2.4.1.2. Regression Analysis

Regression analysis is a popular technique used by the researchers dealing with predicting customer satisfaction. It provides a first step in model development. 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) use a binomial logit model to determine subscriber churn in the telecommunications industry, based on discrete choice theory. Discrete choice theory is the study of behaviour in situations where decision makers must select from a finite set of alternatives. According to Au et al. (2003) regression analysis is fine for determining a probability for prediction; however it is unable to explicitly express the hidden patterns in a symbolic and easily understandable form.

Hwang et al. (2004) discovered that logistic regression performed best for predicting customer churn when compared with neural networks and decision tree. It should be noted that Hwang et al. (2004) were investigating a prediction of the customer lifetime value (CLV), with the intent of including customer churn; they suggest that logistic regression was the best model for their purpose. The authors believe that many factors could influence these results such as the neural network parameters chosen and the data that the experiment was based on. The data used for experimentation may have been more suited to a logistic regression model than that of a neural network or decision tree.

Datta et al. (2001) used simple regression to initially predict churn but later experimented with KNN, decision trees and neural networks. The overall model used to develop the churn prediction platform was done using a neural network. Their research could not establish a best method. They have stated future directions as including an explanation of customer behaviour because their model was unable to predict customer churn accurately. They also suggest that information stored externally to the organisations' database should be included, such as the state of the telecommunications market and current competing offers etc. The model suggested by Datta et al. (2001) fails to distinguish between loyal customers, valuable customers and less profitable customers. They suggest that future research should include a more financial orientated approach by optimising payoff. They further suggest that by concentrating on payoff rather than churn the developed model would weight those customers bringing in higher profits over those bringing in lesser profits.

Baesens et al. (2004) used Bayesian network classifiers for identifying the slope of the customer lifetime value (CLV) for long-life customers, but used simple linear regression on the historical contributions of each customer to capture their individual lifecycles. The slope was then separated into either positive or negative classes to represent increased or decreased spending. This variable was then used as the dependent variable in their study. The CLV was also the focus of research by Rosset et al. (2003) who used the Kaplan-Meier estimator to estimate the value, and Stahl et al. (2003) who link CLV to company shareholder value.

2.4.1.3. Other Traditional Methods

Other techniques dealing with predicting customer behaviour include a semi-Markov process as used by Jenamani et al. (2003). They 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. A semi-Markov process was also used by Slotnick and Sobel (2005). Prinzie and Van Den Poel (2004) introduce a mixture transition distribution (MTD) to investigate purchase-sequence patterns. The MTD is designed to allow estimations of high order Markov chains, providing a smaller transition matrix facilitating managerial interpretation. Markov chains were also used by Ma et al. (2008) for relationship marketing to obtain an estimation of the customer's lifetime value (CLV).

Auh and Johnson (2004) use five firm-level variables named as customer satisfaction index scores (SAT), customer loyalty index scores (LOY), year of data (YEAR), relative quality importance (RQI) and the average ease of comparing quality differences (EQ), to create a general linear model. Prinzie and Van Den Poel (2004) do not offer a precise explanation of how these variables were selected, the authors have determined the following definitions from the literature. SAT has been defined as a cumulative evaluation of a customer's purchase and consumption habits. YEAR represents the year that the data is relevant. RQI determines the relative impact that the perceived quality and value has on customer satisfaction. Estimations were established using a model based on quality-versus-price as defined by the American customer satisfaction index model. EQ is defined as a measure of the ease of judging and comparing quality. It is the average of the response collected by asking the following question: "Thinking about the quality of the product (e.g. mobile telephones) do you consider it easy or difficult to judge what is high versus low quality?" This question would have a score ranging between 1 and 10. A score of 1 would represent very easy while a score of 10 would represent very difficult. LOY is an indication of the customer's intentions to repurchase services or goods. Regression was used as an alternative to test for linear equality restrictions.

Chiang et al. (2003) introduce their own algorithm for identifying potential churners, which they have named 'goal oriented sequential pattern'. This work uses association rules, defined as a technique that identifies relationships amongst variables. A simple example given by their research suggests that people who purchase milk at a supermarket will also buy bread at the same time. The paper defines two steps for finding out association rules. The first step is to detect the large itemset and the second is to establish the association rules by exploiting the large itemset.

For the second step, an *a priori* algorithm is used for the exploration of association rules. The *a priori* algorithm determines rules using a straightforward sequential process to determine relationships in the database.

Tao and Yeh (2003) document two database marketing tools named USC (Usage Segment Code) and NRE (Net Revenue Equation) for the purpose of understanding customer behaviour. It is claimed that customer retention is one of the non-marketing activities that these two tools can be used for. These tools are not used for predicting potential defectors, but to help in making a decision on the marketing strategies to be used if a customer calls the company wishing to cancel the subscription.

Madden and Savage (1999) documents a churn model for retaining customers in the Australian ISP (Internet Service Provider) industry. It could be argued that this work is no longer relevant because it was carried out when ISPs charged a monthly fee for a certain amount of online dial-up time. The development of broadband and flat-rate dial-up offers could see this early work as obsolete. However it is still of interest to examine the techniques used for identifying customer churn and the variables used. The variables were grouped into four categories, economic variables, usage variables, ISP choice variables and demographic variables. It is stated in the research that the demographic variables were the ones that had been found the most important in other research. As stated by Madden and Savage (1999) "a binominal probit model was used to relate to the probability of a subscriber leaving their current ISP".

Support vector machine (SVM) is another technology that is worth investigating for its suitability for use with customer churn management; however a thorough search of the literature has revealed little investigation into this method. The closest research to CRM identified by the author is the mining of customer credit scores (Lee et al., 2006). This research used regression trees as a final model; however SVM was also investigated. The paper states that the SVM approach has emerged as a new and promising approach for the process of data classification. The technology is described as methodical and inspired by statistical learning theory. The experiments performed in the research took advantage of Matlab's SVM toolbox. The experiments showed similar results to neural networks. It is documented that on a standard PC, it took roughly 20 hours to analyse a dataset of 8000 records, suggesting a major drawback of the technology (Lee et al., 2006).

Naïve Bayes is another classification technique which has been over-looked for its potential suitability for customer churn management. This is surprising since the Naïve Bayes is a popular technique for text categorisation. The Naïve Bayes consists of a training dataset and a set of possible features. For each category few features are chosen which contain the largest information with respect to that category. All the features are extracted from new information to determine the probability of which category the new information belongs to. Drawbacks of the Naïve Bayes approach are: the first approximation only considers the features that are present in the new data, and second approximation assumes that the presence of features is independent (consequently this is where the Naïve in Naïve Bayes comes in) (Keren, 2003).

The final technology that should be mentioned but has received very little attention within the CRM research community is k-nearest neighbour (KNN). KNN is another classification technique that is widely used within the text categorisation community. To classify unknown information, the KNN ranks the training information using class labels of *k* most similar neighbours to predict the class of the information (Tan, 2006).

2.4.2. Soft Computing

Soft computing is a consortium of methodologies (such as fuzzy logic, neural networks, and genetic algorithms) that work synergistically and provides, in one form or another, flexible information processing capabilities for handling case 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 (Pal and Ghosh, 2004). Technologies that fall in the category of soft computing are evolutionary computation (EC), artificial neural networks (NN), fuzzy logic (FL), probabilistic computing and their combinations, for example, neuro-fuzzy systems.

Evolutionary computing consists of various computational techniques that have been inspired and developed on the evolution theory as proposed by Charles Darwin. He was one of the leading intellectuals of 18 th century England who proposed in his publication 'On the Origin of Species by Means of Natural Selection', that evolution was an inevitable process through the mechanism of natural selection and survival of the fittest. These theories of natural selection and survival of the fittest have been computationally mimicked to effectively solve real life problems in the forms of such techniques as genetic programming (GP) and genetic algorithms (GA) (Tan et al., 2005).

GA is an evolutionary computing technique that is used to solve optimisation problems. An initial population made up of as chromosomes (solutions) is randomly generated. Each part of the chromosome is known as a gene and the chromosomes are used to spawn new generation through crossover and mutation operators. The fitness of each solution (chromosome) is evaluated against a specific fitness function. The fittest of these chromosomes are selected to become the parent solutions and the whole process repeats. This process can run over hundreds or thousands of generations depending on the complexity of the problem, and eventually the near optimum solution is reached (Do et al., 2008). A flow chart clearly illustrating the components of a GA is presented in Figure 5:

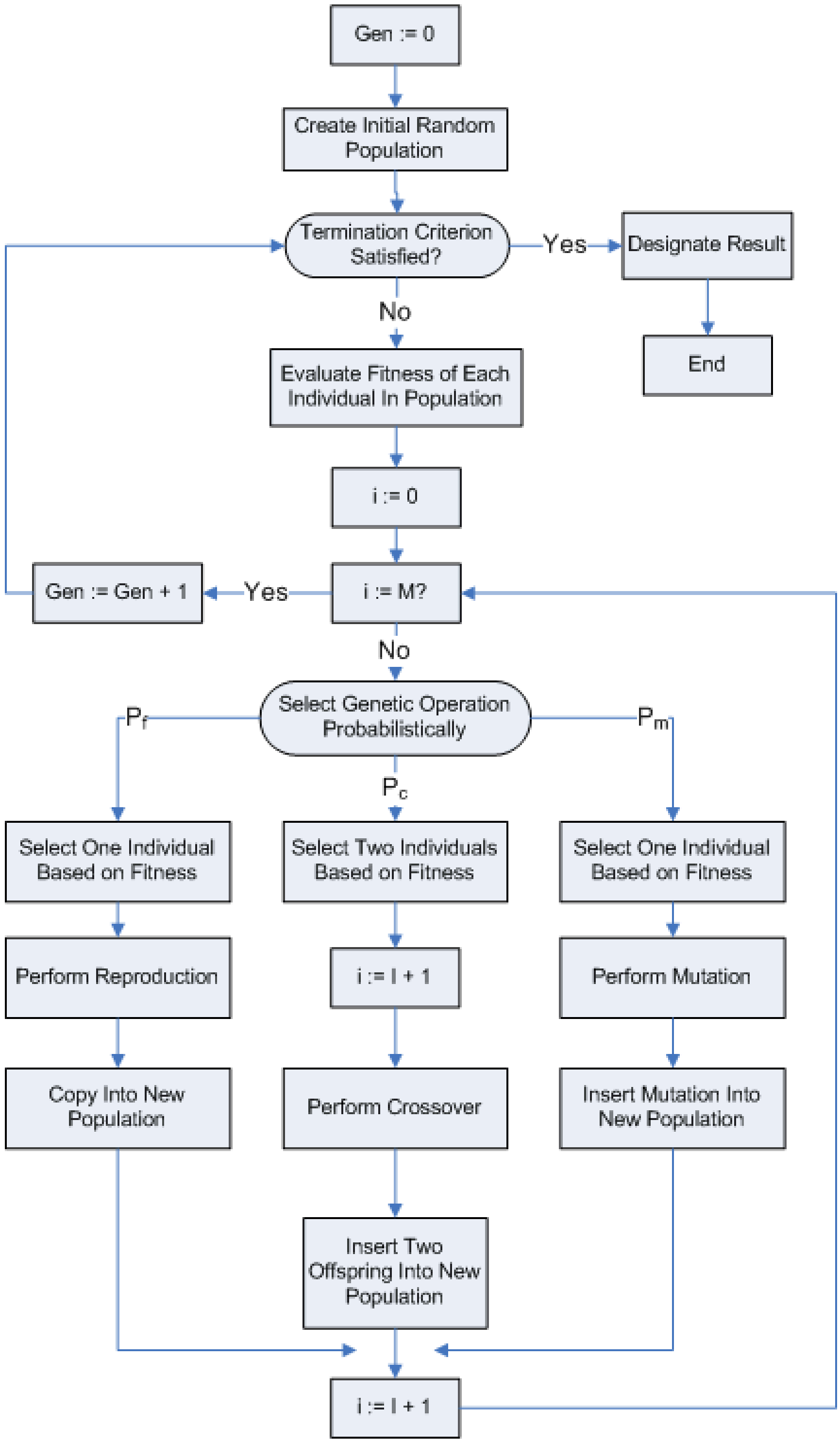


Figure 5: Flow Chart of The Conventional GA (Koza, 1992)

The flow chart presented in Figure 5 does not explicitly show the creation of a mating pool. This is because the diagram has been simplified for presentation. The mating pool stage has been replaced with the selection of 1 or 2 individuals on the basis of fitness. Four stages have been identified that should be addressed when preparing to use a conventional GA on fixed length character strings (Koza, 1992):

1. Determining the representational scheme.
2. Determining the fitness measure.
3. Determining the parameters and variables for controlling the algorithm
4. Determining the way of designating the result and the criterion for terminating a run.

GP is a technique most well known for possessing the capability of generating mathematical equations that can be used as models, where the functional form and numerical coefficients are found through an evolutionary process (Do et al., 2008). It is a generalisation of the GA (Aytek and Kisi, 2008). GP incorporates many of the conventional GA ideas to structures that are more complex than character string patterns. In particular GP operates with very general, hierarchical computer programs. GP can be implemented using any computer programming language that is capable of manipulating computer programs as data and can further compile, link and execute the new programmes. In GP populations of hundreds or thousands of computer programs are genetically bred using Darwin's principle of survival of the fittest, as described for the GA. In conjunction with a genetic recombination (crossover) operation appropriate for mating computer programmes (Koza, 1992).

GP begins with an initial population of randomly generated computer programmes. The programmes consist of functions and terminals specific to the problem. These programmes can consist of standard arithmetic operations, Standard mathematical functions, logical functions or standard programming operations. A typical GP architecture is shown in Figure 6:

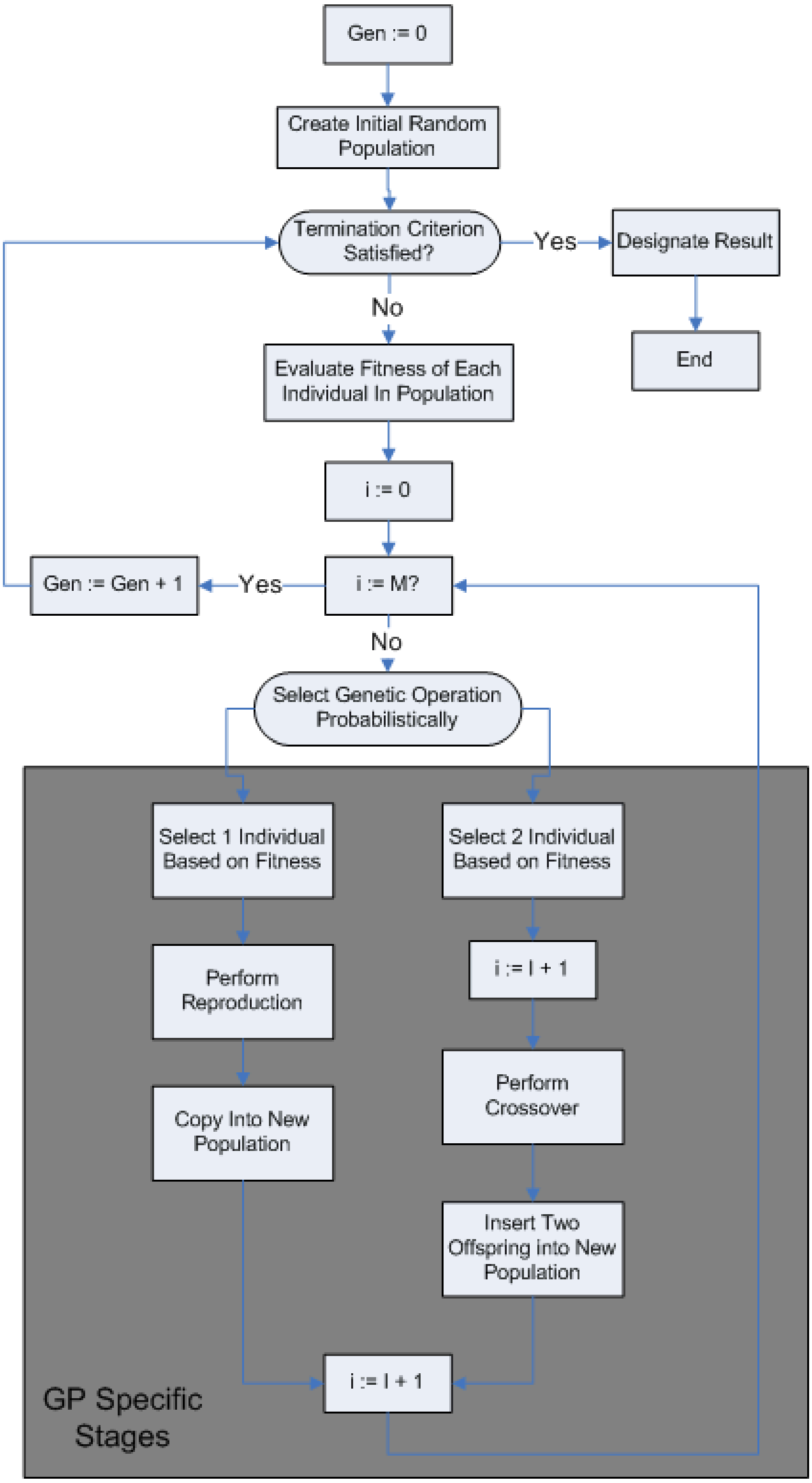


Figure 6: Flow Chart for the GP Paradigm (Koza, 1992)

The GP paradigm as shown in Figure 6 breeds computer programmes to solve problems by executing the following steps (Koza, 1992):

1. Generate an initial population of random compositions of the functions and terminals of the problem (computer programs).
2. Iteratively perform the following sub-steps until the termination criterion has been satisfied:
    a. Execute each program in the population and assign it a fitness value according to how well it solves the problem.
    b. Create a new population of computer programs by applying the following two primary operations. The operations are applied to computer program(s) in the population chosen with a probability based on fitness.
        i. Copy existing computer programs to the new population.
        ii. Create new computer programs by genetically recombining randomly chosen parts of two existing programs.

Investigations by Au et al. (2003) have discovered that two GA approaches have been developed for rule discovery. These approaches are known as the Pittsburgh approach and the Michigan approach. The difference between these approaches is that the Michigan approach represents a rule set for the entire population, while the Pittsburgh approach represents a rule set for an individual chromosome. The research then states, "although GA based rule discovery can produce accurate predictive models, they cannot determine the likelihood associated with their predictions, preventing these techniques from being applicable to the task of predicting churn".

Au et al. (2003) proposed a new algorithm called 'data mining by evolutionary learning' (DMEL). DMEL uses non-random initial population based on first order rules. Higher order rules are then obtained iteratively using a GA type process. When identifying interesting rules, DMEL uses a measurement of the object which does not require user interaction. 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 intended to predict churn within the telecommunications industry. In respect to customer retention and customer churn management little documentation can be found for GP.

2.5. Neural Network Development

Artificial neural networks are loosely mathematically based on what is known about physical, biological cognition and learning. Although much is understood regarding biological systems, biological brains are morphologically and chemically extremely complex, with invasive studies ethically prohibited. Because of this, huge portion of information are missing from the mystery of the human mind. Therefore even the most complex neural mathematical models should not be regarded as accurate, however can be viewed as providing the smallest number of conflicts with respect to what is currently known (Wythoff, 1993).

Artificial neural networks have been defined as by Cevik and Guzelbey (2008) as "a massively parallel distributed processor that has a natural propensity for storing experiential knowledge and making it available for use. The main component of this model is the structure of its information processing unit".

Artificial neural networks have been successfully used to estimate intricate non-linear functions. A neural network is an analogous data processing structure that possesses the ability to learn. The concept is loosely based on a biological brain and has successfully been applied to many types of problems, such as classification, control, and prediction (Behara et al., 2002).

Neural networks are different from decision trees and other classification techniques because they can provide a prediction with its likelihood. Various neural network approaches have emerged over time, each with varying advantages and disadvantages (Liao et al., 2004), however greater detail into these variances is beyond the scope of this thesis. Research suggests that neural networks outperform decision trees and logit regression for churn prediction (Au et al., 2003). According to Rygielski et al. (2002) neural networks provide a more powerful and accurate means of prediction. However there is a potential risk of finding sub-optimal solutions and over fitting when compared with a decision tree. Another important factor to be aware of when considering the use of neural networks is that they do not uncover patterns in an easily understandable form (Au et al., 2003).

Rygielski et al. (2002) discuss neural networks as data mining technique for CRM. According to their work, neural networks provide a more powerful and predictive model than other techniques. They are also documented to be applicable to a wider area of applications. However other disadvantages should be taken into account, such as clarity of output, implementation and construction of the model.

Boone and Roehm (2002) have applied neural networks to segmentation of customer databases in the retail service sector. Vellido et al. (1999) use neural networks to segment the online shopping market; more specifically they use a self organising map (SOM) which is an unsupervised neural network. A SOM has also been analysed by Shin and Sohn (2004) for segmenting stock trading customers according to potential value. An investigation by Datta et al. (2001) revealed that neural networks were only being used by a few companies. They state that a possible reason for this could be the lack of clarity of output.

Baesens et al. (2004) attempt to estimate whether a new customer would increase or decrease future spending. The paper recognises this problem as a classification task and proposes a Bayesian network for the prediction. A Bayesian network is defined as a probabilistic white box which represents a joint probability distribution over a set of discrete stochastic variables.

2.6.Overview of Techniques for Predictive Modelling

Many of the technologies mentioned in this section are examples of data mining. The most popular being decision trees, primarily because they uncover classification rules for classifying records in the form of *if-then* rules, which are easy to understand. Soft computing techniques do not provide this easy to understand classification and do not clearly convey the underlying patterns in an easily understandable form.

This section provides a discussion on the various predictive modelling techniques identified from literature. Much work has been done in the area of CRM and predicting customer behaviour, although it appears that to date, customer churn management has not received the attention that it requires. Figure 7 shows a chart of the known predictive and segmentation methods, compiled by Verhoef and Donkers (2001). A chart illustrating the most popular research techniques as identified from the literature survey reported in this thesis is presented in Figure 8. A comparison of these charts provides an understanding of how research in predictive analysis as altered over time.

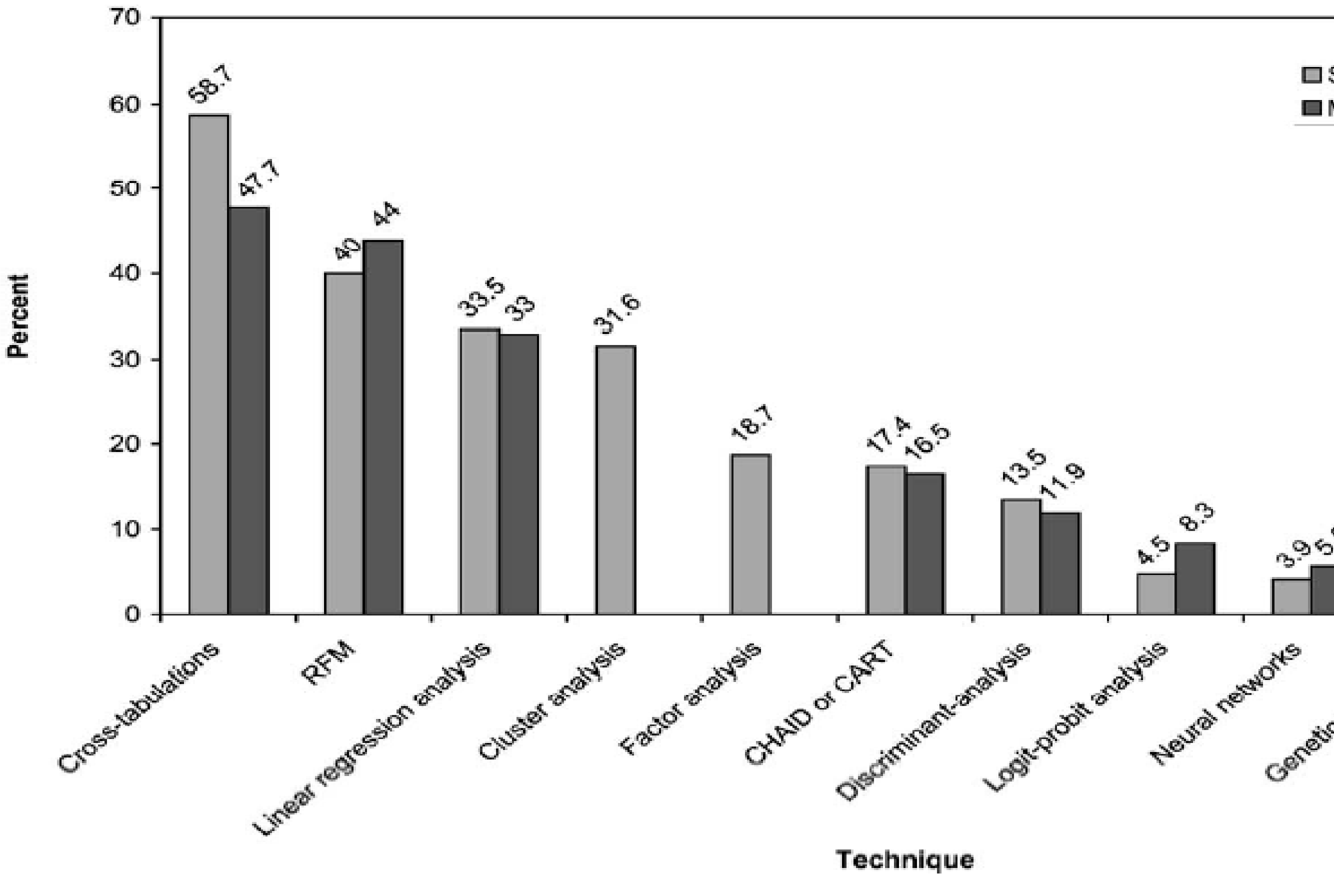


Figure 7: Predictive Modelling and Segmentation techniques (Verhoef, 2001)

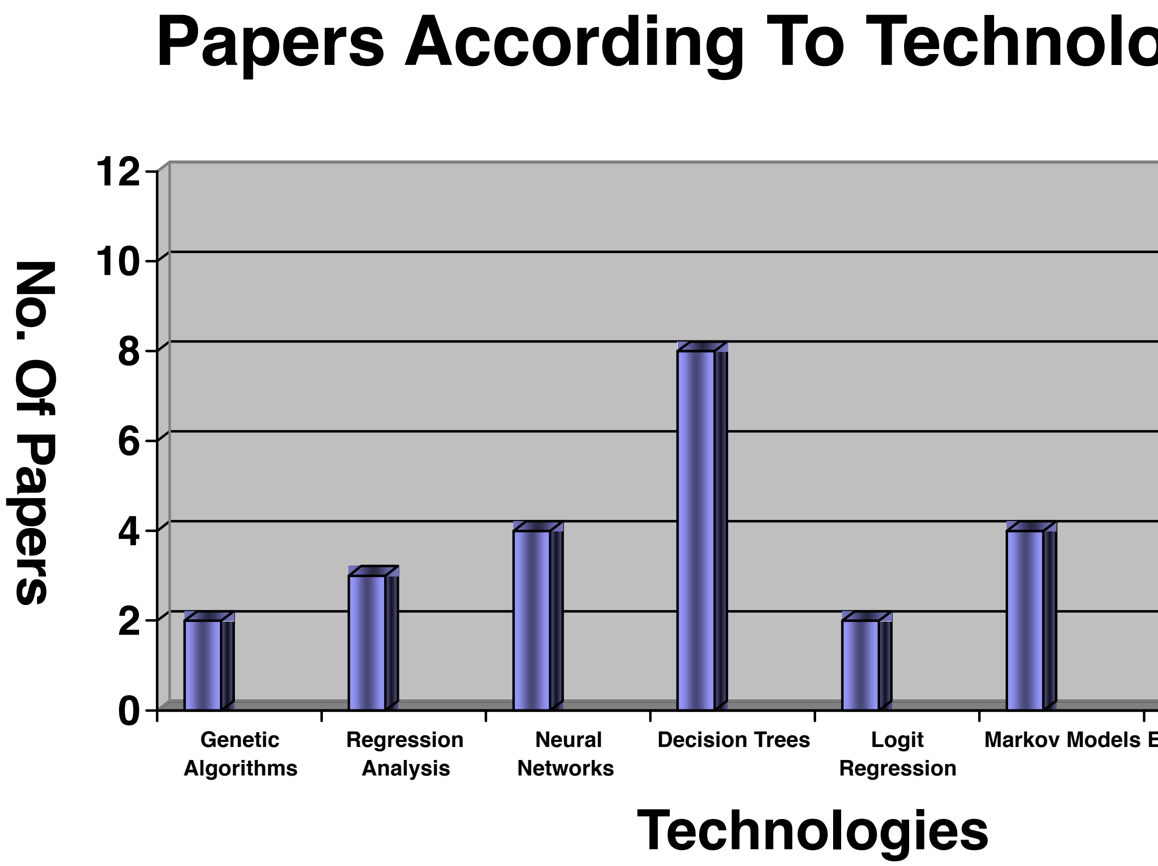


Figure 8: An Analysis of Technology Used in Identified Customer Retention Focused Researc

It can be observed from Figure 7 that cross-tabulation has been the most popular segmentation and predictive modelling technique in the past. The next column shows research, frequency and monetary variables (RFM). The authors argue that RFM defines the variables for use by segmentation and modelling techniques rather than being a technology of its own. Methods such as weighting RFM variables have been reported in the literature by Liu and Shih (2004a), and it is agreed that weighting the variables could provide better results. This is the type of modelling that is classified as RFM by Verhoef and Donkers (2001).

Linear regression has been another popular technique for performing both prediction and modelling, while cluster analysis and factor analysis have been shown specifically for segmentation. Interestingly, according to Verhoef and Donkers (2001), CART, logit analysis and neural networks, the three powerful technologies for prediction, have been the least used methods according to literature. Out of the three however, they report that CART has been preferred over neural networks and logit regression.

There are clear differences between Figure 7 and Figure 8 figures in the result of the research conducted by the authors that have focused mainly on publications during the last five years. Therefore, it updates the finding of Verhoef and Donkers (2001). Cross-tabulation is an older modelling technique which has been overtaken by more advanced and accurate methods. Furthermore, the differences between Figure 7 and Figure 8 are also because the focus of Figure 7 is on segmentation and predictive techniques in general, and the focus of Figure 8 is on technologies for predicting customer behaviour. One of the major differences is that regression analysis has taken the lead for the preferred method, followed by decision trees and Markov models, and then neural networks. This means that four of the least used methods reported by Figure 7 have become four of the most popular as reported by Figure 8.

2.7. Validation Methods

Validation is extremely important for data mining models as it is the only real way to ensure that the model is not simply remembering each data instance that was used in training (Bellazzi and Zupan, 2008).

There are several methods documented for validating a customer churn model. Some popular methods are discussed below:

- Cross-fold validation: 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. Cross-fold validation is most suitable in those cases in which there is a scarcity of data.
- Several cross-validation methods have been proposed in the literature, three examples follow (Dudoit and Van Der Laan, 2005):

  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 for training and validation.

  Using a separate validation dataset: Datta et al. (2001) validated their model by comparing their results against a simple regression model, decision trees and KNN classifiers. An independent validation dataset, containing 17,000 records, was used for validation. Bloemer et al. (2002) used an independent set of data that had no connection to the training or testing sets to validate their model, Prinzie and Van Den Poel (2004) also used an independent validation dataset to validate their work. This method of validation is most suitable in those cases where data availability is not an issue.
- Empirical Study: In some cases the only real way to effectively validate is to reproduce the target working environment and monitor results over a sequence of time. This form of validation was performed by Hsu and Chien, (2007).

2.8. Research in the Area of Customer Retention

This section begins with looking at what research has been done in the area of customer churn and retention within the telecommunications industry and then gradually broadens this investigation by looking at the same subject across alternative service sectors. This should help to build a clear picture on the advancements and limitations of current research with the aim of establishing the necessary steps in creating an enhanced churn prediction methodology.

There has been a reasonable amount of research into constructing a methodology for the prediction of customer churn within the telecommunications industry. The telecommunications industry in particular is challenging because of the multiplicity of services and variances of customer type. Hahm et al. (1997) have approached the problem by constructing a decision support system which makes use of customer satisfaction data. They give customer satisfaction a similar definition to that provided in section 2.2. Their research involves segmenting customers into certain groups by applying a four step procedure. The first step documented is the process of segmenting customer variables according to their relevance for particular services, e.g. a conventional telephone service has demographic variables such as age, sex, occupation and education level and also behavioural variables such as usage data and benefits sought attached. Step 2 involves determining the search space. This stage is basically deciding how each variable should be limited. E.g. Location could be limited to city/street and ages could be grouped into ranges such as 18-25, 26-35, etc. Step 3 takes the information identified from step 2 a little further and identifies which variables are most critical for segmentation of the market. Step 4 is the process of understanding the correlation between the critically identified variables so that a matrix showing the characteristics of each segment can be constructed. Correlation analysis is performed on the identified segments and measured over a period of time. It identifies the determinants of overall satisfaction scores.

Research performed by Hung et al. (2006) also attempts to predict customer churn within the telecommunications industry. Their research focuses on the wireless telecommunications industry in Taiwan. Because of the geographical area where their research is based they are not tied to the same restraints as the author and have been able to base their research on usage and demographic data. They compare back propagation neural networks and C5.0 decision trees to see how each technology performs for the prediction of customer churn. K-means clustering is also used to segment customers into groups according to their billing amount. These groups are used to approximate customer value. A predictive model is created for each cluster using both decision trees and neural networks. They report that neural networks provided a slightly better performance on the segmented clusters but much better performance on the full un-segmented data. They have been clever in reporting their results, basing results on the top 10% of customers with the lowest loyalty index scores for each cluster. Actual churn capture is compared against misclassifications for this 10% cluster only. These customers are reported to provide an over accuracy of 90% churn capture accuracy.

Using customer transaction and billing data, Ahn et al. (2006) investigate the determinants of customer churn within the telecommunications industry based on large scale customer transaction and billing data using logistic regression models. No real churn prediction results are provided for this research; however the main factors leading to churn have been identified. They find that the main contributor of customer churn within the wireless telecommunications industry is dropped calls, i.e. calls that are disconnected mid conversation due to poor signal coverage.

Coussement and Van Den Poel (2008) have also focused their research on making an advanced prediction of customer churn; however their service area is in newspaper subscription services. The techniques that have been the main focus of their study are grid search and cross validation. The most important variables in their research are length of subscription, length of time since last renewal and month of contract expiration showing that their research was mainly based on billing and subscription information. Like Hung et al. (2006) they measure the accuracy of the top 10% of customers most likely to churn. Within this group they claim a 95% confidence in their churn prediction model however they do not state how far into the future (if at all) these customers actually churned after being identified as likely defectors. Another paper discusses churn capture within the financial service sector (banking and insurance). Again the model is built on customer demographics and purchase/billing data. The study mainly assesses the relationships of the variables in relation to churn using the Kaplan-Meier estimate of the survival curve (Van Den Poel and Lariviere, 2003).

Seo et al. (2008) developed a two-level model of customer retention for the US wireless telecommunications market. This study was based on demographics data. More specifically they had access to a dataset containing the information for 31,769 customers from which 46% where churners. The main variables used in their analysis were ‘gender’ and ‘age’. The model first used the statistical approach of mean and deviation *f*-test for linearity, to test if a linear relationship between customer satisfaction and customer retention existed. This test proved positive. Next binary logistic regression was used to test if their main variables had a direct or indirect relationship to customer retention behaviour. These variables proved to have an indirect influence. Finally three models were developed using linear regression. They do not provide exact prediction results in their paper, however they claim that all independent variables significantly predicted customer retention behaviour with an overall group retention average of 0.988.

2.9. Commercial CRM Software

Further to the review of literature, a survey was carried out on the identified most popular CRM software tools. All information was acquired directly from each of the vendor’s websites, with an additional search of available journal databases in the attempt of identifying published work related to each of the vendor’s products. An extensive research phase has lead to a collection of product brochures, specifications, and whitepapers for each of the products. Some journal papers have also been identified for certain vendors but it has been assumed after a thorough investigation that not all software have related publications.

The main observation drawn from this investigation is that most of the CRM vendors do not cater for churn prediction. Most have focused primarily on customer satisfaction. It is important that customers are identified before they actually churn, to provide the service provider a chance to retain them. The small numbers of vendors that have taken churn into consideration do not offer a specific churn management tool.

It is identified that CRM products mainly focus on cross-selling, up-selling, and competitor analysis. Competitor analysis can be regarded as customer churn management; however it should not be confused with customer churn capture. Competitor analysis is a business planning strategy and targets churn prevention. Identifying competing services helps to plan business strategies for the business to stay competitive and potentially win new custom. The research developed in this thesis aims at capturing churn after an existing customer has decided to defect. Due to the complexities and uncertainties involved in churn prediction it is likely that the CRM products mainly focus on churn prevention, cross-selling and up-selling opportunities because they have been unable to devise sound strategies for detecting churn. A full report regarding the investigation into CRM software is provided in appendix B of this thesis.

2.10. Discussion and Research Gaps

We have seen from the literature that the main trends for developing models to predict customer behaviour include regression analysis, neural networks, decision trees and Markov models. Decision trees appear to be the most popular choice, while neural networks have received slightly more interest than regression analysis. Focusing specifically on churn prediction, it can be observed that neural networks, decision trees, genetic algorithms, statistical methods and some new techniques have been mentioned widely.

It has not been surprising to discover that minimal research in the area of customer churn management has been carried out using certain powerful techniques. The amount of research that has been done using a GA is very small and no research at all can be found involving fuzzy logic for either churn management or CRM investigations. It is understandable that GA have received little attention as a genetic algorithm is more an optimisation technique than predictive technique. Using a rule based method such as fuzzy logic could be suitable for determining customer churn; however a lot would depend on the data, resulting in a methodology that is very dependent on the data and consequently on the environment and service sector in which the research had been based.

Through the investigation of literature it has become clear that the overall area of CRM has received a significant amount of research. Papers have been identified that attempt to predict repeat buying and determine customer satisfaction. Some researchers have focused on customer behaviour and some on loyalty reward schemes, all using varying technologies as the basis of their models. Most research has the common goal, of reducing customer churn and increasing customer spending.

From the limited research surrounding the prediction of customer churn, most authors agree that NN provide the best results. Many factors could either hinder or accelerate the usefulness of a predictive model. Factors to account for include data quality, data types and number of variables. Experiments using three of the most popular predictive techniques will be carried out to determine which is most accurate at predicting churn for the targeted service sector and the selected data.

Most papers report only the accuracy determining churn; however an effective churn prediction should provide high accuracy for the prediction of both churn and non-churn. The accurate prediction of customer churn is a difficult task, and segmenting predictions into the top 10% of customers with the strongest probability of defecting is a good way of limiting possible misclassifications, although, this method also limits the total number of possible correct classifications. I.e. if the test set only contains 500 data points, churn capture is limited to a total of 50 customers. The research presented in this thesis will be developed with the focus of accurately identifying customer churn while minimising misclassification levels, with no limitations on churn capture volumes. Segmenting the top 10% of customers most likely to churn and only reporting the results from this group appears to be a popular technique.

The research gaps from this literature review can be identified as follows. The first gap deals with predicting churn sufficiently in advance. The research has shown that only a minimal amount of work has actually focused on attempting to develop a churn model to predict churn sufficiently later than the occurrence of the event that has led to the customer's intention to defect. In the best cases churn is predicted a maximum of one month ahead. With usually limited specialist retention staff available to a company, this may be insufficient time for the successful deployment of a retention campaign.

Therefore the first major identified gap in research is the need for a methodology that is capable of providing a significantly greater time between event and occurrence.

The second gap in current research is the fact that most methodologies have been based on demographics and usage data. The problem with large service industries is that quite often data is spread over multiple departments, sites and even countries. This data is clearly very good for basing a churn prediction methodology but usually access to a lot of these variables is restricted by both internal management and in the case of larger companies, monopoly regulations. Therefore a solution should be developed targeting data that is accessible by all businesses regardless of size or service sector control.

A third gap in current research has been identified as previous methodologies appear to be developed in such a way that implementation in a business would require batch processing style architecture. A methodology is required that would allow real time monitoring of the customer base to maximise the time between the churn event and churn occurrence.

The fourth gap in research is the large misclassification of non-churners. Misclassifying non-churners as churners can be extremely costly. Therefore the methodology should consider a method for minimising the total number of misclassifications while preserving the total number of correct classifications. This in itself would be a considerable contribution to the field of customer retention as no identified research to date has attempted to tackle minimising misclassification rates.

The customer profiling methodology proposed in this thesis aims to address all of the above research gaps by:

- Predicting churn months in advance of the actual churn occurrence.
- Using repairs and complaints data for churn prediction instead of the demographic and billing data commonly used in current research.
- Providing a more accurate prediction of both churn and non-churn compared to the accuracy values reported in literature

3. Research Aim, Objectives and Methodology

The literature has identified how costly churn is to industry, and how benefits of a churn model can be hampered through misclassifying non-churners as churners. Development of a methodology that is capable of accurately capturing churn in advance of its actual occurrence could prevent significant loss of revenue to industry and largely boost profits.

From the survey of literature it has been identified that current methods have several limitations. Current predictive techniques do not provide adequate time between detection of churn and the actual churn event to successfully deploy a retention campaign. Companies who are churn aware typically designate a specialist retention department, consisting of individuals who have been specifically trained to contact customers to attempt to prevent customer defection. A nationwide company would have one retention department to cover the entire country and that department may have a total of between 20 – 30 staff. The more time the department has before customers churn, the more effective the retention campaign will be.

Literature has also suggested that researchers typically target high churn capture from their models. There has been a lack of documentation that targets the control of misclassification. Misclassifying non-churn as churn add to the numbers that need contacting by the retention department and as customers who never intended to churn are being contacted with expensive retention offers, real churners are left to defect.

To generate the longest possible timeframe between event churn detection and churn occurrence a real-time monitoring system should be in place so that customers can be flagged as defectors at the earliest possible time. Current techniques require data extraction and processing that is can take time and reduce the maximum distance between detection and event. Finally the success of churn prediction is extremely dependent on the predictor data. It is favourable in research to use demographic and usage data; however analysis using these data types could result in a model that breaks service regulations. Media businesses are regulated by a governing body called Ofcom. It is the purpose of Ofcom to promote fair competition and prevent monopolisation by one provider. Usage and demographic data could breach these regulations by encouraging price wars. For this reason an alternative source for predictor data is identified.

This chapter attempts to identify the aim and objectives of this research. The following will be discussed:

- Research Aim
- Objectives
- Scope
- Methodology

3.1. Research Aim

The aim of this research is to develop a customer profiling methodology for predicting churn in advance, while keeping the misclassification rates to a minimum.

3.2. Research Objectives

There are a number of objectives that need to be addressed in order to develop a methodology that will accomplish the research aim. These objectives have been identified as follows:

- To develop a quantitative model for churn index based on repairs and complaints data.
- To develop a customer profiling methodology that incorporates time element in the prediction of customer churn.
- To develop a customer profiling based technique for maximising future churn capture by identifying a potential loss of customer at the earliest possible point.
- To develop a customer profiling based technique for reducing misclassified customers, so that errors can be controlled.
- To identify and carry out three case studies for validating the proposed methodology using repairs and complaints data.

- To compare the results from the proposed methodology against popular churn prediction techniques reported in literature.

3.3. Research Scope

Based on the objectives mentioned above, the scope of this research can be summarised as follows:

- Domain: This research targets the services that fall under the umbrella of the telecommunications industry, i.e. telephone and mobile services.
- Data: This research focuses on faults and complaints data as an alternative to demographic and usage data as to avoid conflicts with monopoly regulations.
- Modelling techniques: As identified in the previous chapter there are many modelling techniques available for forecasting an event. This research will analyse the three most popular techniques (Neural Networks, Regression Trees and Linear Regression), and develop a methodology for future churn prediction, through a series of experiments.
- Literature survey: The literature survey focuses on the field of customer loyalty, satisfaction and churn in the service sector. This survey analysis is not limited to the telecommunications industry only.

3.4. Research Methodology

This section discusses the methodology that has guided the main activities of this research. As can be recognised from the aim and objectives, the research is typically quantitative in nature. The problem of predicting customer churn is a data mining problem, with results obtained by the use of statistical methods and decision support computing systems. The results will further be presented in the form of statistical comparisons. Qualitative research does not attempt to quantify results through statistical summary or analysis. Instead qualitative research typically involves techniques that include interviewing methods, observations without any formal measurement, and case studies that are typically an in-depth examination of one person

(Marczyk et al., 2005). This research has therefore focused on quantitative research methods. The research methodology applied to this research can be viewed in Figure 9:

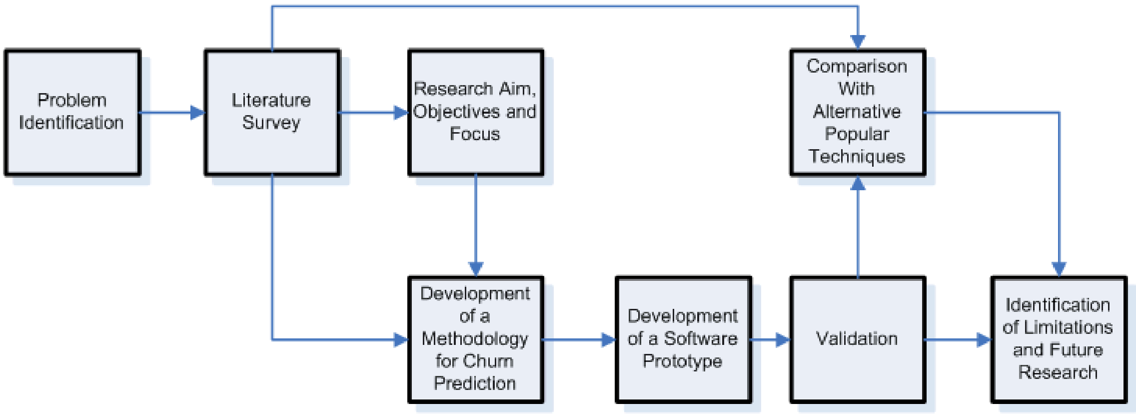


Figure 9: Research Methodology

It can be seen from Figure 9 that the research begins with an investigation of the problem definition. The results from this investigation provide a foundation on which to base a literature survey. The research aim and objectives are identified through the survey of literature. With the aim and objectives defined, development of a methodology for churn prediction is possible. Once a methodology has been created a software prototype is constructed. Validation of the methodology is performed using the software prototype. The results obtained through validation are compared with two popular churn prediction techniques identified through the literature survey. Limitations of the research are concluded through validation and churn prediction methodology comparison.

3.4.1. Problem Identification

As mentioned in chapter 1, it is becoming common knowledge in business that retaining existing customers is an important strategy to survive in industry. Once identified, these customers can be targeted with proactive retention campaigns in a bid to retain them. These proactive marketing campaigns usually involve the offering of incentives to entice the customer into carrying on their service with the supplier. These incentives can be costly, so offering them to customers who have no intention to defect results in lost revenue. Also many predictive techniques do not provide significant time to make customer contact. This time restriction does not allow sufficient time for capturing those customers who are intending to defect. Therefore, the aim of this research is to develop a methodology for predicting customer churn in advance, while keeping misclassification rates to a minimum.

3.4.2. Literature Survey

An extensive literature survey was carried out as part of this research in order to analyse and classify the state-of-the-art churn prediction techniques. Several areas from literature were identified for investigation. The review of literature was planned as follows:

- Begin with an investigation into the subject area of customer churn management to identify the different types of customer churn.
- Previous research into the areas of customer satisfaction and loyalty was investigated and analysed to identify any known links between customer loyalty, satisfaction and churn.
- An attempt was then made to narrow the literature survey specifically targeting state-of-the-art research in the area of customer retention.
- The required steps for the creation of a churn prediction methodology were investigated.
- The most common modelling techniques as favoured by literature, were investigated for customer churn prediction.
- An identification of the main research gaps in the area of customer churn prediction was then possible.

3.4.3. Identification of Research Aim, Objectives and Focus

The survey of literature highlighted the main research issues that need to be addressed for handling the problem statement identified in Chapter 1. This enabled the precise identification of the research aim and objectives that can address the identified issues. The literature survey also enabled the identification of the drawbacks of the existing churn prediction techniques when addressing future prediction while limiting misclassification rates. These research gaps in existing churn prediction techniques define the focus of this research.

3.4.4. Development of a Methodology for Churn Prediction

A novel customer churn prediction methodology is proposed by this research to address the limitations of existing techniques. Development of this methodology is carried out in a systematic, step by step fashion, adding further features as research progresses. These features are those which are not adequately addressed by current research and should enhance the predictive qualities of the methodology. The best modelling method for predicting an event will be used as the starting point of the methodology development to enable the research to inherit all the strengths of this method while further enhancing it to create a methodology for predicting customer churn in advance with minimum misclassification rates. The stages for developing the proposed churn prediction methodology are as follows:

- Identification of the challenges in creating a churn prediction methodology
  - Problems regarding data (missing values, incorrect formats, etc)
  - Regulated restrictions (such as monopoly regulations)
- Determination of the best modelling technique through an experiment based analysis and comparison of:
  - Neural Networks
  - Regression Trees
  - Linear Regression
- Development of the churn prediction methodology using the following stages
  - Initial NN experiments (to determine a starting point)
  - Evolve the methodology based on the initial results
  - Investigate a technique based on profiling the customers
  - Develop the profiling technique
  - Address the misclassification problem using the concept of high/low risk profiles

Testing of the overall methodology using manual experiments

As shown above, development of the churn prediction methodology begins with an investigation into the challenges involved. These challenges include the limitations imposed by Ofcom, the telecommunications industries governing body, and issues that need to be accounted for regarding data issues. This should provide a foundation on which to base the research, with an understanding of what can be investigated and what should be avoided.

The next stage is to investigate the best modelling technique to base the methodology. Three popular techniques have been identified from literature: NN, CART, and LRM. Experiments are conducted to determine which method provides best results for churn prediction. Once it has been determined which technique provides the best results for churn prediction, work can begin on the development of the proposed churn prediction methodology. To initiate the research, experiments are performed on the best identified technique only. The results are analysed and development is planned based on the findings. At this stage investigations will begin on how to proceed with the methodology development. By analysing the results generated from the initial experiments and identifying the strengths and weaknesses of the initial results it should be possible to identify the best direction in which to continue.

An initial profiling based methodology is developed and the results are examined. Identification of the strengths and weaknesses of these results provide the direction in which development of the methodology can be further enhanced. A method for reducing misclassification numbers is also identified at this stage. The developed methodology is then tested manually.

3.4.5. Development of a Software Prototype for Validation

In this research, a software prototype is developed to enable systematic and controlled validation of the proposed methodology. Development of the software prototype is carried out as follows:

Investigation into software development methodologies structured software development lifecycle model (SDLC), extreme programming (XP), and rapid application development (RAD)

Identification of the most suitable software development methodology for progressing the development of the churn prediction software prototype (the extreme programming software development methodology was chosen after this analysis)

Applying the steps of the extreme programming software development methodology for the development of the churn prediction software prototype

- Vision Statement
- System Metaphor
- User Stories
- Gantt Chart
- Prioritising Development
- Programming the Churn Prediction Software Prototype
- Merging Software Releases
- Determining an Advanced Prediction
- Summary

As shown above, the development of a software prototype based on the proposed methodology begins with an initial investigation into the popular software development methodologies. By comparing the features of each software development methodology against the requirements of software prototype for this research, the most suitable software methodology for proceeding with software development is identified. The extreme programming (XP) methodology is identified as best suiting the requirements for software development for the churn prediction software prototype. The XP methodology is based on identifying and creating various software releases. Each release enhances the functionality of the software platform until the desired product is developed. The software development followed the steps of the XP methodology, and is discussed in full detail in chapter 6.

3.4.6. Validation Using the Software Prototype

The validation of the proposed methodology is carried out using the developed software prototype. Three case studies from alternative telecommunications services are analysed using the software prototype for automatic generation of results. These results are also validated and compared against two other popular classification techniques from literature. Validation of the proposed methodology is carried out as follows:

Case study development
Neural network development
Validation plan
Analysis of the three case studies
Generation of churn index values
Applying the proposed profiling methodology
Making churn predictions and comparing with actual churn/non-churn data

As shown above, validation of the churn prediction methodology begins with securing and developing three case studies from three separate areas of the telecommunications industry. These case studies include residential mobile phone service, residential broadband service and business landline service. The best NN architecture for each of the case studies is identified through an experiment based study. NN development is achieved through the use of Matlab's NN toolbox. This toolbox is investigated and discussed to understand how NN creation is achieved.

Next a validation plan is developed beginning with a description of the three case studies, then discussing the necessary steps taken for preparing each case study for analysis. The most suitable data from each of the case studies is identified for profile generation, and then development of three NN architectures is presented and the results for each experiment discussed and compared. The best NN architecture is used for the generation of loyalty index values. The customer profiling methodology is then applied. The identification of the most accurate and least accurate profiles for future churn prediction is discussed. The high risk profiles are used to make churn predictions and the results are compared with actual churn/non-churn data.

3.4.7. Comparison With Other Classification Techniques

The results are benchmarked against two other popular classification techniques found in literature. This should help to determine the advantages and weaknesses of the proposed churn prediction methodology. The following steps are carried out for each case study:

- Generate churn predictions for the methodology proposed by Hu (2005) and compare with actual churn/non-churn data
- Generate churn predictions for the methodology proposed by Hwang et al. (2004) and compare with actual churn/non-churn data
- Generate churn predictions for the proposed methodology and compare with actual churn/non-churn data
- Compare accuracy of the three methodologies for predicting customer churn in advance, while keeping misclassification rates to a minimum.

The results from the proposed methodology are compared against two other popular classification techniques from literature.

3.4.8. Identification of Limitations and Future Research Directions

Finally, the limitations of the proposed methodology are identified. Based on these limitations, the generality of the research and its contribution to knowledge are established, and the corresponding future research directions are proposed.

3.5. Summary

This chapter has discussed the following:

- It has stated the research aim
- It has discussed the research objectives that address the aim of the research

It has summarised the scope of the research based on the objectives

It has finally discussed the methodology that has guided this research. This methodology has eight main parts:

- Problem identification.
- Literature survey.
- Identification of research aim, objectives and focus.
- Development of a methodology for churn prediction.
- Development of a software prototype for validation.
- Validation using the software prototype.
- Comparison with other classification techniques.
- Identification of limitations and future research directions.

As stated in this chapter, the aim of this research is to develop a methodology for predicting customer churn in advance, while keeping misclassification rates to a minimum. The next chapter discusses the stages of developing this methodology.

This Page is intentionally
Left Blank

4.Customer Profiling Methodology for Churn Prediction

Chapter 2 investigated and reported all current and historical research in the area of customer churn management, identifying the most powerful and common methods for predicting customer churn and providing direction on how further research should proceed if a significant contribution to this field is to be achieved.

This chapter basis the research on the finding from chapter 2, documenting the stages of research that should lead to a generic, improved, powerful churn identification system that is capable of handling the complex issue of determining loss of contractual service. The chapter attempts to achieve the following:

To identify a test study on which the development can be based. The data contained within the case study should be non industry specific in order to achieve a generic methodology

To develop a customer churn methodology that is capable of generating an advanced, accurate prediction of customer churn.

In addition, the proposed customer churn prediction methodology should conform to the following guidelines:

Generic – so that it is not restricted to a specific problem or industry

Modular – so that is can be configured to assess potential churn from multiple sources

Performance - Better performance than the current state-of-the-art techniques as identified from literature

- Increase churn capture accuracy
- Lower misclassification rates
- Longer time between initial identification of churn and actual occurrence of churn

4.1. Challenges in Creating a Customer Churn Prediction Methodology

Identification of a customer's intention to end his/her contractual service is a complex problem. It has been identified from chapter 2 that many researchers working in this field have opted to base their research on demographic, usage and billing data. For a contractual service in particular, demographic, usage and billing data is the most convenient choice for basing a customer churn prediction methodology for the following reasons:

- By analysis of a customer's usual usage behaviour and monitoring activity, sudden reduced usage can be interpreted as a customer's intention to churn.
- By monitoring a customer's spending behaviour, sudden reduced revenue from a particular customer can be interpreted as a customer's intention to churn
- By segmenting customers by demographic location a clear understanding of typical location behaviour can be determined. This would be more beneficial to the wireless telecommunications industry where poor coverage in a particular area could signify poor retention rates
- By segmenting customers by age group, it is possible to determine the behaviour of certain age ranges. E.g. age group 18-24 could value technology most while a 35-45 age group could be more favourable towards service performance.

Several problems have been determined with the use of demographic, usage and billing data. The most significant of these problems was established through an initial requirements analysis with British Telecom (BT). BT is the UK's leading telecommunications service provider responsible for over 28 million telephone lines in the UK. It was advised by BT that as a leading service provider they had to conform to specially imposed government regulations as outlined by Ofcom (formerly known as Oftel) (Wikipedia, 2008a). Ofcom is the independent regulator and competition authority in the UK for the telecommunications industry, receiving its full authority from the communications act in 2003. It has a statuary duty to act in favour of the public by promoting competition within the telecommunications industry so as to provide competitive consumer options (Wikipedia, 2008b).

The BT brand became established in 1980 as the official name of Post Office Telecommunications, becoming a state owned organisation independent of the post office. At this time BT was the UK's only telecommunications provider. This changed however in 1982 when their monopoly was broken by Mercury Telecommunications who was also granted a licence to provide telecommunications services. Presently there are multiple telecommunications services in operation throughout the UK (Wikipedia, 2008a).

The Ofcom regulations prevent BT undertaking investigations into any areas where the outcome could hinder competition and lead to a business monopoly. Because usage and billing data in particular are so closely connected to pricing strategies BT are reluctant to pursue research based on these cases is as to avoid breaching any Ofcom regulations. With this information in mind, a generic methodology is required that will be deployable while conforming to Ofcom regulations.

An assumption regarding the use of demographic, usage and billing data is that basing a model on these types of variables would result in a specific methodology with little scope for exporting to other services. Repairs and complaints are common in a huge variety of sectors whereas usage data is specific to very few. Narrowing this down even further, the types of usage data between alternative service industries is very specific to the industry. For example 'call frequency', 'length', 'type' etc would be specific to the telecommunications industry while 'kilowatt-hours' would be specific to the electricity supply industry. 'loss of service' or 'equipment fault' could be applicable to both industries despite the fact that the actual nature of the businesses are different. Because of these reasons both from a research point of view and an industrial point of view repairs and complaints data was deemed to be the most suitable choice. Therefore a dataset comprising of customer complaints and repair data was obtained from BT for initial experimentation. This dataset contains 25 variables which represent repairs and complaint information for 18453 customers for BT's residential broadband service. The data for each customer is provided for all months through January to October 2004; however the churn data is only available for the month of October. So that this dataset can be used for validation of the final full methodology later in the research, a small number of the customers have been extracted into two new datasets for training and validation. The training dataset contains 202 customers (50% churners and 50% non-churners) and the validation dataset contains 700 customers (30% churners and 70% non-churners). Greater details regarding these datasets will be provided later in this chapter. An example of the variables contained in the dataset can be seen from Table 2:

Table 2: Initial Experimental Data

Variable Name	Variable Description
LIFETIME	How long a customer has been with the company
PROMISE DURATION	Provisionally the number of days promised to the customer until resolution
REPAIR DURATION	Number of days the repair to be resolved
MU COUNT	No. of internal queues the job has come through
COMPLAINT DURATION	Duration of the complaint in days
COMPLAINT RESOLUTION OVERDUE	No. of days the promise resolution date ran over
COMPLAINT CODE	Code of complaint type
REFUND AMOUNT	Money refunded to the customer for loss of service
PROMISE COUNT	No. of provisions made
REPAIR COUNT	No. of repairs made
REPAIR OVERDUE	No. of days the promise resolution date ran over
NUMBER OF COMPLAINTS	No. of complaints made

4.2. Strategy for Methodology Development

Various stages have been identified for the development of a customer profiling methodology for churn prediction. These stages are outlined as follows, and will be discussed in greater detail later in this chapter:

Determining a predictive model – This stage will identify the most suitable predictive model for churn prediction through experimental analysis and comparison of.

- NN
- RT
- LRM

Initial churn prediction experiments – Initial experiments will be carried out to predict customer churn to provide greater knowledge of the complexity of the problem and determine further direction.

Validate results – The results from all experiments will be validated and analysed.

Evolve the methodology – The methodology will be evolved based on the results generated at the validation stage.

The stages outlined above are illustrated by the flow chart shown in Figure 10 to provide a clear understanding of how the methodology creation is progressed:

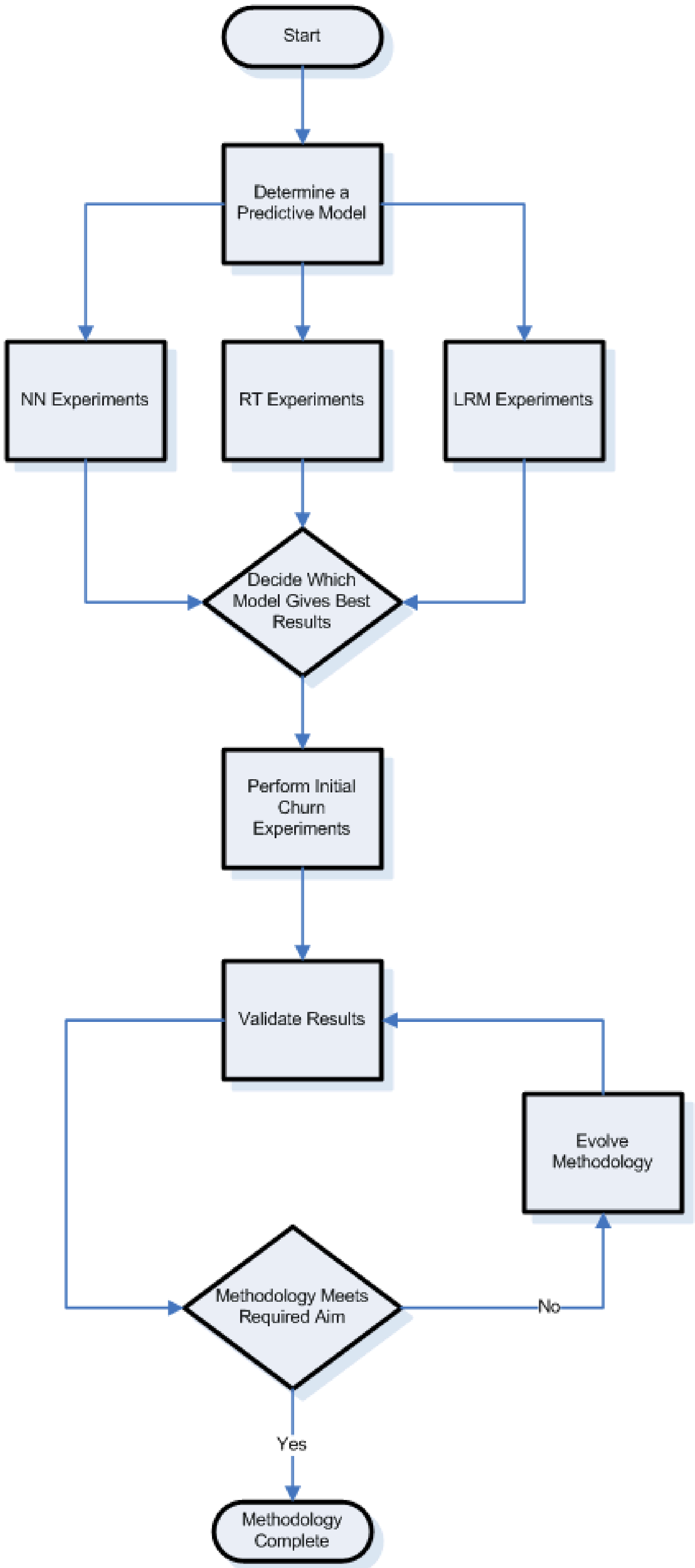


Figure 10: Methodology Development

4.3. Determining a Predictive Model

Three models were identified from the literature as the most popular for predicting customer churn, neural networks, regression trees and linear regression. The development of these models is discussed in this section.

4.3.1. Neural Networks

Several experiments were performed using feed-forward, back propagation neural networks as there are many variations available. The architecture remained the same however different activation functions where applied to assess which would performed the best for predicting customer churn. All architectures consisted of twenty five inputs, two layers with 1 neuron per layer, and an output layer. An illustration of the architecture can be viewed in Figure 11: -

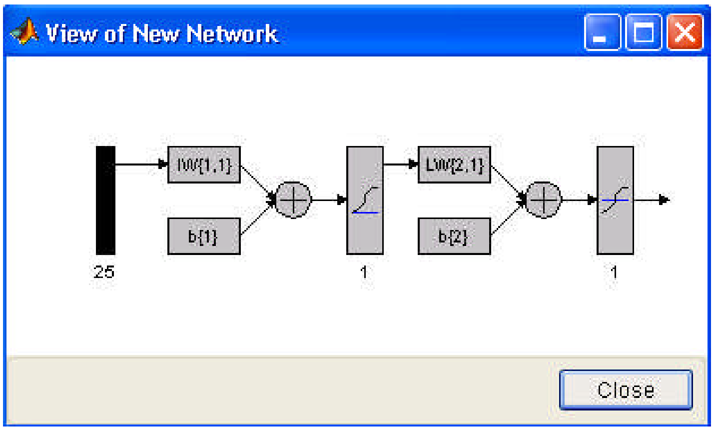


Figure 11: Neural Network Architecture

Figure 11 illustrates how data is passed through the created neural network architecture. From left to right, the diagram displays the input layer. As can be noted from the diagram each input has its own weight associated to it (IW[1,1] = input weight), with a general bias (b[1] = input bias) applied. The number underneath the first box indicates the total number of inputs. All inputs are then summed. Next the Logsig activation function is applied (this stage is symbolised by the sigmoid sign). The number underneath the box containing the sigmoid sign represents the total number of neurons contained within the hidden layers in that particular neural network architecture. The resulting value again has a weight (LW[2,1] = layer weight) and bias (b[2] = Layer Bias) applied and the churn rate is finally output via the output layer. The number underneath the final box indicates the total number of outputs.

Several experiments were performed with various neural network activation functions. These experiments are detailed in chapter 6 – validation. The Logsig activation function displayed best results from these experiments so this is the activation function used when configuring a new network. The NN is then trained using a training dataset containing a churn/non-churn ratio of approximately 50:50, totalling 202 customers. This ratio was selected for experimentation purposes as it was assumed that providing sufficient churn points would aid in NN training. This ratio was used for technology comparison only.

The performance of the neural network was validated by instructing it to predict the churners from a validation dataset containing a total of 700 customers consisting of 70% non-churn and 30% churn. Churn is indicated in the original target dataset as a 0 for non-churn and a 1 for churn. The neural network generates a decimal value between 0 and 1 for all customers so it has to be determined which value ranges represent churn. The higher the value, the higher the likelihood of defection and it was found that a threshold of 0.7 for classifying customers as churners provided the best results. A further strength of the NN has been identified, as the ability to manually define a churn threshold. This could be regarded as 'fine tuning' the output from the NN model. The varied churn rates obtained from the neural network are illustrated in Figure 12:

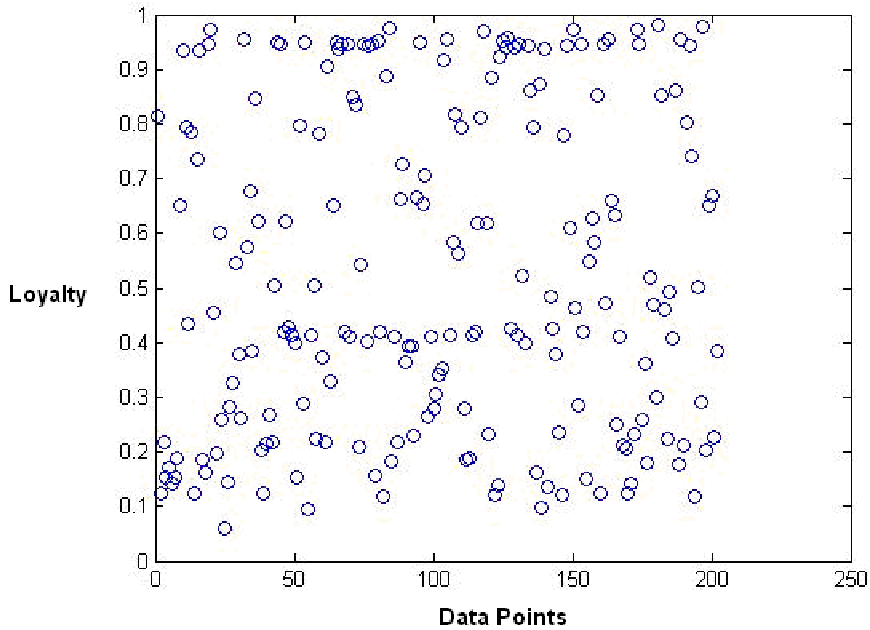


Figure 12: Churn Rate Predictions

Figure 12 helped to generate a greater understanding regarding the NN outputs and led to an idea that even though most of the churn is captured in the values 0.7 and higher, maybe the customers with NN values 0.6, 0.5, 0.4 etc were on their way to churn.

The confusion matrix in Figure 13 illustrates the results obtained from the neural network, using a Logsig activation function and defining a churn threshold setting of 0.7: -

True Labels	Estimated Labels 0	1	Totals
0	54	40	94
1	10	98	108
Totals	64	138	202

Figure 13: Churn Results on Training Dataset

It can be observed from the confusion matrix in Figure 13 that the results were accurate. This accuracy was further analysed by using the simple formula of dividing the predicted churners by the actual churners. The feed-forward, back-propagation neural network using the Logsig activation function has provided 90% accuracy of churn capture and 57.4% accuracy of non-churn capture. Since the churn capture is deemed the most important result at this stage, the results are considered good.

Secondly a neural network was created using the Purelin activation function and Satlin activation function. The Purelin activation function also provided 90% accuracy but the misclassification accuracy fell to 54%. The Satlin activation function only provided 41% churn prediction accuracy with a 56% non-churn capture. These results can be seen clearly by the table in Table 3:

Table 3: Activation Function Comparison

	LOGSIG	PURELIN	SATLIN
Churn Accuracy	90.00%	90.00%	41.00%
Non-Churn Accuracy	57.40%	54.00%	56.00%

Based on the results displayed in Table 3, it was decided to construct the neural network using the Logsig activation function with an underlying Bayesian architecture.

An analysis of the weights established for the twenty five variables as provided by the neural network have provided an understanding of the seven variables that hold the most significance for predicting customer churn. These variables are as follows: -

1. Cash Conceded – C_CASH_CONCEDED
2. the company - LIFETIME
3. How long the repair took – R_DURATION
4. No. of repairs made – R_COUNT
5. Promised resolution time – P_DURATION
6. Indication that an order for a repair has been placed – R_FLAG
7. No. of days the promise & resolution date ran over) - C_OVERDUE

4.3.2. Regression Trees

Classification and regression trees (CART) are constructed by recursively splitting the instance space into smaller subgroups, 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 (Muata and Bryson, 2004b).

Classification and regression trees were used to establish if they would offer a more accurate method of predicting customer churn using the complaints and repairs data. Regression trees rather than classification trees were selected because classification trees generate an output that mimics the format of the target that has been used for training. As we have previously discussed, churn is represented by the value of either 0 or 1 so a classification tree will generate outputs as these values. The regression tree however, returns a decimal value between 0 and 1, similar to the output from the NN. This makes the regression more suitable for analysis because the decimal value between 0 and 1 can be directly used as a probability of churn.

A regression tree was trained using the training dataset of 202 customers, comprising of a 50:50 churn/non-churn ratio. Figure 14 illustrates the resulting regression tree:

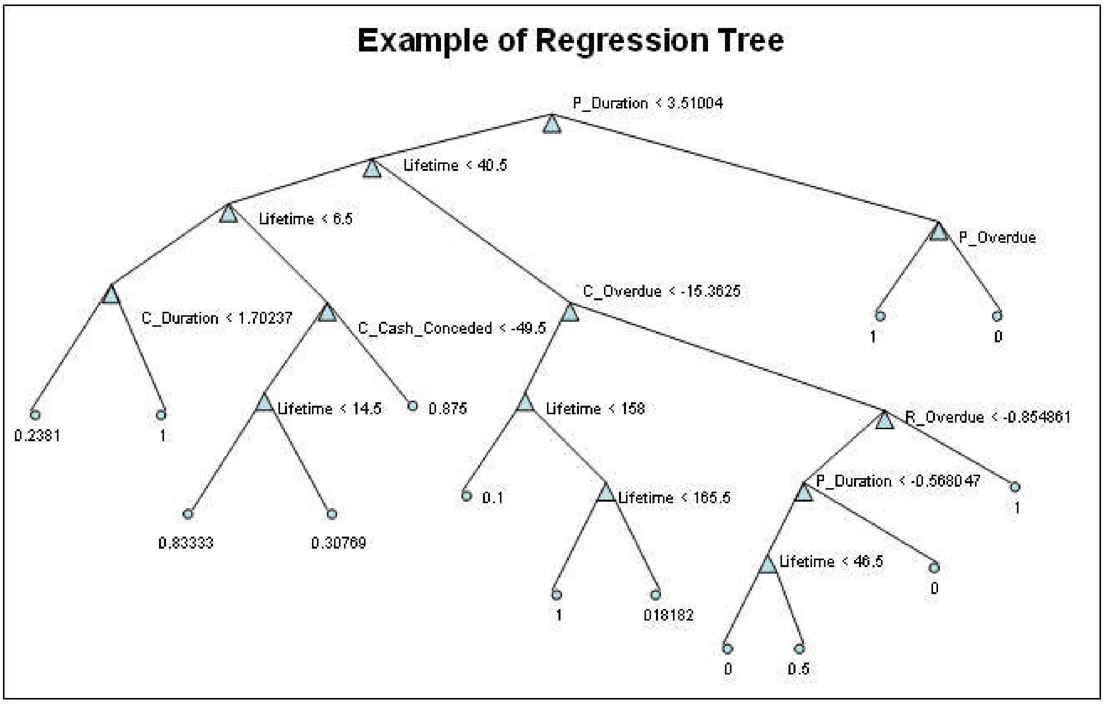


Figure 14: Regression Tree

Key: - R_OVERDUE (No. of days the promise resolution date ran over), P_DURATION (Provisionally the number of days promised to the customer until resolution), C_OVERDUE (No. of days the promise & resolution date ran over), P_OVERDUE (Number of days by which promised delivery was late), C_DURATION (Duration of the complaint in days), C_CASH_CONCEDED (Money refunded to the customer for loss of service), LIFETIME (How long a customer has been with the company)

The regression tree in Figure 14 illustrates the influencing variables as determined by the regression tree, and shows the various conditions that result in the prediction of churn.

Twenty five variables were used to train the regression tree; however according to the constructed tree as shown in Figure 14 only seven variables were selected as strong influences. These variables are as follows: -

1. Cash Conceded – C_CASH_CONCEDED
2. No. of days the promise resolution date ran over - R_OVERDUE
3. Provisionally the number of days promised to the customer until resolution - P_DURATION
4. No. of days the promise & resolution date ran over - C_OVERDUE
5. Duration of the complaint in days - C_DURATION
6. Number of days by which promised delivery was late - P_OVERDUE
7. How long a customer has been with the company - LIFETIME

The confusion matrix in Figure 15 shows the results of the regression tree. As with the NN experiment the regression tree was used to predict the targets of the initial training dataset of 202 customers. The churn threshold was set to a level of 0.7.

```
       | Estimated Labels
 True  |
Labels |    0      1   | Totals
-------|---------------|-------
0      |   63     31   |   94
1      |   30     78   |  108
-------|---------------|-------
Totals |   93    109   |  202
```

Figure 15: Regression Tree Analysing Training Dataset

As can be seen from Figure 15, when generating predictions for the training dataset of 202 customers using the regression tree, the results are good, providing 72% churn accuracy and 67% non-churn accuracy. This shows that the regression tree is capable of generating good churn predictions and the clearly defined rules that are generated via training make it easy to understand why they have been a popular technique amongst researchers. However when compared with the same experiment performed by the neural network illustrated in Figure 13, it shows that the neural network significantly outperformed the regression tree.

4.3.3. Linear Regression

Linear regression analysis is a popular technique used by the researchers dealing with predicting customer satisfaction. This is usually the first stage of developing more complex models (Kim and Yoon, 2004). SPSS (statistical package for the social sciences) was used for generating a linear regression model based on the same training set that was used for the previous NN and regression tree experiments. Once the data was analysed using SPSS the results shown in Table 4 where obtained: -

Table 4: SPSS Linear Regression Results

Variables	Standardised Coefficients
(Constant)	.480
LIFETIME	-.074
P_DURATION	6.311
P_OVERDUE	-8.715
R_DURATION	-2.314
R_PROMISE_LEVEL	1.699
R_MU_COUNT	-.002
C_DURATION	2.913
C_OVERDUE	1.225
C_ISSUE_TYPE	-1.449
C_CASH_CONVERTED	.098
P_COUNT	2.835
R_COUNT	.617
R_OVERDUE	.016
C_COUNT	-2.589

Table 4 displays the standardised coefficients as generated by SPSS. The linear regression formula can be created from these calculations. The actual linear regression formula is displayed in Equation 2:

$$
\begin{aligned}
&Churn \quad 0.480 \quad (\ 0.074)*(Lifetime) \quad (\ 6.311)*(P\_Duration) \quad (\ 8.715)*(P\_Overdue) \\
&\quad (\ 2.314)*(R\_Duration) \quad (\ 1.699)*(R\_\mathrm{Pr}omise) \quad (\ 0.002)*(R\_MU\_Count) \\
&\quad (\ 2.913)*(C\_Duration) \quad (\ 1.225)*(C\_Overdue) \quad (\ 1.449)*(C\_Issue) \\
&\quad (\ 0.098)*(C\_Cash\_conceded) \quad (\ 2.835)*(P\_Count) \quad (\ 0.617)*(R\_Count) \\
&\quad (\ 0.016)*(R\_Overdue) \quad (\ 2.589)*(C\_Count)
\end{aligned}
$$

Equation 2: Linear Regression Model

Table 5 describes each of the variables that have been favoured by SPSS in generation of the linear regression model:-

Table 5: Variable descriptions

Variable Name	Variable Description
LIFETIME	How long a customer has been with the company
P_DURATION	Provisionally the number of days promised to the customer until resolution
P_OVERDUE	Number of days by which promised delivery was late
R_DURATION	Number of days the repair took to be resolved
R_PROMISE_LEVEL	No. of promises made (> 1 means a promise was broken)
R_MU_COUNT	No. of internal queues the job has come through
C_DURATION	Duration of the complaint in days
C_OVERDUE	No. of days the promised complaint resolution date ran over
C_ISSUE_TYPE	Code of complaint type
C_CASH_CONCEDED	Money refunded to the customer for loss of service
P_COUNT	No. of provisions made
R_COUNT	No. of repairs made
R_OVERDUE	No. of days the promised repair resolution date ran over
C_COUNT	No. of complaints made

As shown in Table 5, out of the twenty five initial variables presented to SPSS, only fourteen were used in the final linear regression model. Seven of the variables holding the most significance for predicting customer churn using linear regression are identified as follows: -

1. Number of days by which promised delivery was late - P_OVERDUE
2. Provisionally the number of days promised to the customer until resolution - P_DURATION
3. Duration of the complaint in days - C_DURATION
4. No. of provisions made - P_COUNT
5. No. of complaints made - C_COUNT
6. Number of days the repair took to be resolved - R_DURATION
7. No. of promises made (> 1 means a promise was broken) - R_PROMISE_LEVEL

The confusion matrix in Figure 16 shows the results of the linear regression model when using to predict churn for the training dataset of 202 customers:.

```
       | Estimated Labels
 True  |
Labels |    0      1  | Totals
-------|--------------|-------
0      |   90      4  |   94
1      |   46     62  |  108
-------|--------------|-------
Totals |  136     66  |  202
```

Figure 16: Linear Regression results for Training Data Set

As with the neural network and regression tree models, the results for linear regression when analysing the training dataset of 202 customers is accurate showing that the model works. However, when compared to the same experiments performed using neural networks and regression trees on the same data it is clear that linear regression is the least accurate of all three technologies for capturing churn, however it does provide the least number of misclassifications. Table 6 provides a comparison of the results obtained from each of the technologies.

Table 6: Technology Comparison

	Neural Network	Regression Tree	Linear Regression
Churn	90.70%	72.20%	57.40%
Non-Churn	57.40%	67.00%	95.70%

4.4. Validation of Results

To further test the performance of each technology a validation dataset of 700 customers has been presented to all the three models. The validation dataset contains customer data that is completely independent of initial training set. The churn/non-churn ratio for this dataset is 70% non-churn and 30% churn. All validation results are reported in this section.

4.4.1. Neural Networks Results

Both the Bayesian and standard feed-forward back propagation neural networks were trained using the training dataset and validated using the validation dataset. The results are displayed in Figure 17: -

Bayesian Architecture

True Labels	Estimated Labels 0	Estimated Labels 1	Totals
0	371	119	490
1	63	147	210
Totals	434	266	700

Standard Architecture

True Labels	Estimated Labels 0	Estimated Labels 1	Totals
0	386	104	490
1	93	117	210
Totals	479	221	700

Figure 17: Neural Network Validation

As can be viewed in Figure 17, the Bayesian architecture correctly identified 70% of the churners while the standard feed-forward architecture only identified 56% of the churners. However the standard architecture correctly identified 78% of the non-churners while the Bayesian architecture identified 75% of the non-churners correctly. The real difference between the two architectures can be seen when the prediction accuracy is calculated. The prediction accuracy can be calculated by dividing the total number of correct predictions by the size of the dataset (correct churn + correct non-churn / dataset size). This shows that the Bayesian methodology has a provided a prediction accuracy of 74% while the standard methodology has provided a prediction accuracy of 71%.

4.4.2. Regression Tree Results

When the tree was used to predict churn for the validation dataset the following results were obtained, as shown in Figure 18: -

True Labels	Estimated Labels 0	1	Totals
0	435	55	490
1	71	139	210
Totals	506	194	700

Figure 18: Regression Tree Churn Prediction

It can be observed from the confusion matrix in Figure 18 that the regression tree provided good results for churn prediction correctly predicting 66% of the churners and 88% of the non-churners, with an overall accuracy of 82%. Another advantage of using a regression tree is that the extracted rules could be used for creating a fuzzy clustering model.

4.4.3. Linear Regression Results

The linear regression model was set to predict churn for the validation dataset. The results can be viewed in Figure 19: -

True Labels	Estimated Labels 0	1	Totals
0	465	25	490
1	103	107	210
Totals	568	132	700

Figure 19: – Results for Linear Regression

It can be seen from the confusion matrix in Figure 19 that linear regression correctly predicted 51% of the churners and 95% of the non-churners, with an overall accuracy of 81%. It is interesting hat the linear regression model outperformed all other technologies for predicting non-churners but was not as accurate at predicting the churn customers.

4.4.4. Discussion of Results

It is apparent that each of the technologies provides varying results. A comparison of the results for all the methods can be viewed in Table 7: -

Table 7: Comparison of Technologies

	Accuracy %		
	Predicted Churn	Predicted Non-Churn	Overall Accuracy
NN FF/BP Bayesian Architecture	70%	75%	74%
NN FF/BP Standard Architecture	55%	79%	72%
Regression Tree	66%	88%	82%
Linear Regression	51%	94%	81%

It can be observed from Table 7 that neural networks with a Bayesian methodology perform the best for predicting customer churn while linear regression is very accurate for predicting non-churn. Overall the best performing technology was the regression tree and the poorest was the standard neural network.

The most difficult class of customers to predict and the most important are the churners. This means that the most accurate method for the author's research is the neural network using Bayesian architecture. Table 8 compares the most significant variables of all three technologies: -

Table 8: Most Significant Variables

	RT	NN	LRM
P_COUNT			
C_CASH_CONCEDED			
R_COUNT			
R_OVERDUE			
C_COUNT			
P_DURATION			
C_OVERDUE			
C_DURATION			
P_OVERDUE			
R_DURATION			
LIFETIME			
R_FLAG			
R_PROMISE_LEVEL			

As can be viewed from Table 8, several of the variables were used by more than one of the technologies. The best technology for predicting customer churn is the neural network with the Bayesian architecture. This technology has been selected as the best because it has shown the best results for capturing customer churn. At this point it is anticipated that an improvement on misclassification rates will emerge throughout the rest of the research.

The best technology to advance the research is a neural network with Bayesian architecture. Although regression trees show the best overall accuracy for predicting churn and non-churn, and linear regression shows the best accuracy for predicting non churners, the neural network with the Bayesian architecture displayed the best results for predicting churn. The profiles that will be developed in the next section will be based on the neural network with Bayesian architecture.

The majority of the work presented in this section was published in the conference paper (Hadden et al., 2006b). A portion has also been published in the conference paper (Hadden et al., 2006a). The identified best technology can be used for the development of a customer churn prediction methodology.

4.5. Development of a Customer Churn Prediction Methodology

The identification of the best model for predicting customer churn in advance went through several iterations with a significant amount of trial and error in the early stages. With the final identification of a feed-forward back-propagation neural network using Bayesian regularisation as the best technology for continuing the research, and the research gaps identified from an extensive review of literature the research is ready to move onto the next stage. It was investigating how the identified technology can be used to determine a prediction of future churn. The following section documents the experimental stages that evolved through the research process until a suitable methodology was created.

4.5.1. Initial Experiments

Initial experiments began with investigating how to achieve a future prediction of customer churn, and involved simple neural network experiments to help understand the complexity of the problem. These simple experiments were deemed as starting point for establishing the best direction in which to proceed with the research.

It had already been determined how accurately a neural network could be used to predict churn as it occurred, e.g. predicting churn for October using the input data from October, as reported in the comparison of techniques section 4.3.

It was felt that the next logical step would be to determine if the neural network could predict churn in the future by e.g. offering input data from September to predict the target churn for October. These experiments continued for some time, using alternative sets of data and alternative neural network configurations. The results in each case were very poor and it became clear that this was not the way to proceed with the research and an alternative approach was required. An example of the results obtained through these experiments can be seen in Figure 20. These experiments did however; help with understanding the complexity of the problem.

True Labels	Estimated Labels 0	1	Totals
0	212	278	490
1	141	69	210
Totals	353	347	700

Figure 20: Future Churn Model 1

Unlike a decision tree, neural network outputs are variable between 0 and 1 providing flexibility for analysis. The closer the NN output is to '1' the more likely the customer will churn. Not all churners are graded with the same value by the neural network therefore a 'churn threshold' is required, stating that all customers with NN values over a certain value should be classified as potential churners, and all customers with NN values under that threshold classified as non-churners. The churn predictions could be capped by setting the churn threshold high, aiding in the reduction of misclassification rates while preserving capture of correct classification rates. The objectives of the proposed framework are to:

- Determine a framework that could support prediction of churn.
- Achieve the maximum possible future prediction
- Base the methodology on data that is available across multiple service sectors and accessible for analysis by any business regardless of size or monopoly status.
- Minimise the total number of misclassifications from the predictions to reduce retention costs to the business.

No previous research could be identified for predicting future churn so the aim was to use the churn index values that had been determined for each customer using the NN for predicting future churn. It should be noted that loyalty index equals 1 – churn index.

At this point in the research the question arose, 'how can the values generated by the NN be used effectively to build a methodology that can identify an advanced prediction of customer churn?". After some thought it was realised that the NN provided a customers propensity to churn on a monthly basis. A theory emerged that possibly customers with similar propensity measures over a set period of time would also display similar characteristics with regards to how long it takes them to terminate their service with their provider. This lead to the theory of a customer profile based methodology.

It was clear that the decline from loyalty to dissatisfy required further research. It seemed necessary to come up with a way to visually see what was happening with the customer loyalty index values over time and the most suitable method of doing this was by building a graphical representation of the customer's loyalty value history.

4.5.2. Evolution of a Customer Profiling Framework

At this point of the research the data described in section 4.1 was the only data available to base experimentation, therefore development continued with this dataset.

The first stage of the analysis was to create graphical profiles for each customer. To achieve this, a NN output was generated for each customer on a monthly basis. These NN values were then flipped using the equation *loyalty = 1- churn* and stored by month in an excel worksheet. Graphs were then generated from the monthly sequence of each customer. An example can be seen in Figure 21:

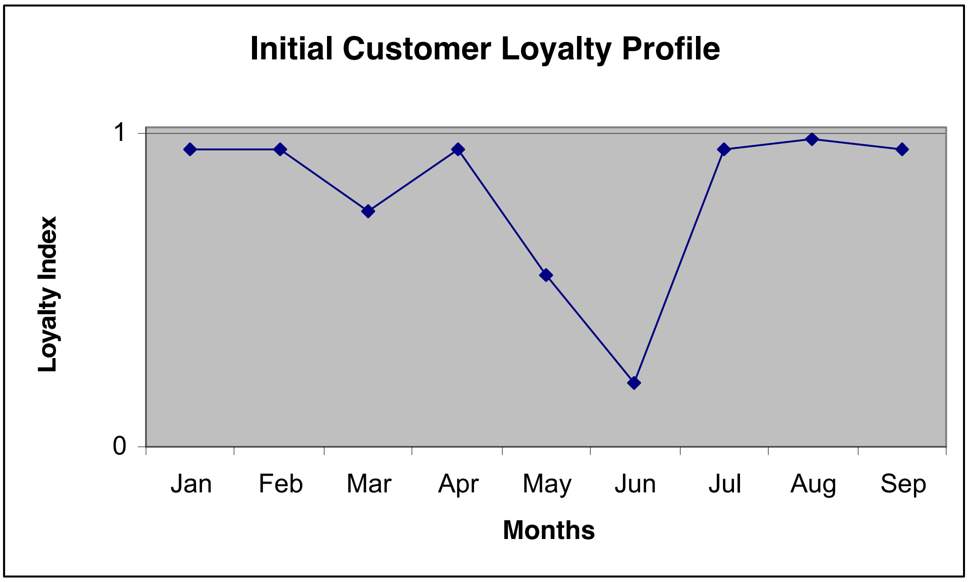


Figure 21: Initial Customer Loyalty Profile

Customer loyalty profiles such as the one displayed in Figure 21 are created from the loyalty index values. An attempt was made to manually match profiles between customers. Initial evaluations of profiles suggested that a comprehensive comparison between customers would be impossible with the profiles in this raw format, as there were what appeared to be hundreds of variations. At this point it was important i) to limit the number of possible profiles, ii) to enable a comparison between customers and iii) to allow a common comparison strategy to be applied. These points were addressed through consideration of the following issues:

1. What should be regarded as separate profiles? An example can be seen from the 2 graphs in Figure 22:

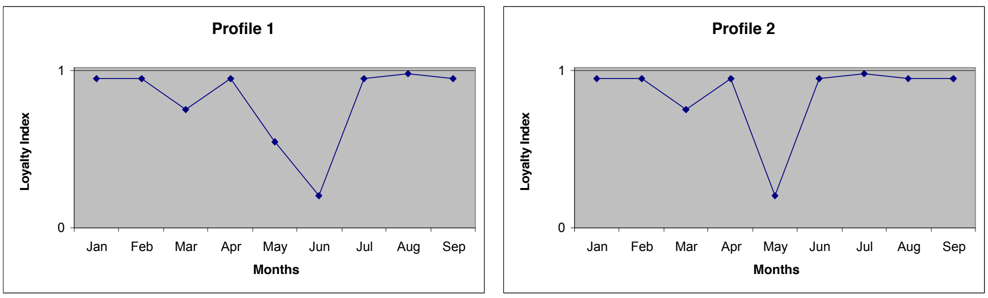


Figure 22: Profile Comparison

Should profile 1 on the left be classified the same as profile 2 on the right or does the 2 stage decline shown in profile 1 (between April and June) justify it as being a completely different profile?

2. What level of loyalty index decline should be regarded as a significant contributor for which to base profiles? Again using Figure 22 as an example, should the slight decline between August and September in profile 1 be considered insignificant and disregarded or is this significant enough to base profiles?

3. How could the total number of generated profiles be reduced to a controllable size?

4. Is it feasible to consider loyalty index values to freely move up or down?

5. Churn is flagged once the loyalty index value falls below a user defined threshold. Because during the initial NN experiments it was empirically found that churn index values above 0.7 appeared to be a suitable threshold, the loyalty index churn threshold becomes 0.3. Once a customer profile falls below this threshold, should we still monitor and classify values that continue past this point?

The first question considered was question 4, should profiles be allowed to freely move up and down? At this point an assumption was made that it was inappropriate for the loyalty index values to be allowed to move back up once they had declined. The reason this happens is because when the NN generates the churn index values it does not take care of any previous activity. Each month is generated as if it were a brand new case. Therefore it was assumed that loyalty index values should be recorded in a continuously falling sequence.

Preventing the loyalty index values to move back up solves some of the identified issues with profile analysis. It would address issue 3 by limiting the total number of profiles that could be generated. Furthermore, monitoring past the churn threshold would not be viable because once the profile had reached zero loyalty monitoring would not be possible. This addresses issue 5. A suitable rule had to be decided about how to determine the decline of customer loyalty. It was assumed that 0.1 should count as a decline, e.g. loyalty 0.9 to loyalty 0.8. This decision was made to prevent possibly hundred of profiles being generated from minor fluctuations. All less significant movements should be disregarded. This addresses issue 2. Issue 1 is also solved through this decision because now there is a strategy in place as to what is regarded as activity and what is not.

It was decided to record loyalty in a continuous declining fashion. The next question was how could this be achieved? When the NN generates churn index values it treats each month as an individual case. No consideration is taken into account for any previous activity. By applying a decreasing loyalty approach the example profile in Figure 21 would become the profile displayed in Figure 23. The details of this transformation are explained in section 4.5.2:

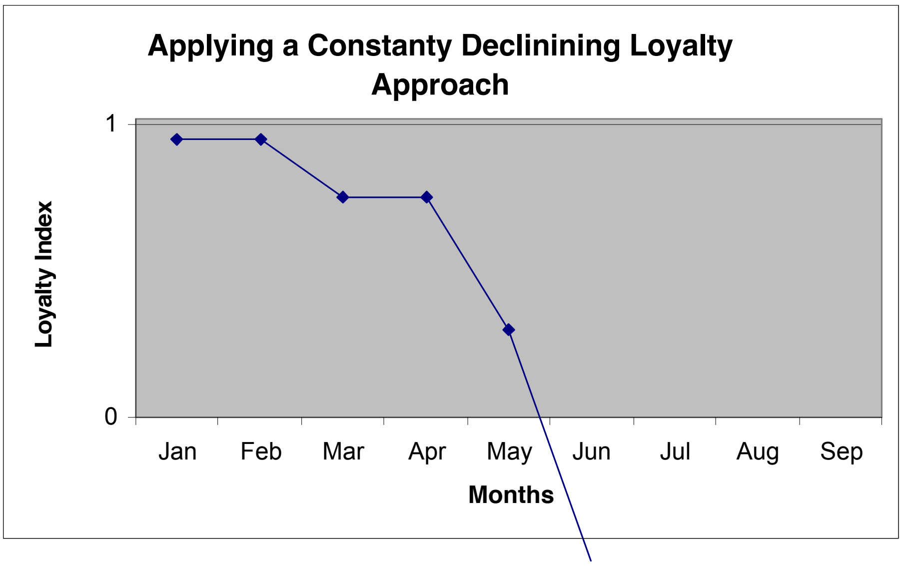


Figure 23: Decreasing Loyalty Values

As can be seen from Figure 23, once the concept of declining loyalty is applied, the customer's loyalty index falls below the level of 0 after May. 0 should be the minimum value so once the value falls below 0 it is set to 0 and any further activity passed this point is disregarded. This means that the example from Figure 21 has been transformed into the example show in Figure 24:

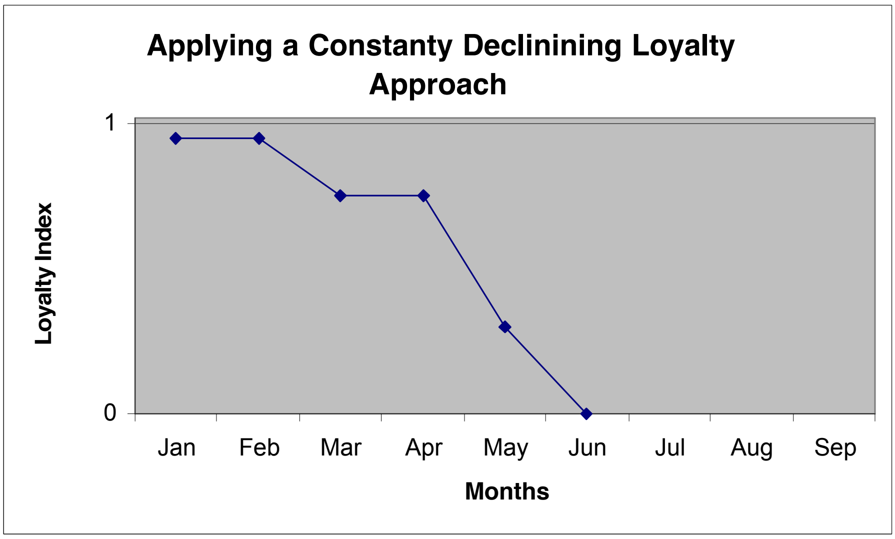


Figure 24: Complete Customer Profile

The final profile is created as displayed in Figure 24.

4.5.2.1. Automating the Constant Loyalty Decline Approach.

The stages of converting the customer profile from the one in Figure 21 to that of Figure 24 is illustrated by the flow chart in Figure 25: -

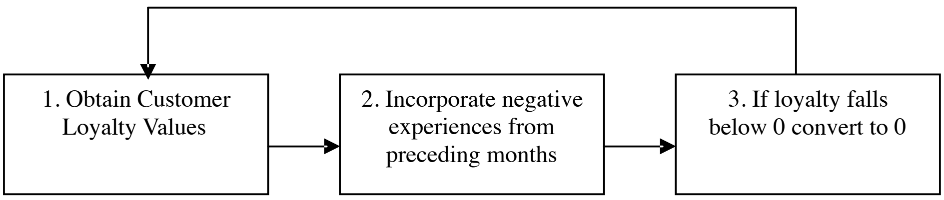


Figure 25: Stages of Calculating Customer Loyalty throughout the Customer Lifetime

Stage 1

The first stage of the flow chart in Figure 25 was described previously. It is the process of estimating loyalty values using the neural network model.

Stage 2

The second stage is the process of creating a declining representation of the customer's loyalty values. The flow chart in Figure 26 shows the rule that was created for this purpose: -

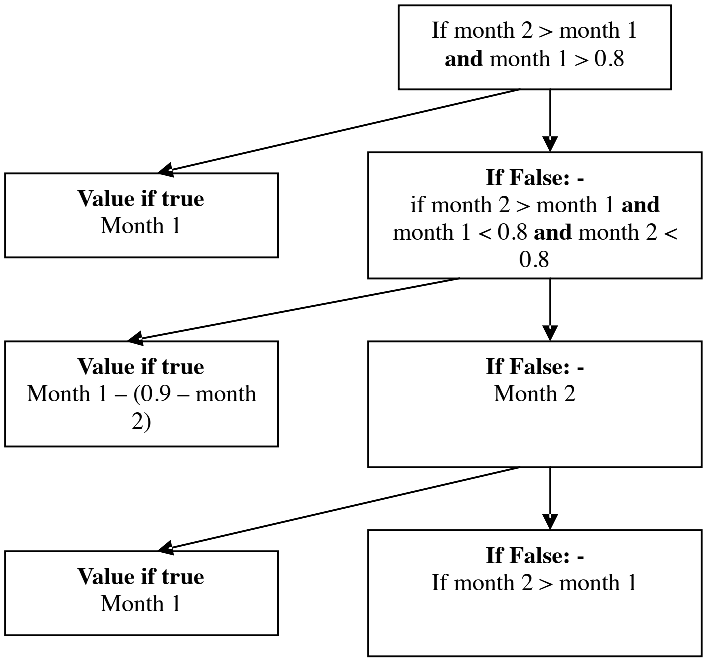


Figure 26: Rule for Calculating Customer Loyalty

Figure 26 describes the rules for establishing the customer's loyalty for each month by account for the values for preceding months. The first part of the rule states that if the second month has a higher loyalty value than the first month and the first month has a loyalty value of over 0.8 then the loyalty value of the first month should be carried over to the second month. The rule has been defined this way because after analysing all the values, the author has concluded that any decrease over 0.8 suggests no negative activity. Fluctuations above the customer loyalty values of 0.8 are only noise. The preceding loyalty value is carried over to the following month if no negative activity has occurred, and if the current loyalty value of the customer higher than 0.8

The second month has a higher value than the first month, and if both months have loyalty values lower than 0.8, then subtract the value of the second month from a maximum loyalty value of 0.9 in order to find the size of the negative activity, then subtract the answer from the value of month 1. 0.9 is regarded as maximum loyalty because due to the nature of the neural network it is unlikely that a customer will ever receive the maximum value of 1.0. Setting the maximum loyalty value to 0.9 provides a degree of lenience for small errors generated from the customers data as a natural result of the NN model. If the values for months 1 and 2 are equal then either value can be carried over to the second month

Stage 3

The final stage of creating the customer profiles can be seen via the flow chart in Figure 27: -

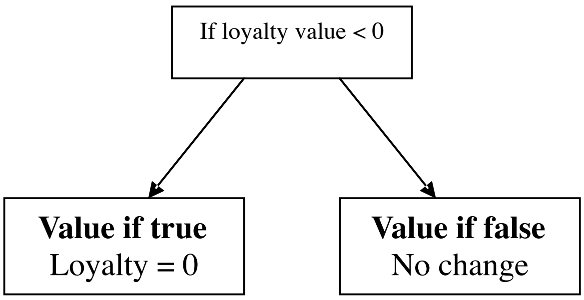


Figure 27: Ensuring Values Do Not Fall Below 0

The stage of the loyalty calculation displayed in Figure 27 is a basic check to ensure that values are not allowed to fall below 0. The rule states that if the customers loyalty is less than 0, replace the value with a 0 else do not make any change.

4.5.2.2. Profile Analysis

300 sample churn customers were analysed manually for profile detection. It was decided that 300 was a manageable number for manual analysis. At this stage of the research manual clustering had to be performed as investing time to develop a software tool to automate the task could not be justified without first identifying the potential of the hypothesis. The data described in section 4.1 was used for this analysis. 10 profile clusters were generated. These profile clusters have graded with the term ‘master profile clusters’. Each master profile cluster was analysed and the results compiled into a two graph representation. The first graph shows the customer’s predicted loyalty

index values over their lifetime with the company. These loyalty values have been processed using the techniques and stages previously discussed.

A second graph has been added to the master profile categories illustrating approximately how long the customers will continue their service with the company. The approximation is a calculation based on all customers belonging to that master profile cluster who have actually churned. It is the average time taken for each customer to churn after their loyalty index value fell below the specified churn threshold. For example, if 8 out of 10 customers churned 2 months after their loyalty index values fell below the churn threshold then the second graph would show an 80% chance of retention after 2 months. The hypothesis made for the second graph is that as customers are successfully identified as belonging to a specific profile they will churn in a similar time frame to previous churners who belonged to that master profile cluster, informing the provider with, for example an 80% confidence that there is time to retain the customers belonging to that cluster. This hypothesis will be further tested in this research.

The profiles had already been subject to a rule to ensure that loyalty was recorded in a continuously declining fashion. Converting the numeric profiles into string representations made it possible to compare and match profiles on a large scale.

A string representation is generated for the profiles for all customers. For example, going back to the profile in Figure 24, as a string it would be represented as *"LDLDD"*.

The letter representations to describe loyalty basically states that loyalty has remained static if represented by the letter "L" or loyalty has fallen, shown by the letter "D". It is apparent that string profile versions should be easy to match. A further decision was made that all master profiles groups should be based on historical churn data only. The logic behind this decision is that the resulting methodology is aimed at determining churn. Therefore a master profile cluster should only exist if actual churn is connected to that profile.

All profiles belonging to all actual churners were manually compared for similarities. All similar profiles were then grouped, and all unique profiles recorded as individual clusters. Four master profiles (1 – 4) are detailed as follows to help in an understanding of the analysis:

Master Profile 1

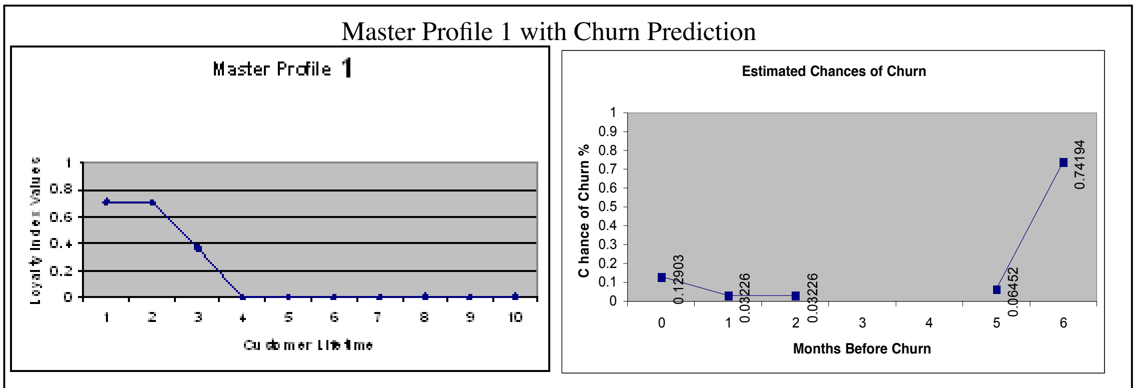


Figure 28: Master Profile 1(Initial Experiments)

The first master profile seen in Figure 28 illustrates those customers who churned after two continuous events, bringing the loyalty index value below the churn threshold of 0.3. Out of the 300 customers analysed 31 fell into this category. The estimated chances of churn show a 13% chance that a customer will churn immediately, a 3% chance the customers will churn after 1 months, a 3% chance the customers will churn after 2 months, 0% of customers churned during months 3 or 4, 6% of customers churned after 5 months and a 74% of customers churned after 6 months.

Master Profile 2

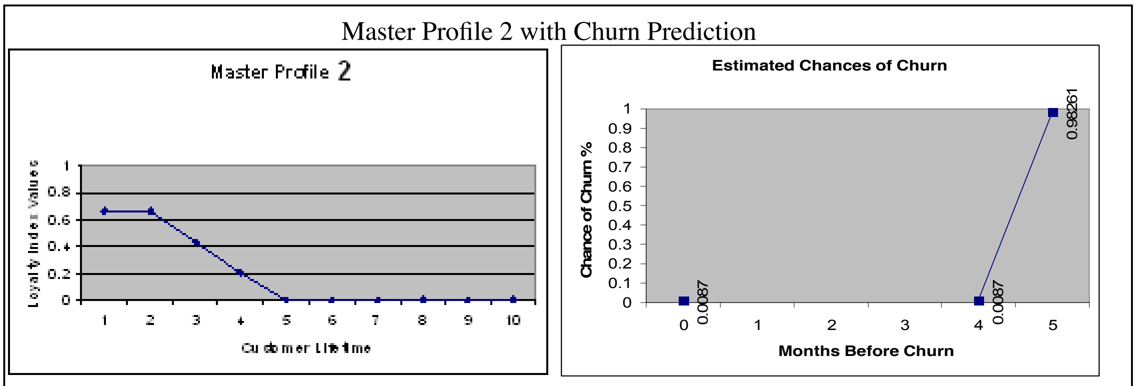


Figure 29: Master Profile 2 (Initial Experiments)

Master profile 2 is displayed in Figure 29 and illustrates those customers who churn after three consecutive events. Out of the 300 customers analysed 115 customers fell into this category making it the most dominant of all categories. An interesting observation of this category is the fact that more than 98% of all customers that have churned with this profile have churned after a five month period. Only less than 1% have churned immediately and less than 1% has churned in 4 months. This means that if a customer is predicted to fall into the category of master profile 2, the company has a maximum 5 month window for retention efforts to be deployed.

Master Profile 3

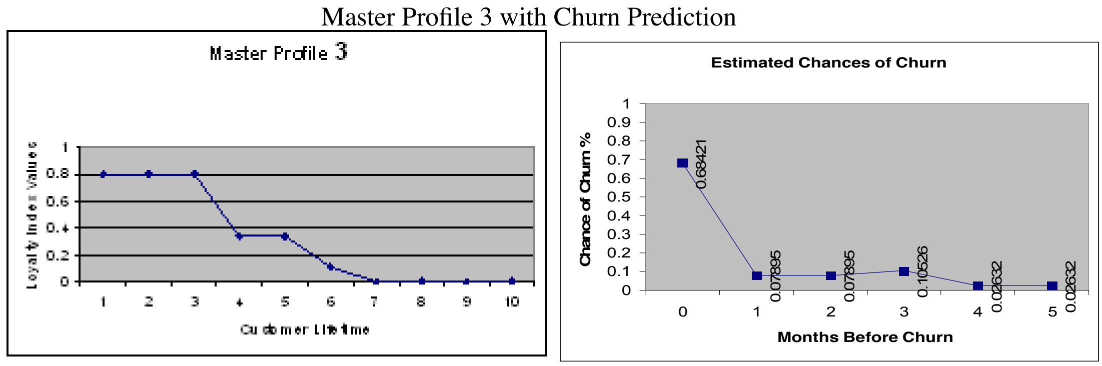


Figure 30: Master Profile 3 (Initial Experiments)

Master profile 3 seen in Figure 30 is different from the other master profiles. All previous master profiles show a consecutive number of events. This profile shows events in steps, e.g. one or two months in a row show a decline in customer loyalty then loyalty will be stable for the next month and then begins declining again. Master profile 3 shows customers whose loyalty declines, however once during the fall towards churn the loyalty stabilises. From the 300 customers analysed, 38 fell into the category of master profile 3.

It is apparent from the estimated chances of churn that customers who fall into the category of master profile 3 are more difficult to predict for churn. 68% of these customers churned immediately, nearly 8% churned after 1 month, nearly 8% churned after 2 months, over 10% churned after 3 months, over 2% churned after 4 months and over 2% churned after 5 months. Although companies may find it difficult to retain customers that fall into this category, this research provides them more insight to channel their customer retention efforts.

Master Profile 4

Master Profile 4 with Churn Prediction

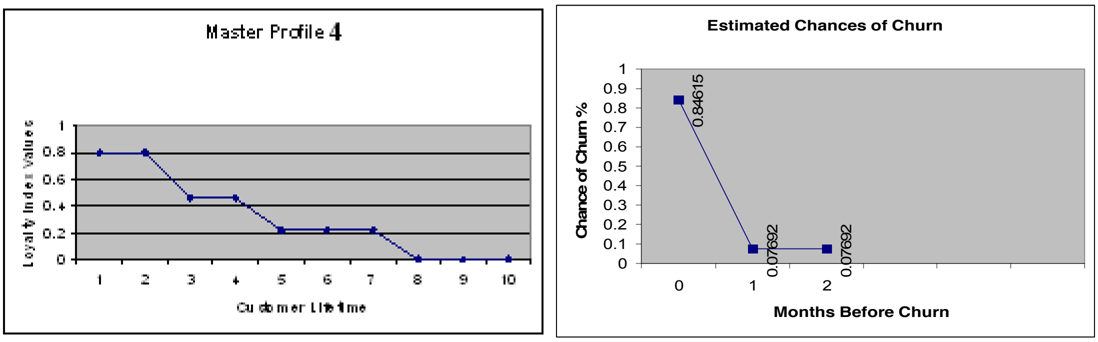


Figure 31: Master Profile 4 (Initial Experiments)

Master profile 4 seen in Figure 31 is similar to master profile 3 however customer loyalty values stabilise twice in their decline towards churn. Out of the 300 customers, 13 fell into this profile. Again it would be difficult for the company to retain the customers that fall into this category because nearly 85% of them churn immediately. Only 7% churn after 1 month and 7% churn after 2 months. The validation of the proposed methodology is detailed below.

4.5.2.3. Initial Validation

Initial validation for the dataset was performed by manually matching a validation set of 500 customers was constructed using the profiles of 250 churn customers and 250 non-churners against the identified master profile clusters from the model creation stage. This experiment expected to find that the majority of the churn customers would match the master profile clusters while the majority of non-churn customers would not. The results for the dataset ended with 415 customers being classified to churn profiles. This resulted in a large misclassification rate; however further investigation suggested that the main reason for the massive misclassification rate was the constant declining loyalty index. Most customers encountered some type of decline in loyalty within the monitored time period. The loyalty index had been restricted to a constant decline with all negative activity being deducted from all previous negative activity. A large portion of the customer base was being captured as churners because the sum of all the small declines in loyalty added up to a reduction in the customer's loyalty index that fell below the churn threshold.

4.5.3. Evolving the Profiling Methodology

It was apparent from the initial experiments that an alternative strategy was required. Attempts were made to enhance the results of the profiling methodology, by creation of a hybrid model. This idea came from referring back to Table 7 where it was identified that NN generated the best results for predicting churn while linear regression was best at predicting non churn. It was attempted to develop a strategy where the linear regression results and NN results could be used in partnership to counteract the misclassifications, however these experiments failed. It was found that the results from the linear regression model cancelled out accurate churn results from the NN model resulting in a poor churn capture and high misclassification rate.

The reason why the initial profiling analysis did not work is analysed here. When the NN generates churn index values it does not regard any previous activity, and treats each new month as an entirely new case. Trying to force the churn index to remember previous activity by applying a rule to disallow loyalty to increase once it had decreased has not worked. Loyalty therefore has to be allowed to fluctuate with activity between months being treated as independent events and it should be these independent events that are used to build customer loyalty profiles. Again referring to the example profile in Figure 21, the profile is created directly from this original state before the constant decline rule has been applied. Converting this profile to string format it becomes *"DUDD"*. The string loyalty profile has a third letter representation from the previous example in the form of the letter "U", representing an increase in loyalty. It should be noted that the last two months are disregarded because the profile falls below the churn threshold in month 5 so at that point the customer coupled with this profile is flagged as a probable churner. Any further activity is therefore deemed irrelevant. The churn threshold can be regarded as classification criteria. It was reported in the literature survey that popular classification techniques involve classifying all customers as churn with a loyalty value below 0.5. 0.3 was selected for this research as investigations showed that most churners had loyalty values of 0.3 or lower, so 0.3 would maintain a high level of accuracy while helping to control misclassification rates.

Tests were performed using the same data used in the initial experiment. The results from this methodology were much better with a high correct churn classification achieved while misclassification rates were reduced by a massive amount in comparison with the initial experiments. These results are shown in

```
        | Estimated Labels
 True   |
Labels  |    0       1   | Totals
--------|----------------|-------
0       |    84     166  |  250
1       |    88     162  |  250
--------|----------------|-------
Totals  |   172     328  |  500
```

Figure 32: Rising Loyalty Experimental Results

As can be seen from Figure 32 the initial tests displayed 64.8% churn capture accuracy, 33% non-churn capture accuracy.

Although better than the initial theory that included a constant decline in loyalty index values, the misclassification was still too high. To determine exactly what was happening with the profile clusters the actual results for each profile were broken down so that a clear understanding could be gained. It was identified from this analysis that not all profiles could be regarded as churn sensitive. In fact most of the captured churn was captured within only 3 master profile clusters. The remaining clusters contained a very small number of churners but accounted for the majority of misclassifications. Therefore it became apparent that clusters should be further sub-grouped into high risk and low risk churn groups. The 3 groups that contain the majority of churn are as follows:

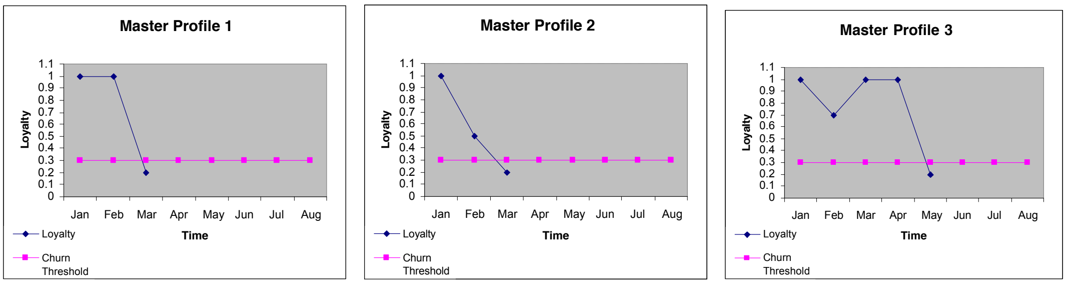


Figure 33: Significant Master Profiles

The most significant master profile from the three shown in Figure 33 was master profile 1. 87 customers were matched to master profile 1. 53 of these customers actually churned 2 months in the future. 59 customers were assigned to master profile 2 and 54 actually churned. 64 customers were assigned to master profile 3 and 37 actually churned. Therefore it can be determined that from these 3 profile clusters alone 149 churners were captured who actually churned, while the total cluster sizes of these three profiles combined equals 210. Therefore if just these three profiles are used to predict churn the misclassification rate is reduced to 61 while the churn capture would is only reduced by 13 customers only. Therefore 59.6% of the churn that occurs has been captured, while the misclassification rate has been reduced to a minimum. The stages of the methodology were documented as follows:

4.5.4. Summary of the Methodology

A summary of the stages of the customer profiling methodology are as follows:

Prepare the data

- Ensure the data is numerical, and has no missing or incompatible values
- Take a time sequence of customer data and split the data by customer, by month.

Determine the most suitable for creating the neural network model

- One month of data is required for training the neural network. The training set should contain as many churners as possible so an analysis of the available data should be performed in order to determine which month contains the highest churn. This month should be used as the basis for creating the training dataset.

- Based on the total number of churners in the chosen training month randomly eliminate non-churners from the training set to create a 20:80 churn/non-churn ratio.

Create a neural network suitable for generating a loyalty index

- Using the training dataset create 3 neural networks (1 neuron, 1 layer), (3 neuron, 2 layers), and (6 neuron, 4 layer). It maybe necessary to continue these experiments increasing the number of neurons in each layer?
- By applying each of the NN architectures to the full month that the training set was originally based on, and then comparing the predicted results against the actual churn data, it is possible to determine which NN architecture has converged best. The NN displaying the best convergence should be used for generating churn index values for all other available months.

Generate churn index values

- Using the best neural network, generate churn index values for each individual month for each customer.
- Compile a database containing all churn indices for each month and all churn information.

Apply customer profiling methodology

- Analyse churn indices to establish master profile classes and customer classification into the determined master profile classes.

Identify the high and low risk profile clusters

- By determining how many of the customers contained within each profile cluster have actually churned it is possible to rate each cluster into high-risk and low-risk churn groups.

Determine churn capture accuracy.

- Comparing the high-risk profile clusters with future churn data provides a determination of how many future churners have been captured from the resulting methodology. The profiles that have been determined to hold the highest portion of future churn are classified as high risk clusters. The profiles that have small or zero future churn capture classified as low risk clusters. The time frame for the data being analysed is shifted forward. In business the time frame should move on a monthly basis; however for the research it will move forward 3 months to minimise the risk of duplicate capture. This risk would not be a problem in industry as once a customer had churned they would be automatically be removed from the dataset being analysed. The high risk profiles can then be used as future churn predictors and all customers being assigned those profiles classified as churners.

Methodology Flow Chart

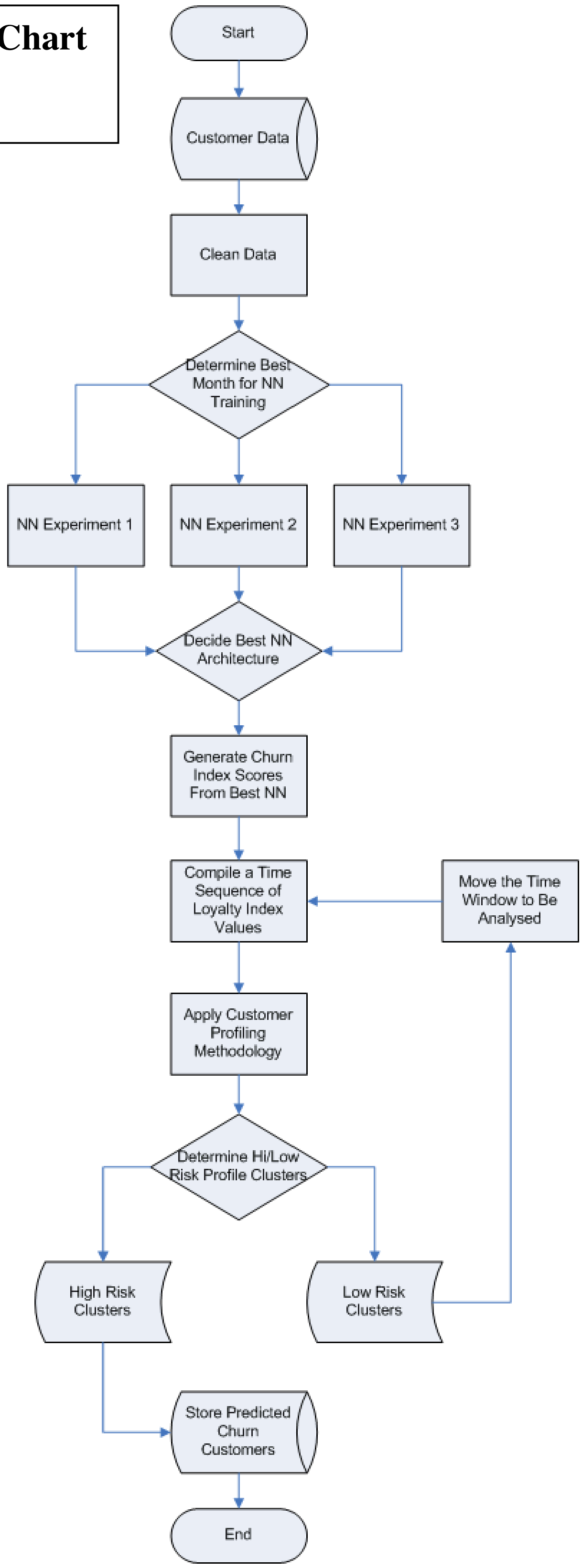


Figure 34: Customer Profiling Methodology Flow Chart

All identified research gaps have been addressed by the customer profiling methodology.

Achieve the maximum possible future prediction.

- The profiling methodology has proven to be capable of predicting churn in the future with the ability of matching customers to master profile clusters to enable early classification, increasing time between classification and the actual churn event. The customer profiling methodology uses several months of data to determine customer churn rather than just 1 month as commonly seen in research. Because of this time window a system could easily be implemented in the form of a continuously moving time window capable of classifying customers to master profiles and flagging churn in advance.

Base the framework on data that is available across multiple service sectors and accessible for analysis by any business regardless of size or monopoly status

- Predictions have been based on customer repairs and complaints data. The benefits of using this type of data source are that it is available for use by any company regardless of the monopoly status and it has proved to be a good source for basing accurate future predictions. The reason this data has no monopoly regulations attached to it is because it is strategic to the business. Churn connected to repairs and complaints is strategic to the specific business. It is the identification of customers who are contemplating defecting because they are not happy with their service so the business is responsible for their dissatisfaction.

Minimise the total number of misclassifications from the predictions to reduce retention costs to the business

- It has been demonstrated that a lot of thought has gone into how misclassification rates can be improved. The customer profiling methodology achieves this task by eliminating customers who fall into weaker churn categories.

4.5.5. The next stage

With the profiling methodology created the next stage is to see if it is possible to implement this methodology into a software based platform. The next chapter documents the process of converting the methodology into a physical application. This software application will then be used to validate using three case studies so as to determine the full power of the methodology

This Page is Intentionally
Left Blank

5.Developing a Prototype System for Churn Prediction

This chapter documents the development of a software prototype for the advanced capture of future customer churn. This software prototype is required to aid in the following:

Analysis of the overall performance of the methodology

The practicality of the methodology

Large scale testing and validation of the methodology

Implementation issues regarding the methodology

The developed software prototype should support all aspects of the proposed methodology. The proposed prototype is expected to incorporate the following features:

Be generic so that it can be applied to multiple data sources

Read data directly from a database for large scale analysis

Offer a user friendly interface

Provide significant decision support regarding churn predictions

5.1.Software Development Using XP

The XP software development stages, and descriptions throughout the remainder of this chapter, have been derived from (Baird, 2003). XP incorporates 4 fundamental principles, communication, simplicity, feedback and courage. Communication is a core value for all software development methodologies. With XP communication is oral instead of the lengthy construction of documentation, reports and plans. Simplicity in XP is defined as 'doing the simplest thing that could possibly work'. XP focuses on what is critically needed right now, and not what *might* be needed. A saying in XP development is "you aren't going to need it" which has been abbreviated to YAGNI.

Feedback with XP is performed by developing software quickly and then demonstrating it to the customer to receive feedback on functionality. This is achieved through the process of 'nagging' the customer with the same questions over the state of the system to build concrete feedback. Finally courage is the confidence to work quickly and develop if required. Courage should be thought of in the context of communication, simplicity and feedback. Without these three values courage leads to chaos. The stages for developing an XP project are as follows:

Exploration
Planning
Iterations
Productionising
Maintenance

5.1.1. Use Case Diagram

Figure 35 displays a use case diagram, designed to illustrate how the software prototype would implement into typical business architectures:

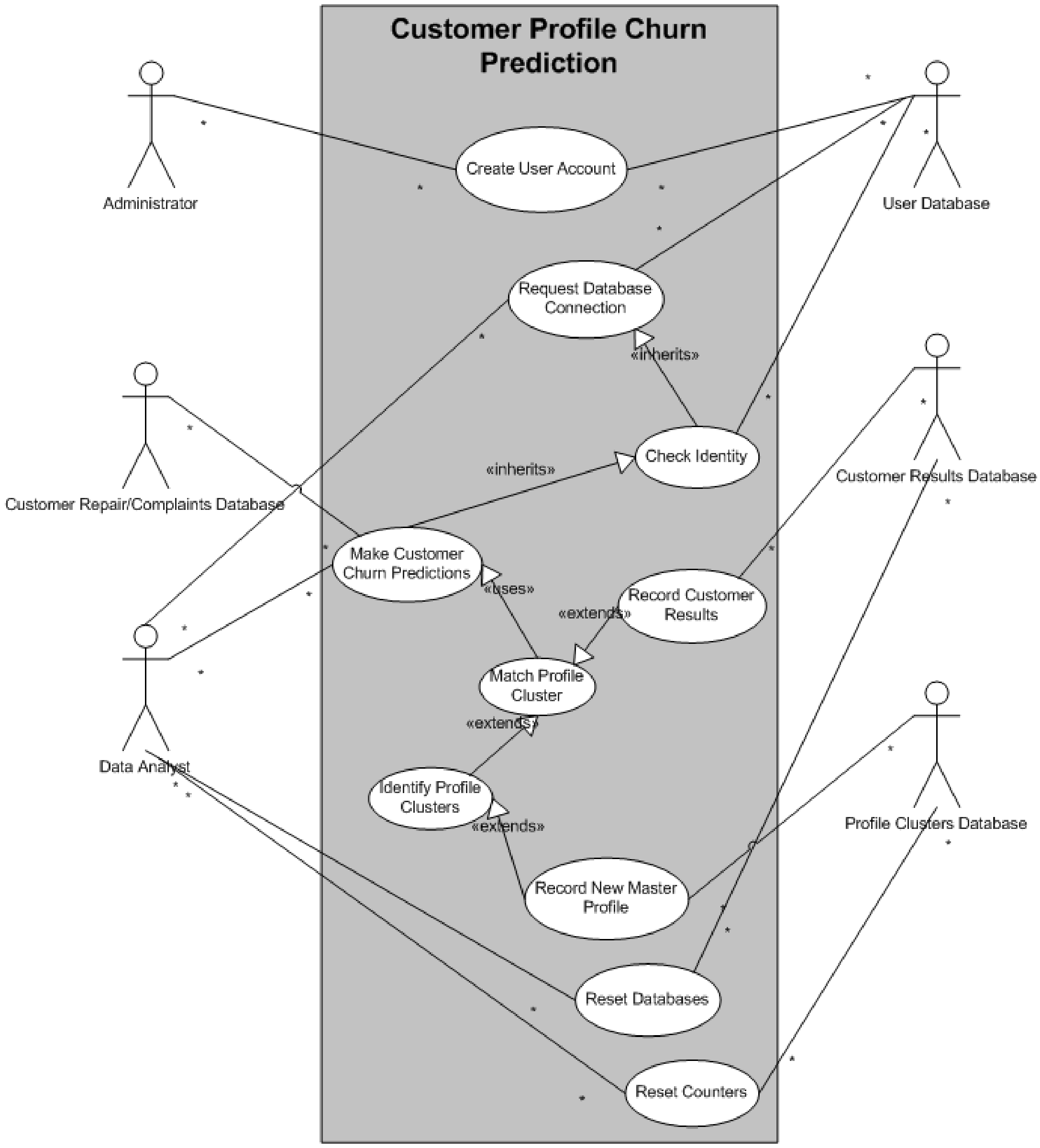


Figure 35: System Requirements Based User-Case Diagram

5.2. Programming the Prototype System

Software development has followed the XP (Extreme Programming) software development methodology. XP focuses on development using a staged release model. Basic functionality is developed in the first release and the complexity of the system is then gradually built through various further releases, until the software includes all required functionality. To ensure this chapter remains understandable no code will be printed. Instead a detailed description of the development including all challenges will be reported.

A decision was made for the first stage of software development to provide the basic software frame including database connectivity. This task can be broken down as follows:

5.2.1. Release 1 – Database Connectivity

Tasks :

Define a desired database connection

Capture the desired time frame to be analysed

Define connectivity to a MySQL database as the source

Automate input as much as possible to minimise the risk of input error

Store the user input for recall as and when needed by the system

Test the database connection to ensure stability.

Development performed using the Java programming language. This is the language the author is most comfortable using. The familiarity with the development environment and the fact that Java offers a visual development environment through JBuilder should ensure that development is completed in the shortest possible timeframe. The resulting software release for stage 1 is illustrated in Figure 36:

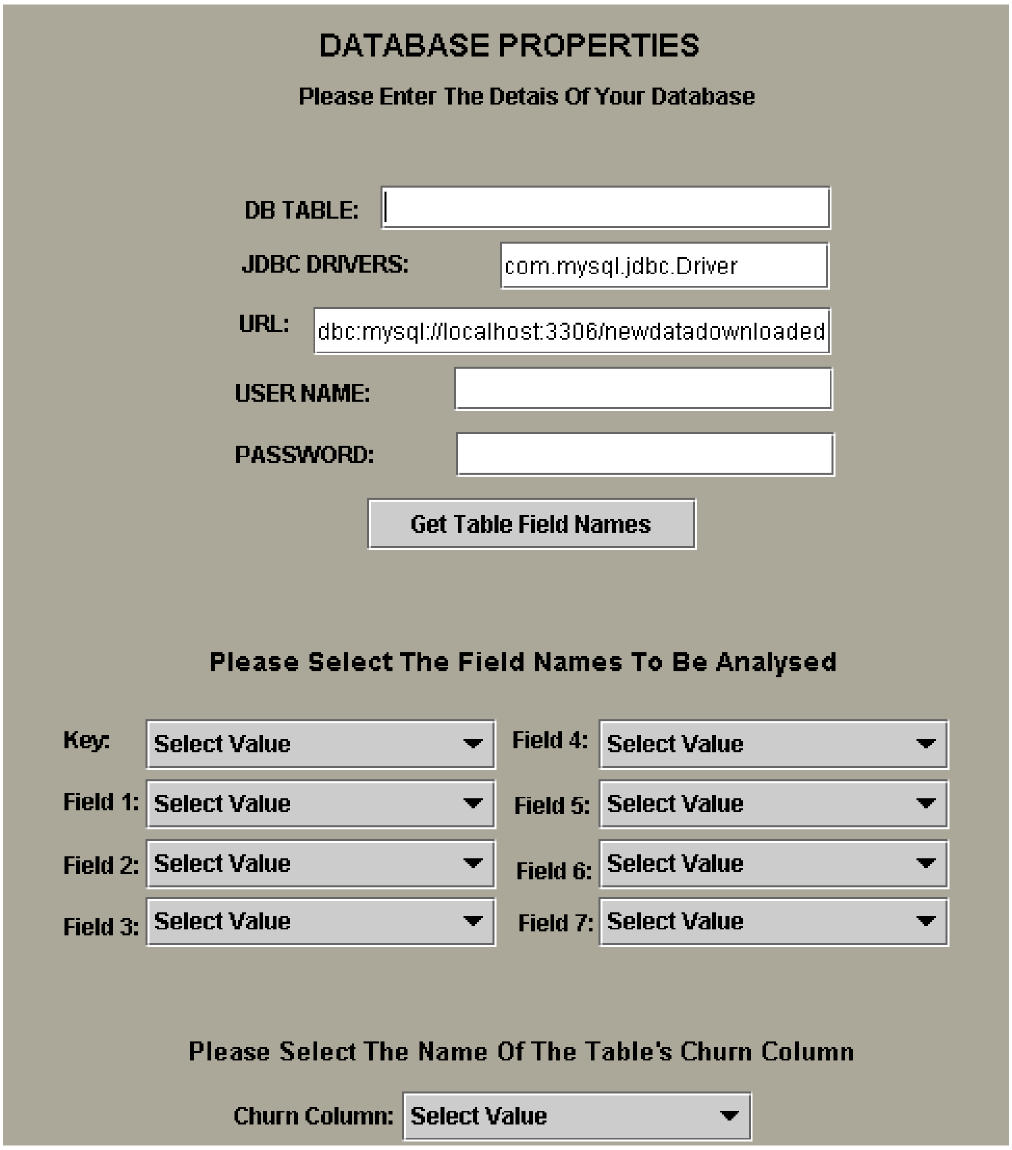


Figure 36: System Database Connection Interface

As can be seen from Figure 36, the properties input box requires several inputs. The first input, labelled 'DB Table' is a part of the URL connection string. The URL connection string is a requirement of the connection/j class as it is this string that instructs the class which database is to be connected with. The 'DB Table' definition part of the connection string has been isolated and given its own input box to make specifying a table less cryptic. The JDBC driver tells Java which driver should be used to handle the connection. Java can handle connections to a multiple database types so the target database type has to be declared. The URL specifies where the database server is located and which database found on the server should be connected to.

Finally MySQL databases require a username and password so these fields are provided for restricted access to the database system.

Once all desired inputs are declared the 'get table field names' button can be selected. This button has two objectives. The first objective is to test that a connection can be made to the database based on the user defined input. The second objective is to populate the database field boxes as seen in the lower half of the database properties window shown in Figure 36, based on the database metadata.

The next step is to find a way to store the input data in such a way that it can be recalled by the rest of the system for analysis. It was decided to save the data in an application 'properties' file. It can be seen from **Error! Reference source not found.** that a 'Save Properties File' button has been added to the GUI. Once the save properties file is clicked all data from the information boxes is written to the .PROPERTIES file and stored directly to the root drive as shown in Figure 37:

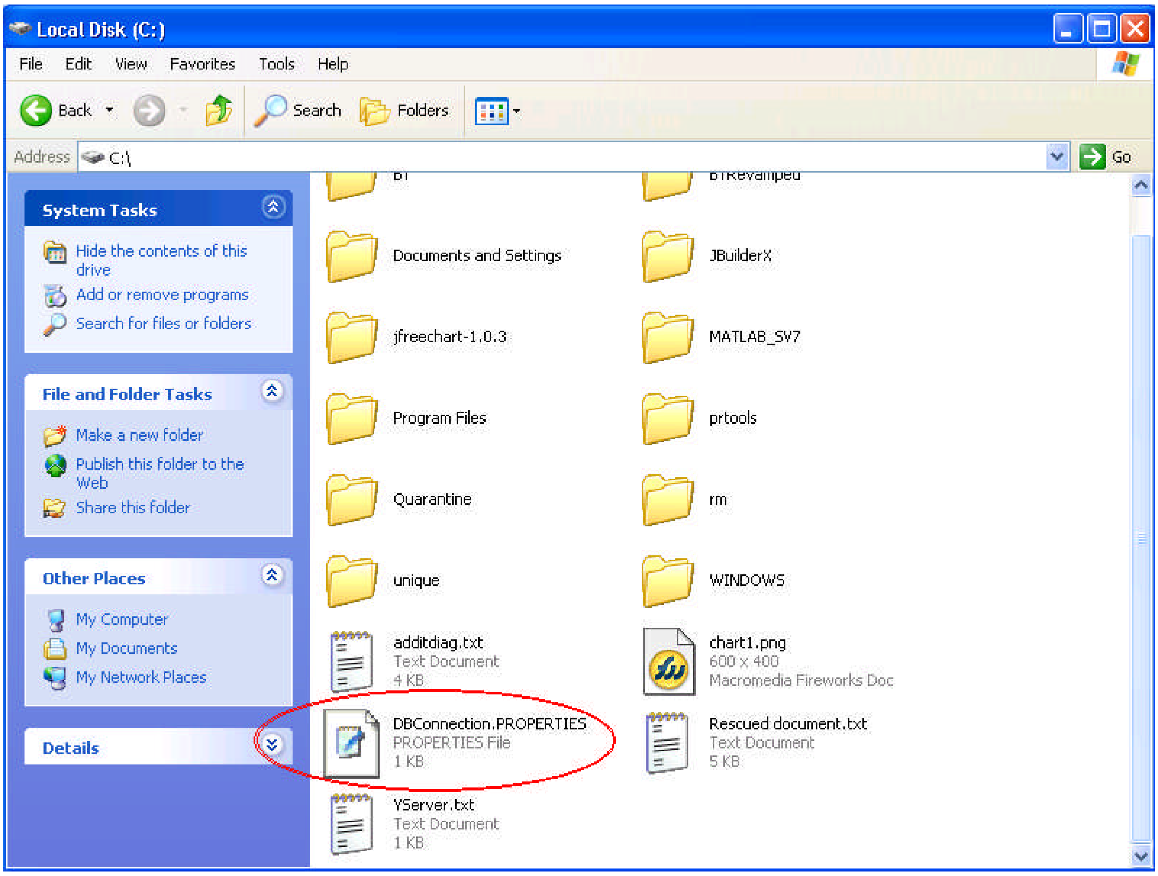


Figure 37: Properties File

The properties file shown in Figure 37 is a text file and stores the database connection properties as string values. An example of the data stored in the properties is shown below:

```
jdbc.drivers=com.mysql.jdbc.Driver
jdbc.url=jdbc:mysql://localhost:3306/valdb1
jdbc.username=root
jdbc.password=joshua
jdbc.table=dataset1_loyalty_inputs
jdbc.table.key=ID
Field1=Oct-04
Field2=Nov-04
Field3=Dec-04
Field4=Jan-05
Field5=Feb-05
ChurnColumn=ChurnMonth
FieldCount= 5
```

The .PROPERTIES file can be accessed and read by the software anytime, enabling a connection with the database whenever required. Now that a method has been created to store and recall a connection to the defined MySQL database the first software release is complete. Work can now begin on the second software release to enhance functionality. The flow chart shown in Figure 38 has been provided for this software release:

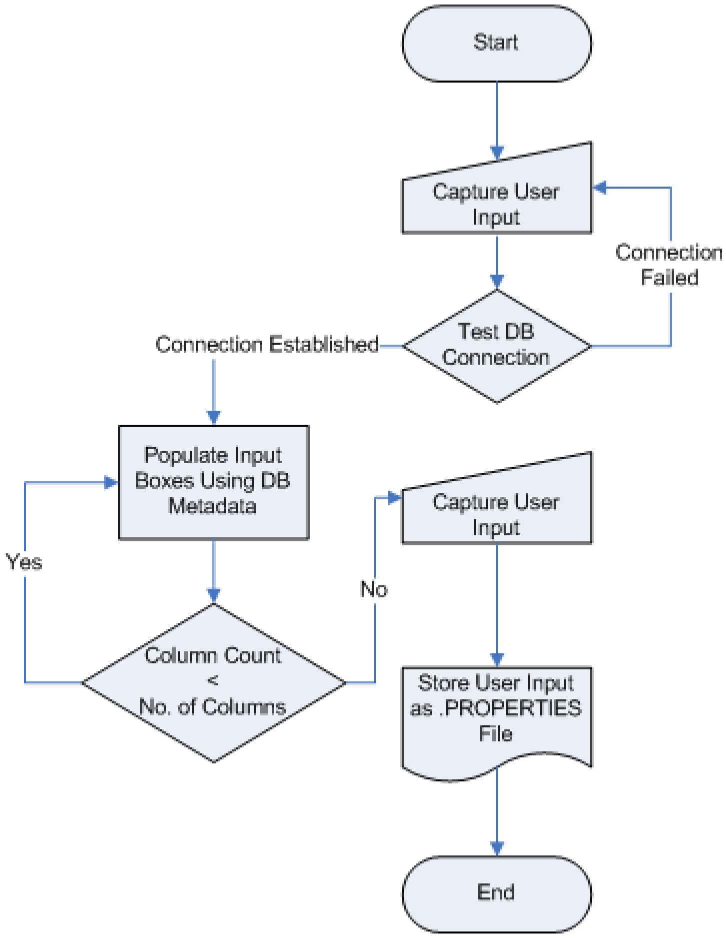


Figure 38: Software Release 1

5.2.2. Release 2 – Churn Index Analysis

The next software release is to include functionality that will enable the creation of loyalty index profiles. There are some challenging aspects of this stage and it will provide functionality rather than a visual form for requesting information.

Tasks:

Read input from the database

Provide support for analysis of a single customer or the full customer base

Initiate analysis through user friendly interface

Inspect the input and create string representation values

Display the customer profile in a graphical representation

Request user profile chart through user friendly interface

5.2.2.1. Read Input from the Database

The first requirement for providing the functionality of churn index analysis is establishing a connection to the database. In the first software release a database connection was defined and stored in a .PROPERTIES file. For the second release, the first stage is to read the values stored in that .PROPERTIES so a connection to the desired database table can be achieved. The steps required to read these values are as follows:

Define a file input stream – To enable reading information from the required .PROPERTIES file.

Define a Properties object of the Properties class – The Java class library contains a class type called 'properties'. The properties class represents a persistent set of properties. The main property methods used in development are:

- GetProperty() – searches for the desired property, defined as a string value.
- Load() – Reads the contents of a .PROPERTIES file as defined by the file input stream.

Define a record set using the MySQL class – The MySQL class types are included with the connection/J library as discussed in **Error! Reference source not found.**. This record set can be used to define and execute SQL statements direct from the code and store the results.

Read the contents of each field in the database in to a variable for use by the software. Convert the churn index value into a loyalty index value by subtracting the field value from 1.

Using the steps listed above, a connection has been created to the defined database table and the contents of that table have been assigned to individual variables which can be used freely by the churn index analysis class. This concludes the process of retrieving values from the database. These values can now be used to determine string representations of the loyalty index values.

5.2.2.2. Support for Single Customer or Full Customer Base

It was thought that a useful feature would be to allow analysis of a single customer or the full customer dataset. It is possible that some situations may arise where it would be inappropriate to analyse the full dataset every time the software runs, such as testing. Testing would be very slow and tedious if the full customer base had to be analysed to test new functionality.

It was decided to construct two separate user interfaces to define how the customer base should be analysed, and provide a mechanism to initiate the analysis process. The first user GUI is for single customer analysis and the second GUI is for the analysis of the full customer base. The single customer GUI is displayed in Figure 39:

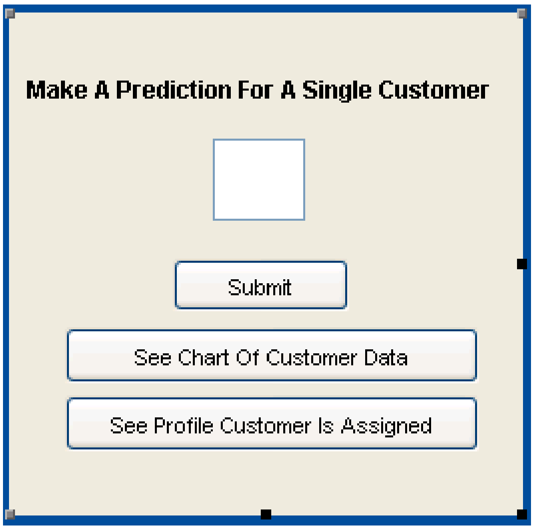


Figure 39: Single Customer Analysis GUI

The single customer analysis GUI is shown in Figure 39. The white box is the input field where a customer number can be entered. It can be seen that support has been provided for various aspects of customer analysis. The first feature is an input box. This input box has multiple functions. The first function is to define a customer to be analysed by the methodology which at this stage means determining a string representation of that customer's loyalty index values. The second feature is to define a customer from the analysed customer base for viewing that customer's loyalty profile. The third feature will be available later in development and it is to show the master profile cluster that a particular customer has been assigned to. The text box is used as input for all three functionalities. The command buttons define which functionality to call.

A GUI was also developed for the analysis of the full customer database. Separate GUI's were developed for single customer analysis and full database analysis so that functions could be grouped and any confusion in user interaction could be avoided. The GUI for analysis of the full customer base can be seen in Figure 40:

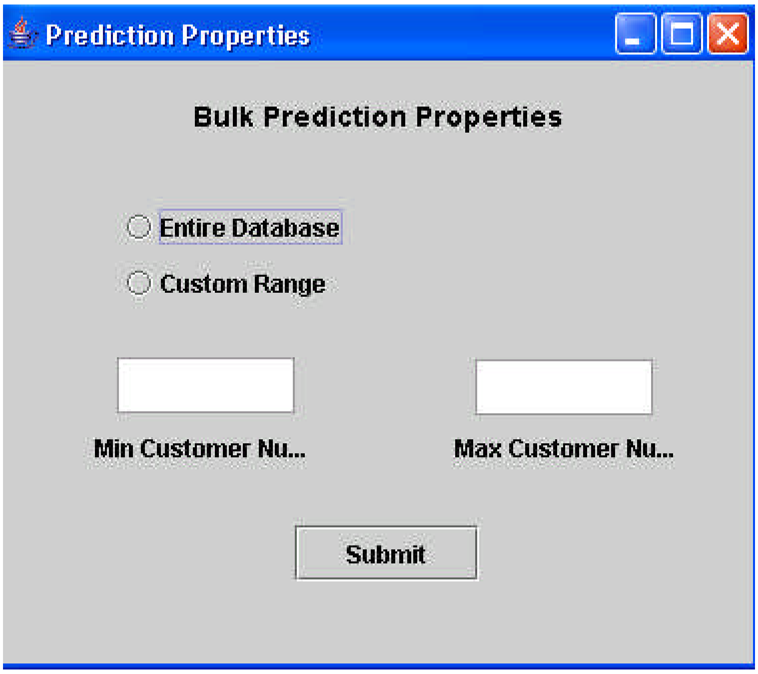


Figure 40: Full Customer Base Analysis

It can be noticed from Figure 40 that analysis of the full customer database can be performed in two ways. First the full database can be defined for analysis by selecting the entire database radio button and hitting the submit command button. Or a user defined range can be input. This is achieved by selecting the customer range radio button and hitting the submit button. The custom range feature was included because its benefits were substantial in enhancing the testing process.

5.2.2.3. Creating String Representation Values

In order to create string representations of the loyalty index values the first stage is to determine if there has been an initial event that has had an impact on customer loyalty. The initial is directly used as the first input into the string calculation.

The second loyalty index value is analysed to see if it is smaller than the first string value. If it is smaller it can be determined that there has been an impact in loyalty and the value '*D*" is recorded as the initial string value. If the second loyalty index value is larger than the first value it can be determined that loyalty has risen and the value "*U*" is recorded as the first string value. If the second loyalty index value is equal to the first loyalty index value it can be determined that loyalty has remained constant and the value "*L*" is recorded as the initial string value. The process repeats for all months to be analysed by the profiling methodology. At this point a string representation of the customer's loyalty index over a time series has been determined. This release can be seen from the flow chart in Figure 41:

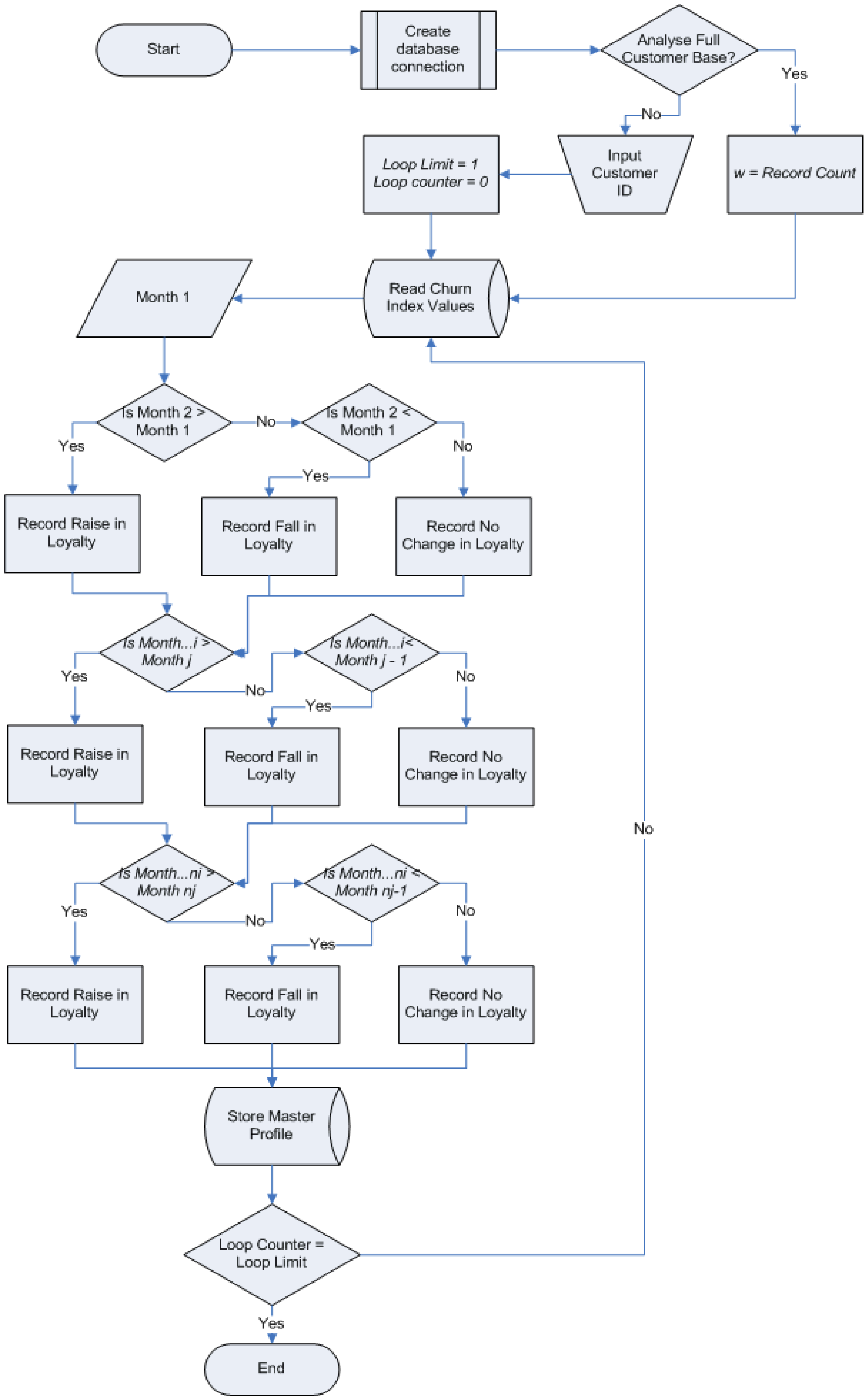


Figure 41: Software Release 2 - Churn Index Analysis

The third software release requires a method for showing a customer's loyalty over the time period in the format of a graphical representation. The stages carried out for this release are documented in the following section:

5.2.3. Release 3 - Graphical Representation of Customer Loyalty Profiles

This stage of development required some research development was done using Borland's JBuilder. JBuilder does not provide any class types for creating charts so the decision had to be made if it would be efficient to manually code a chart class for drawing a chart; or if a class package could be acquired from a third party source?

Tasks:

Identify a third party package for generating charts using Java and JBuilder
Determine how to use the identified package
Implement a chart class in the prototype system
Use the chart class to plot customer loyalty values
Display the chart to the screen

It was decided that manually coding a class for drawing a chart would be very time consuming so investigations began on identifying third party Java chart class packages. There are several solutions available for building charts directly from Java software, some packages need purchasing but there are also free packages available. The Java package JFreeChart is available from http://www.jfree.org. The class package is free; however documentation about using the class packages requires purchase. This documentation was not required. Learning how to use the JFreeChart classes was achieved by the process of trial and error. The JFreeChart class packages are open source so the software packages support many chart types with a large amount of chart functions that can be manipulated based on user definition. A complete description of all JFreeChart classes and functions is beyond the scope and interest of this chapter; however the parts of the JFreeChart package used for providing graphical representation of the customer loyalty profile are as follows:

Read required values into a category dataset as part of the JFreeChart class:

Create a chart object from the desired type

Set compulsory components for a line chart:

Create a chart frame object (the window from which the generated chart will be displayed)

Create a chart plot object – a general plotting class that takes data from the category dataset and renders each data item into the chart.

Create a value axis object – to customise axis settings including the desired axis ranges.

The above components are requirements for constructing a JFreeChart. There are a huge variety of chart types available. For the purpose of displaying a customer loyalty index a line chart is defined. An example of a customer profile drawn using JFreeChart can be seen from Figure 42:

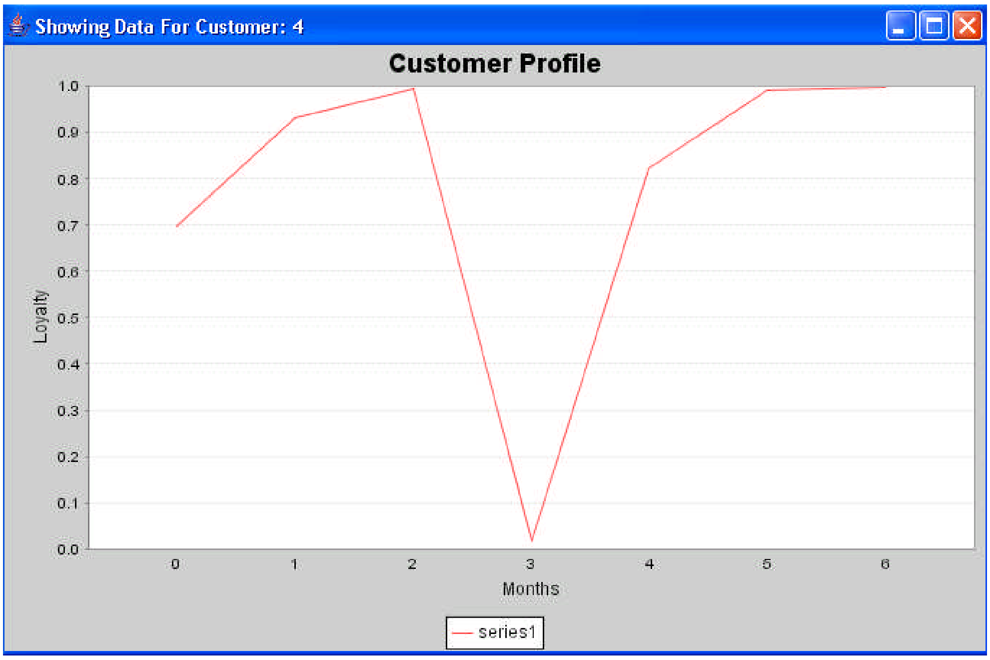


Figure 42: JFreeChart Line Chart – Customer Profile

As shown from Figure 42 the JFreeChart classes can be used to generate professional charts with flexibility for user requirements. With the string representation of the customer's loyalty index determined and the capability of viewing the customer profile as a chart coded, the third software release is complete. This release is presented by the flow chart in Figure 43:

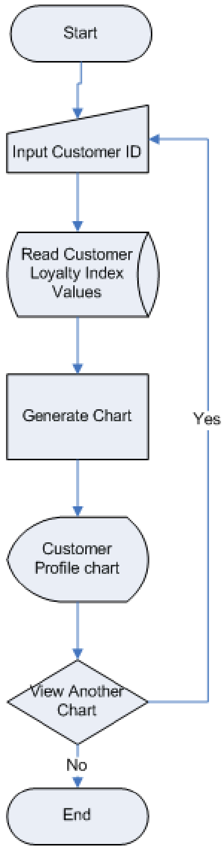


Figure 43: Software Release 3 - Customer Loyalty Chart

By defining interaction with the database, converting the customer loyalty index into a string value and generating a graph of a customers loyalty index values, the third software release is complete. Work is ready to begin on the fourth software release.

5.2.4. Release 4 – Customer Profile Analysis

The fourth software release deals with the problem of comparing customer loyalty profiles. Comparing customer loyalty profiles is not a straight forward task. Fluctuations can begin anywhere along the analysed time line, so a simple task of comparing two string values will not catch all customers who match a specific profile.

Tasks:

Determine a method of removing redundant data

Determine a method of capping the profile after the churn threshold has been met

Record a string representation of a customer profile with the above two points addressed.

To illustrate this, two customer loyalty profiles are shown below for visual comparison. Both profiles should belong to the same cluster group; however as seen from Figure 44 there are some challenges in getting the software to recognise this:

Figure 44: Customer Loyalty Profile Comparison

Comparing the two customer loyalty profiles in Figure 44 it can be identified that they should be classified as a match. The problem is that the decline in loyalty begins at different points along the time line. For the profile on the left, the customer profile shows no event within the first month, no event within the second month, no event within the third month, a slight decline in the fourth month, and a major decline in loyalty in the fifth month that falls below a threshold of 0.3. During the creation of the methodology a value of 0.3 displayed continuously good results for classifying churn, so this will be the threshold set in the software. Any activity that occurs after the detection of churn is considered redundant and should not be regarded for comparison of profiles. The problem that exists is; how should the software be constructed so that it will consider the profile on the left of Figure 44 as the same profile as that on the right?

The first stage for analysing the profiles is to remove any leading information that does not benefit the profile. It can be seen from Figure 44 that the profile on the left contains 3 months of initial data that shows no events. This information does not benefit the customer profile so the software needs to be told to ignore initial zero activity. The process of removing any leading zero activity is defined in the flow chart in Figure 45:

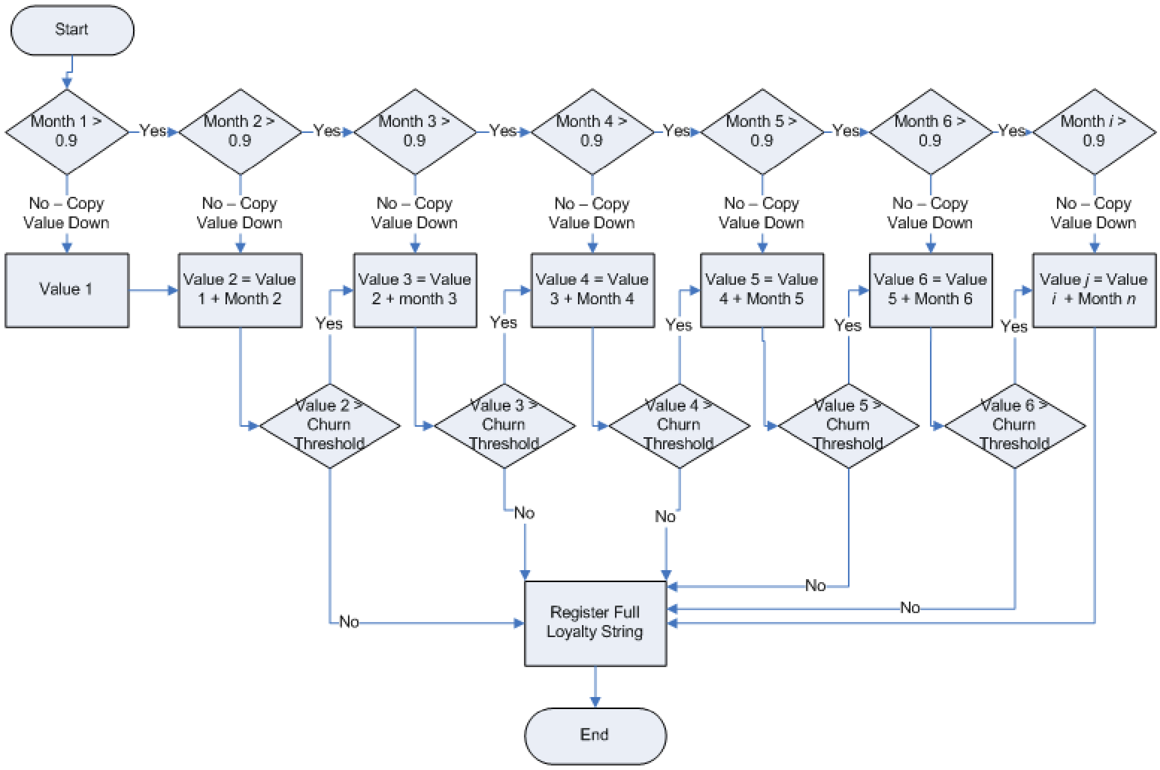


Figure 45: Creating a Loyalty String for Profile Comparison

As shown by the flow chart in Figure 45, the solution to the above problem was identified as checking the first value to see if any event had occurred. If the value of the first month is greater than 0.9 no event is detected so the program moves to month 2. If month 2 has a value below 0.9 then the loyalty value of month 2 equals the loyalty value of 1 plus the string value of month 2. If the loyalty value of month 2 is below the churn threshold then the full customer string is complete and registered. If the loyalty value of month 2 is above the churn threshold then the process repeats with month 3. The cycle continues through to the final month where the loyalty string is registered. If all months have been detected as being above 0.9 then the loyalty string would be written as NULL, the equivalent of no string. This is because a loyalty profile does not exist. It would be interesting to monitor the behaviour of customers with loyalty string values NULL. Customers now have loyalty values that can be compared to one another. This concludes the fourth software release so work can begin on release five.

5.2.5. Release 5 – Master Profile Storage

A method had been created in release 4 for removing redundant information from the beginning of the string and terminating the monitoring of activity after the customer profile has fallen below the churn threshold. The next development stage is to determine a way of storing master profile clusters as they are identified.

Tasks:

Determine the best solution for storing master profiles

Determine the best solution for storing customer results

Determine a method of implementing the solutions

Allow this stage to be cancelled it deemed unnecessary

There are several ways that this could be done. The first method would be to create a database that is continuously used by the software for the storage of master profile clusters. There is a major drawback of using a database for storing profiles. Each time a new dataset is applied to the software the results from the previous analysis will be lost. The alternative is to define a results table specifically for dataset being analysed. To do this it was decided to include an ‘Initiate Tables’ user interface (UI). This UI is shown in Figure 46:

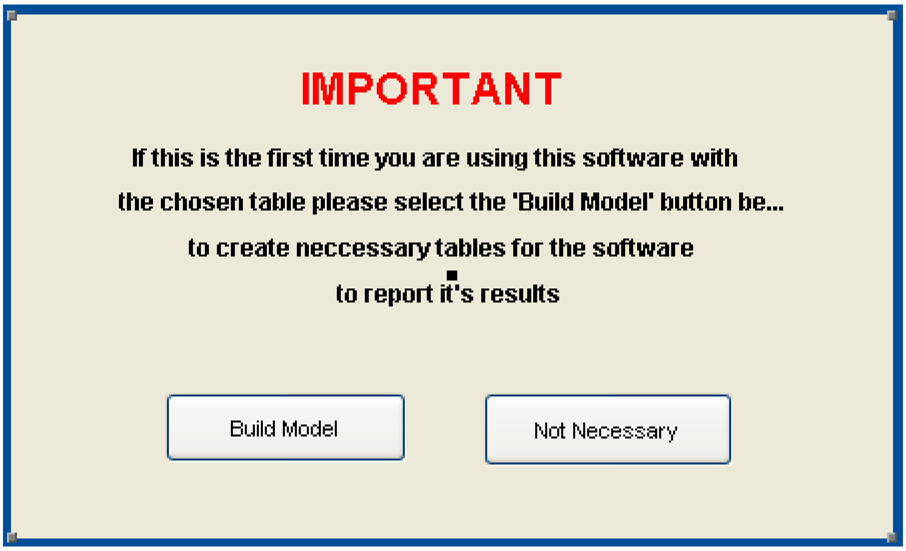


Figure 46: Initialise Tables GUI

As shown from Figure 46, the initialise tables GUI provides two options. The first option is to build a model. The UI advises that if this is the first time it is being used to analyse a particular dataset then the 'Build Model' button should be selected. If this is not the first time the software has been used with the particular dataset then the 'Not Necessary' button can be selected. The not necessary button simply closes the GUI and cancels this stage of the process. The 'Build Model' button does several things in the following sequence:

Establish a connection with the database

Execute a create table SQL statement for storing master profile clusters

Execute a create table SQL statement for storing customer results

It can be seen that the steps for creating necessary tables to hold master profile clusters and customer results is a straight forward process due to the ability of being able to define and execute SQL statements directly from the software. The flow chart in Figure 47 illustrates the process:

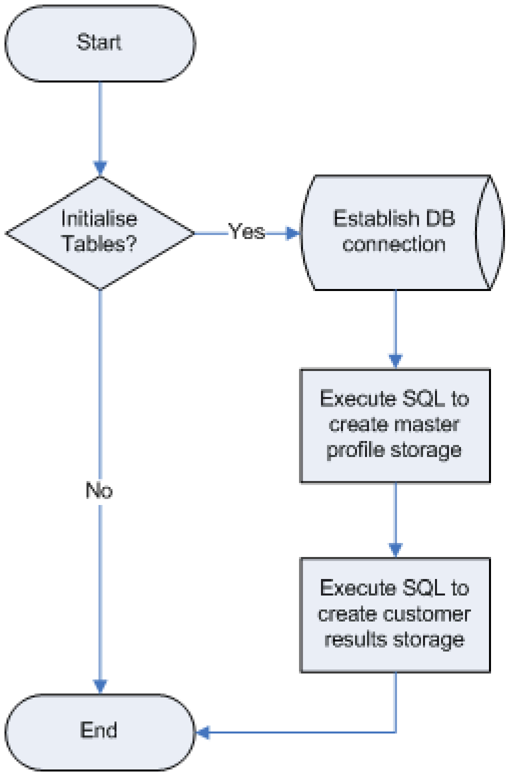


Figure 47: Initialise Tables Process

As shown by the flow chart in Figure 47 the process of creating additional tables for storing profile clusters and customer results is straight forward. With release 5 completed, work can begin on software release 6.

5.2.6. Release 6 – Master Profile Capture

Software release six deals with the problem of master profile capture. It has been determined from the development of the customer profiling methodology for churn prediction that master profile clusters should capture future churn. For the purpose of capturing churn it is logical to ensure that all profiles are created from churners. The following tasks have been identified to ensure that software release 6 is successfully.

Tasks:

Analyse the customer profile as created from software release 4

Compare the customer profile against master profile clusters to see if a match can be made.

- If the customer's profile falls below the churn threshold an exact match is required
    - Determine the time it takes the customer to churn after his/her loyalty index falls below the churn threshold
- If the customer's profile does not fall below the churn threshold a part match is required

If the profile does exist assign that profile to the customer

If the profile does not exist check to see if the customer is an actual churner

- If the customer is an actual churner add his/her profile as a new master profile cluster
- If the customer is not a churner it is identified that this profile does not lead to actual churn so do nothing

It can be noticed that for the software release two important stages are merged; the determination of customer profiles and the allocation of customer base to the customer profile groups. The reason that these stages have been merged is because they are closely related. The master profile clusters are created from customer profiles where the customer has churned and master profile cluster does not already exist. Therefore the first stage for software release six is to see if the customer profile can be matched to any of the master profile clusters that have already been identified and stored. This is done by establishing a connection with the database and using an SQL query to compare the customer's profile string with the master profile cluster strings. There are two SQL statements that can be executed, one for an exact profile match and one for a partial profile match. The statement executed for the customer being analysed depends on the condition of the customer profile. If the customer profile contains a loyalty index which has not fallen below the churn threshold then an SQL statement is executed that contains a wild card character (%). The percentage sign wild card character is used to match zero or more characters. When comparing the customer's string, if an exact match is not found then the percentage wild card will ensure the best match is found. There may be more than one master profile that contains the characteristics of the customer's profile string. For this reason the master profiles are retrieved in ascending order and the results limited to 1. This ensures that only 1 master profile is returned as a match and that master profile contains the shortest possible route that the customer could take to churn. It is important to match the customer to the shortest possible route because otherwise the churn event could be missed. If the customer does not continue to follow the shortest master profile characteristics then a new master profile assignment can be later matched. For example the customer profile in Figure 48 could match either of the master profiles in Figure 49, Figure 50 or Figure 51.

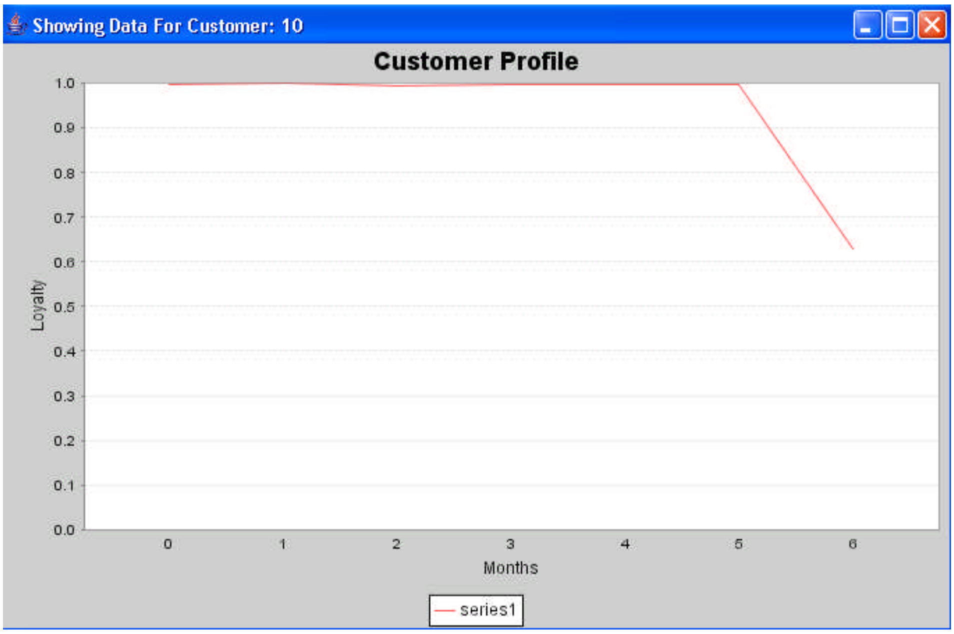


Figure 48: Customer Profile to be Part Matched

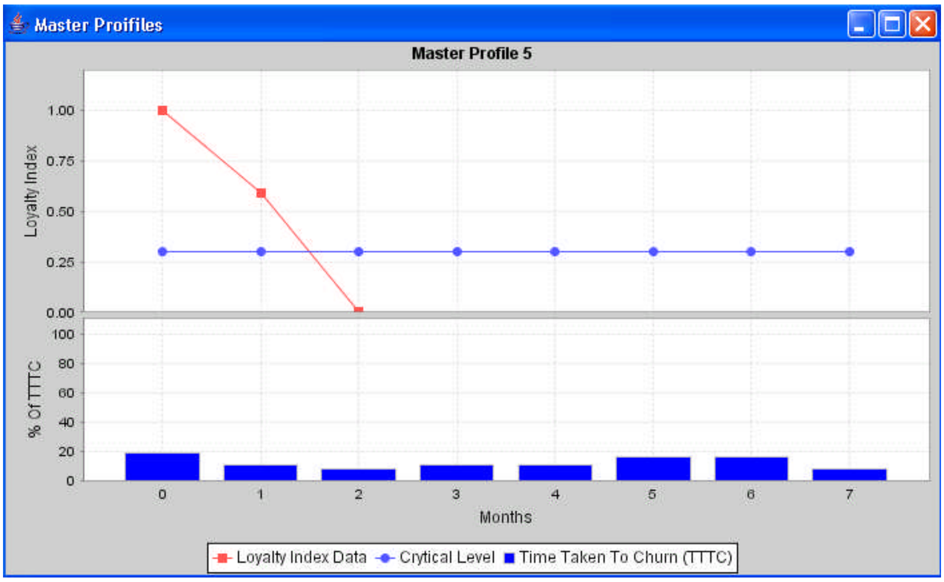


Figure 49: Possible Master Profile Match 1

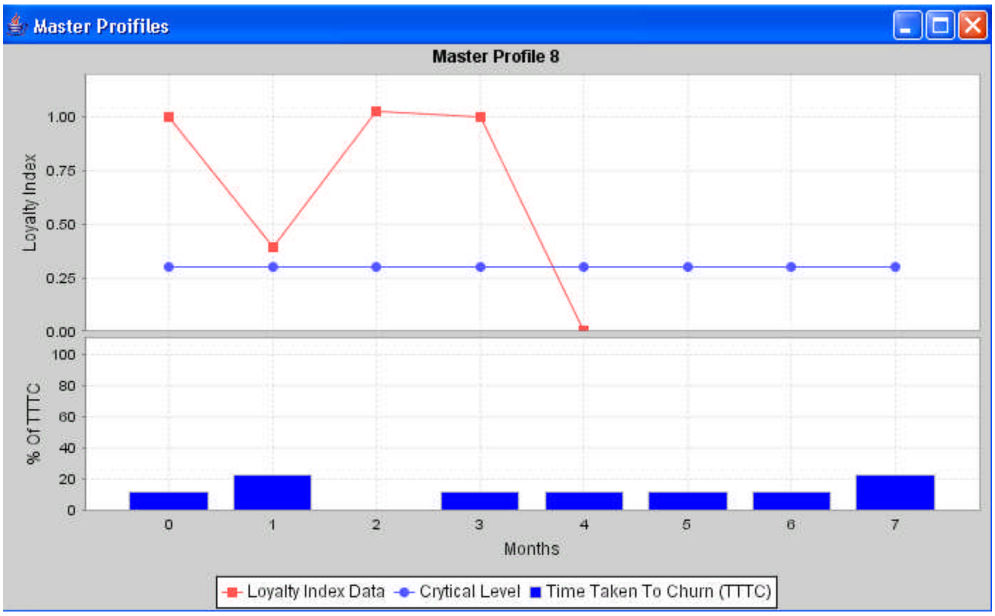


Figure 50: Possible Master Profile Match 2

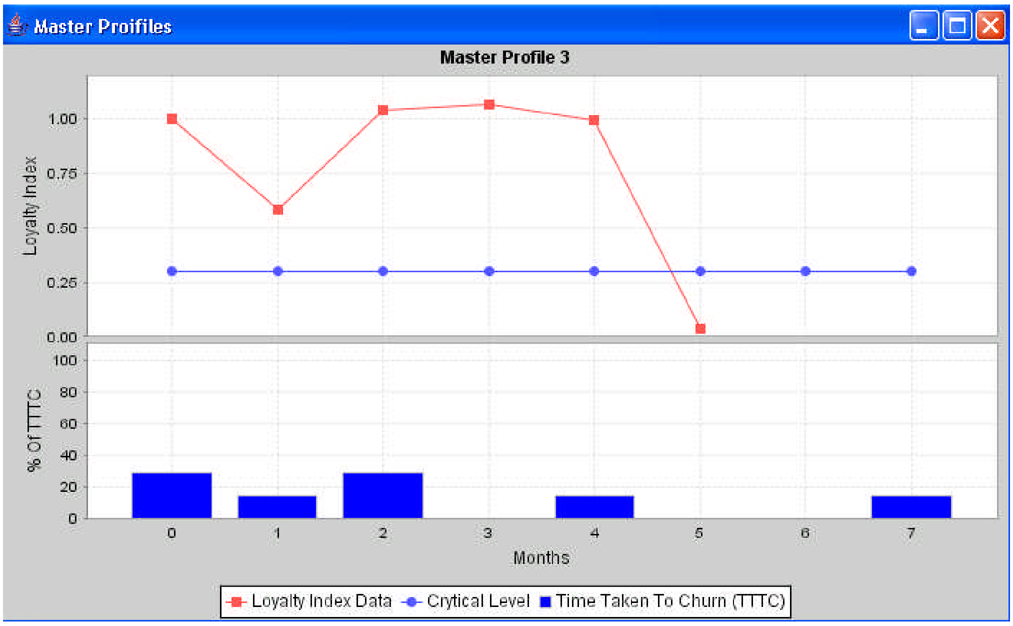


Figure 51: Possible Master Profile Match 2

The master profile in Figure 48 could match any of the profiles in Figure 49, Figure 50 or Figure 51, so which master profile cluster should the customer be assigned? It would be dangerous to assign the customer to the profile in Figure 51 because it will take four months to find out if that customer does actually belong to that cluster. If no other profile matches the customer then this would be fine; however the customer could still match the profiles shown in either Figure 49 or Figure 50. Likewise if the customer was matched to the profile in Figure 50 it would take a further 3 months to know if the customer belonged to that cluster. The master profile cluster shown in Figure 49 would result in the customer churning in 1 month time, therefore this is the worst scenario and should be regarded as the master profile cluster to which the customer should be matched. At this point it would be unnecessary and possibly costly to regard the customer as a churner; however the customer could be placed in a quarantine list and monitored more closely than the customers who have not been detected as potential churners. If the customer's loyalty does fall below the churn threshold in the next month then the customer should be contacted within the shortest possible time. The time it takes the customer to churn is calculated by presenting the actual churn column of the customer base to the software in the form of the number that corresponds to the position in the time series where churn actually occurs. For example if the dataset contains 20 months of data all churn that occurs in the first month is represented as the number '1', all churn that occurs in the second month is represented as the number '2', in the third month number '3', fourth month, number '4', etc. Then if the customers loyalty index falls below the churn threshold in month 2, and on inspection of the actual churn data it is discovered that the customer does actually churn in month 7, the calculation 7 – 2 can be used to determine that the customer actually churns 5 months after his/her loyalty index value falls below the churn threshold. This data is recorded as a count with the master profile so that a probability of retention can be built displaying the average time it takes customers belonging to that cluster to churn after his/her loyalty index falls below the churn threshold.

The second SQL statement is executed if the customer's loyalty index does fall below the churn threshold. In this case the profile has to match exactly to a master profile cluster. If no exact match can be found then the actual churn data has to be checked to see if the customer has actually churned. If the customer has actually churned then it has been determined that this type of customer profile does lead to customer churn and the profile can be added as a master profile cluster itself. If the customer has not actually churned then it has to be assumed that the particular profile does not lead to churn and no action is taken. These processes are displayed by the flow chart in Figure 52:

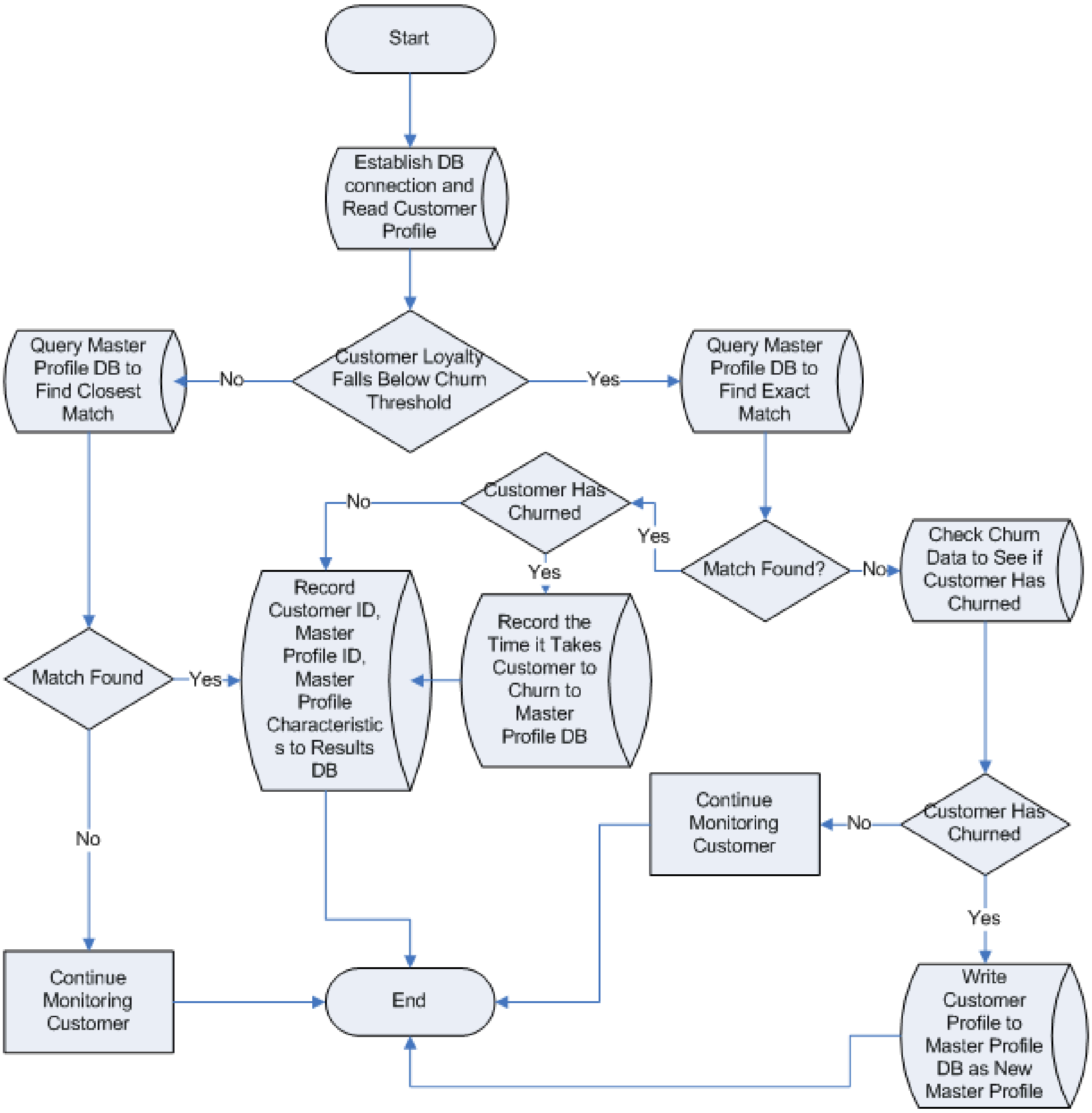


Figure 52: Master Profile Capture\Assignment

Figure 52 concludes the sixth software release. The next stage is creating a visualisation of the master profile clusters.

5.2.7. Release 7 - Master Profile Visualisation

The seventh release is to provide a graphical representation of the master profile clusters. This has been allocated a short development time as the creation charts has already been investigated for release four – customer profile visualisation; however the master profile clusters use a more advanced chart type that was required for displaying the customer profiles.

Tasks:

- Determine a method for displaying two charts, one for the master profile and one for the probability of retention.
- Include a calculation of how long the customer takes to churn after his/her loyalty index value has fallen below the churn threshold (to create a probability of retention)
- Determine a method of displaying the probability of retention with the master profile.
- Create a GUI for selecting a master profile cluster to view

The first bullet point above requires a method to be determined for displaying two graphs in one window. This was achieved by way of trial and error (exploring the available JFreeChart classes). It was identified that JFreeChart contains a hybrid class for combining two chart types. A line chart was selected to display the master profile characteristics and a bar chart type was selected for displaying the probability of churn. An example of a master profile chart is displayed in Figure 53:

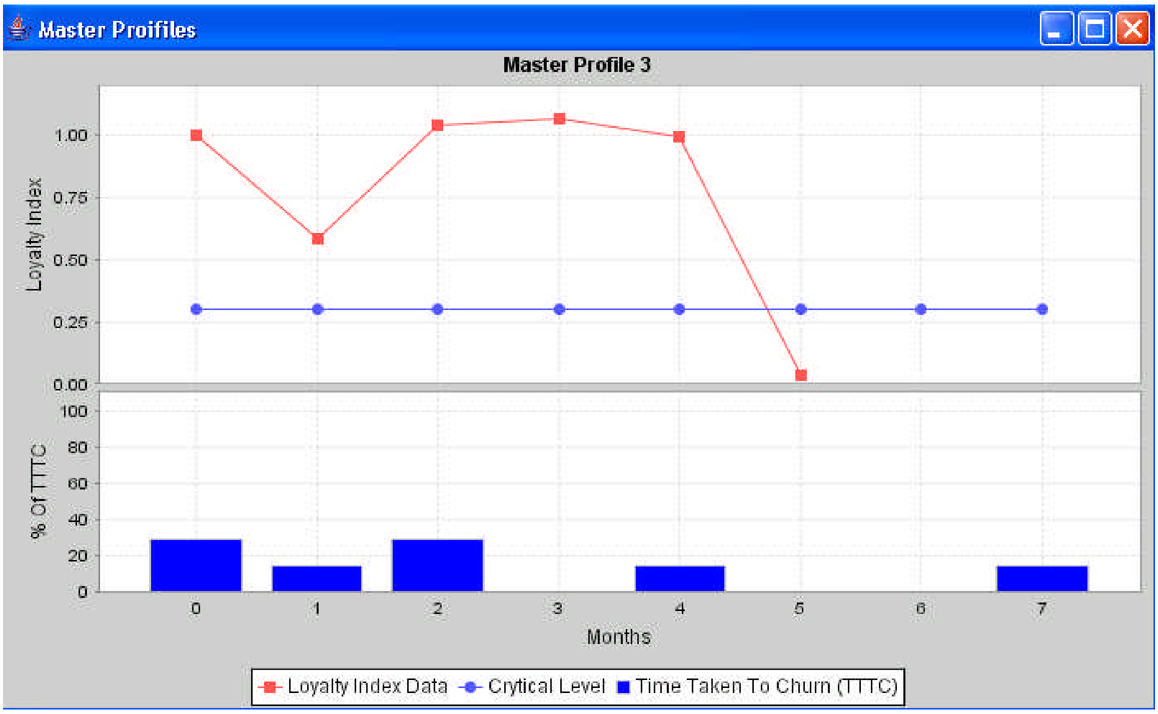


Figure 53: Master Profile Chart Example

As can be seen from Figure 53, the master profile chart contains 2 graphs. The line graph on top displays the characteristics of the master profile cluster as captured from the customer data. The bar graph at the bottom displays the probability of retention. The probability of retention is shown as a percentage rather than a count of customers. This is because a percentage provides cleaner information than the count. For the case of master profile cluster 3 as shown in Figure 53 it can be seen that 25% of churners who have been matched to master profile cluster 3, churned immediately once there customer loyalty index fell below the churn threshold. About 15% of customers churned 1 month after their loyalty index value fell below the churn threshold, about 25% churned 2 months later and so on. It is anticipated that the best way to use this data is to recognise that 25% of customers belonging to this cluster churn immediately, which means there is a 75% chance that customers being assigned to this profile cluster can be retained if contacted by the churn management department within 1 month of the customers loyalty value falling below the churn threshold.

To be able to view the master profile there has to be a method of requesting which master profile to show. This is done by entering the master profile ID into a GUI. The GUI is shown in Figure 54:

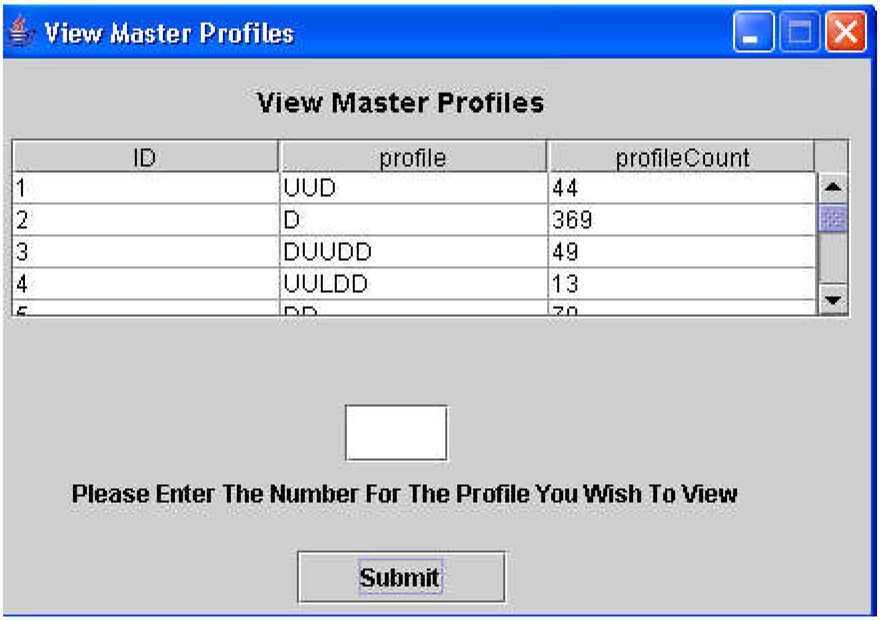


Figure 54: Request Master Profile GUI

As can be seen from Figure 54, the view master profile GUI contains a table of the master profiles. This was added to make it easier to select a master profile to view. The end user can scroll through the table looking at the profile characteristics. When he/she finds one he/she wishes to look at he/she can enter the profile ID number into the text box and hit the submit button. This will then display the requested master profile. The flow chart shown in Figure 55 provides the deliverable for the seventh release.

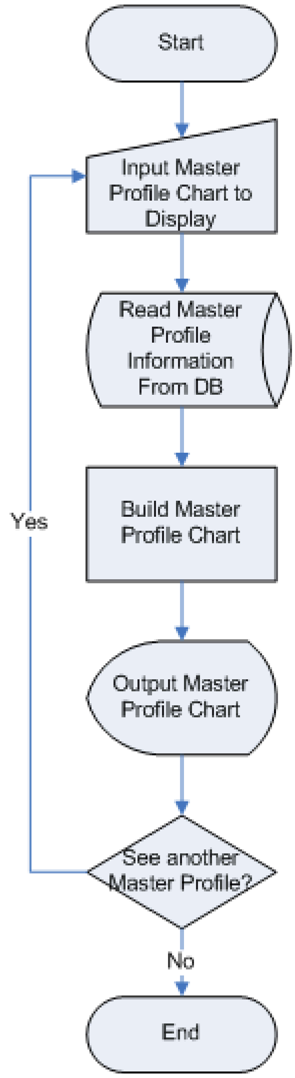


Figure 55: Display Master Profile

With the seventh release completed the software is complete as it provides all necessary functionality. There is an eighth deliverable – customer churn prediction. This deliverable will not be coded in the software. Instead the results generated from the software will be exported to Microsoft Excel for further analysis. Identification of high\low risk profiles and using these profiles for future prediction were not included in the software. The reason it was not included was because these processes can be achieved quickly by an analysis of the results using Microsoft's Excel. Once high risk profiles are identified all future analysis for the dataset has predictions set to 0 unless the prediction corresponds to a high risk profile. The results are then imported into Matlab for automatic generation of a confusion matrix for analysis of the results. This is because the software provides all necessary functionality for automating the creation and assignment of master profile clusters.

5.2.8. Churn Prediction

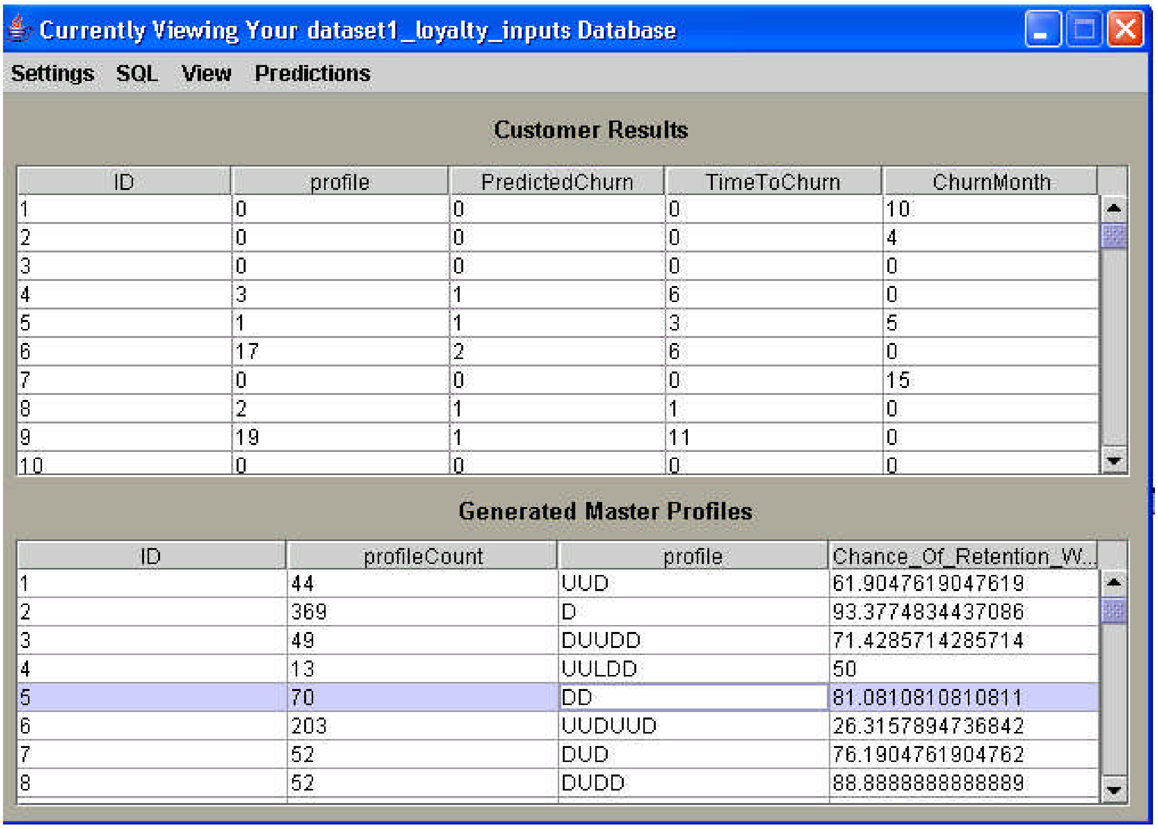
Customer churn prediction has not been developed into the software. The decision was made not to provide this feature because it can be done quickly using Microsoft Excel, while it was anticipated that hard coding it as a feature would take at least 3 months. The processes of performing churn prediction using the results generated from the prototype are as follows:

- Copy the results database into Excel
- Copy the actual churn data for the dataset into Excel
- Look up the churn flag for each customer (1 = churn, 0 = non-churn)
- Apply an auto filter to the spreadsheet
- Filter the spreadsheet by profile number (i.e. 2) and actual churn (Churn = 1)
- Use the count feature in Excel to determine how many customers in that profile group actually churned
- Record the actual churn numbers for the profile cluster
- Change the profile number to record the actual churn of each cluster

The processes defined in the above bullet points provide a determination of how many customers in each profile cluster have actually churned. The software provides the numbers of how many customers have been assigned to each profile cluster. It is then possible to compare the actual numbers against the predicted numbers to assess the profile clusters that have been most accurate at capturing future churn.

5.2.9. Merging All Software Releases

All software releases can be compiled into a single application with a main GUI from which all other functionality can be controlled. The main GUI is shown in Figure 56:



ID	profile	PredictedChurn	TimeToChurn	ChurnMonth
1	0	0	0	10
2	0	0	0	4
3	0	0	0	0
4	3	1	6	0
5	1	1	3	5
6	17	2	6	0
7	0	0	0	15
8	2	1	1	0
9	19	1	11	0
10	0	0	0	0

ID	profileCount	profile	Chance_Of_Retention_W...
1	44	UUD	61.9047619047619
2	369	D	93.3774834437086
3	49	DUUDD	71.4285714285714
4	13	UULDD	50
5	70	DD	81.0810810810811
6	203	UUDUUD	26.3157894736842
7	52	DUD	76.1904761904762
8	52	DUDD	88.8888888888889

Figure 56: Main Software GUI

It can be seen from Figure 56 that the main GUI contains two charts. The top chart displays the customer results from the software while the bottom chart displays the master profile clusters. Across the top of the GUI is a menu bar from which all other functionality can be accessed. The menu bar contains the following options:

Settings

- o Properties – Call properties GUI (Shown in **Error! Reference source not found.**)

SQL

- o Reset Database – Execute SQL master profile and customer results tables
- o Reset Counters – Clear all counters, (i.e. master profile cluster size etc) for new analysis

View

- o View Master Profile – Call view master profile GUI (Shown in Figure 54)

- View Customer Loyalty Profile – Call view customer GUI (Shown in Figure 42)
- Refresh Tables – Refresh the information shown by the tables in the main GUI (Shown in Figure 56)

Predictions

- Predict Entire Dataset – Call predict entire dataset GUI (Shown in Figure 40)
- Predict Single Customer – Call predict single customer GUI (Shown in Figure 39)

5.3. Summary

This chapter has documented the development of a prototype system for churn prediction. The software development used several advanced programming techniques and proved a challenge to complete; however it does validate the ability of being able to transfer the determined methodology into a software prototype for automated churn prediction through large dataset analysis.

The XP methodology was selected for two reasons. The first reason is that it is one of the fastest forms of development methodologies, and the second reason is that most stages of the XP methodology can be applied by a single developer. Creating software as part of a PhD is in most cases a one man job. This is a problem for applying any software development methodology as all software methodologies are targeted towards group development and communication between developers and customers. The other alternative would be to apply the traditional ad-hoc method of programming on the fly. This would most likely lead to an unstable software tool built with unreadable code. Applying a software methodology as closely as possible has helped to develop a quality prototype.

This Page is Intentionally
Left Blank

6. Validation of Customer Profiling Methodology for Churn Prediction

This chapter analyses the performance of the customer profiling methodology for churn prediction using historical data for three separate case studies, from three separate service sectors within the field of telecommunications. The following objectives are the target of this chapter:

- To validate the performance of the customer profiling methodology using three separate real life case studies.
- To report the experimental results as generated using the customer profiling methodology for each of the cases.
- To compare these results with those reported in literature for other methodologies.
- To draw conclusions from these results regarding the performance of the customer profiling methodology.

6.1. Case Study Development

The following section details three historical case studies in the form of customer data. Each case possesses unique characteristics of the specific service sectors from which they were supplied. All cases fall under the generic umbrella of the telecommunications industry. The data for all cases is a direct representation of the actual accessible data from the respective service sectors. Each case will be used to analyse the data over a sequence of time in an attempt to predict the churn of customers at a future point past the latest date used for profile creation. The data documented for each case study has been desensitised in order to protect the sponsoring company's business strategy. The unique characteristics of each of the three cases can be seen from Table 9:

Table 9: Case Study Characteristics

Case Studies	**Case-1 (Residential Mobile Phone Data)**	**Case-2 (Residential Broadband Data)**	**Case-3 (Business Landline Telephone Data**
Customer Repairs and Complaints Data	35 variables 8408 Customers 13 Months of data 29:71 Churn\Non-Churn 13 Months Contain Churn	25 Variables 18453 Customers 10 Months of Data 6:94 Churn\Non-Churn 1 Month Contains Churn	4 Variables 7358 Customers 9 Months of Data 6:94Churn\Non-Churn 9 Months Contain Churn

It can be identified from Table 9 that the data for all cases is specific to customer repairs and complaints. The reason for this is because due to monopoly regulations, large service leaders are restricted by the data that they can use for customer analysis. Customer repairs and complaints data can be freely used and analysed and it is anticipated that this type of data can be a major contributor to customer churn.

6.2. Neural Network Development

To ensure that the most accurate NN (neural network) architecture is constructed for the generation of a customer churn index several experiments will be performed for each case study using various NN configurations.

6.2.1. NN Architecture 1

The first NN is a simple single layer; single neuron NN as keeping complexity of the architecture to a minimum could significantly decrease the analysis time. The architecture for this type of NN can be viewed in Figure 57:

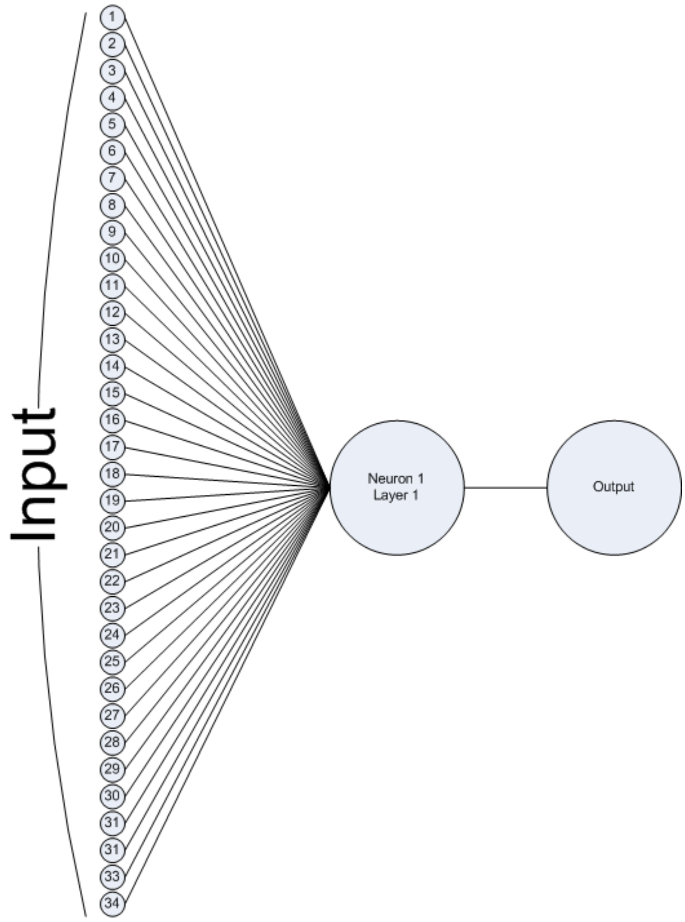


Figure 57: Single Layer, Single Neuron Network

The diagram in Figure 57 illustrates the general neural network architecture. The diagram in Figure 58 displays further details about the neural network configuration: -

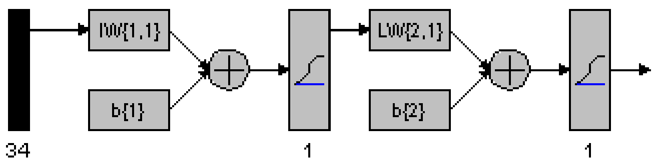


Figure 58: Single Layer, Single Neuron Layer Configuration.

Figure 58 illustrates how data is passed through the created neural network architecture. From left to right, the diagram displays the input layer. As can be noted from the diagram each input has its own weight associated to it (IW[1,1] = input weight), with a general bias (b{1} = input bias) applied. The number underneath the first box indicates the total number of inputs. All inputs are then summed. Next the Logsig activation function is applied (this stage is symbolised by the sigmoid sign). The number underneath the box containing the sigmoid sign represents the total number of neurons contained within the hidden layers in that particular neural network architecture. The resulting value again has a weight (LW{2,1} = layer weight) and bias (b{2} = Layer Bias) applied and the churn rate is finally output via the output layer. The number underneath the final box indicates the total number of outputs.

6.2.2. NN Architecture 2

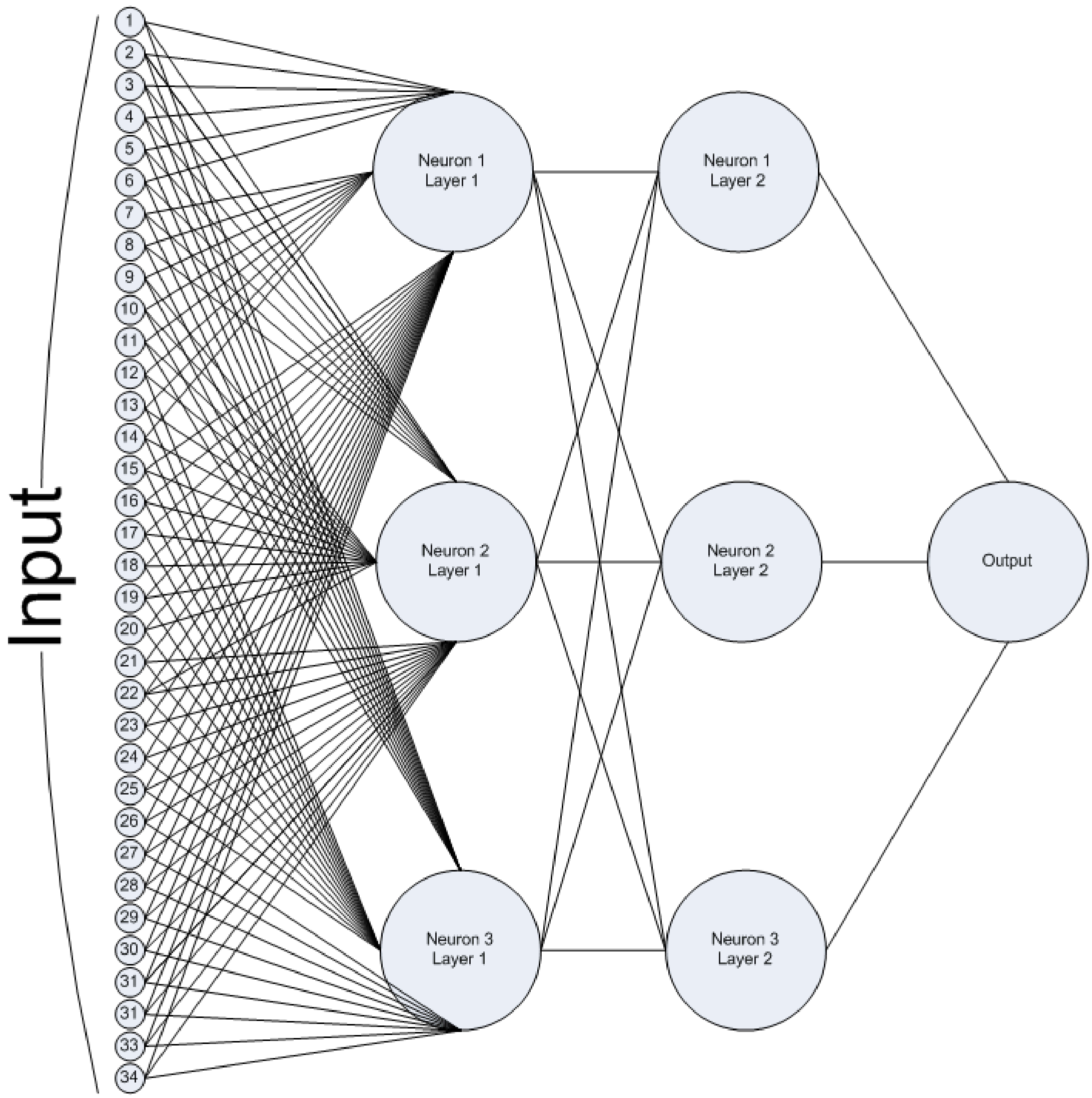


Figure 59: Two Layer, Three Neuron Network.

Figure 59 illustrates a more complex neural network architecture to that shown in Figure 57. The corresponding detailed diagram for the second NN architecture can be seen in Figure 60:

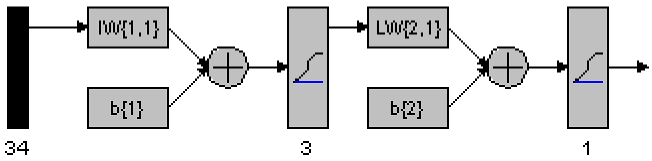


Figure 60: Two Layer, Three Neuron Layer Configuration.

It can be seen by the neuron configuration displayed in Figure 60 that the basic neuron is the same as the one shown in Figure 58; however each hidden layer now contains three neurons instead of just one.

6.2.3. NN Architecture 3

The final experiment analyses the performance of a large neural network. A neural network is constructed with four hidden layers and seven neurons per layer. The neural network architecture can be viewed in Figure 61:

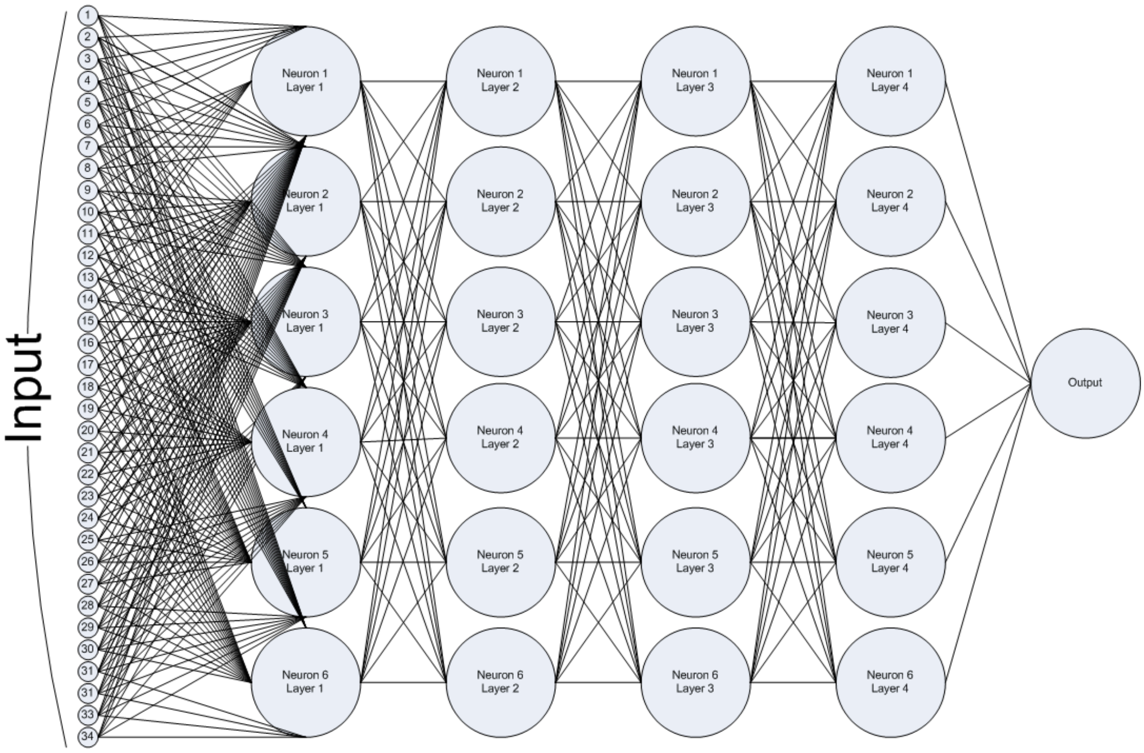


Figure 61: Four Layer, Six Neuron Network.

As can be seen from Figure 61, the final neural network is much larger than the previous two. This experiment has been performed to establish if the complex problem of determining a customer churn rate requires a large neural network. The detailed level diagram for the neural network pictured in Figure 62 can be seen in Figure 62:

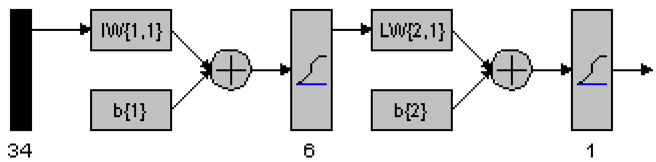


Figure 62: Four Layer, Six Neuron Layer Configuration.

6.2.4. Developing the NN Using Matlab

This section attempts to discover the optimum learning and configuration properties for a feed-forward, back propagation neural network using Bayesian regularisation (TrainBR) in Matlab.

As numerous NN architectures have been developed, the ways in which these architectures perform have also seen developments. Matlab has constructed a neural network toolbox for easy creation and experimentation of neural networks. To provide an example of the number of varying neural networks, the NN toolbox offered by Matlab provides support for 15 different neural network architectures. As an example, within Matlab the feed-forward back-propagation algorithm can be configured to use one of 18 different training functions, 2 different learning functions, 3 different performance functions and 3 different transfer functions.

6.2.5. Neural Network Training Algorithms

A full investigation into all training functions available from Matlab is beyond the scope of this thesis; however an investigation regarding the 'TrainBR' algorithm will be provided.

The desirable outcome for a neural network is to obtain a small error for both training and testing data. A Bayesian regularisation limits the neural network to converge to a set of small weight and bias values, creating a resulting network that provides a smooth response with a decreased chance of over-fitting to the initial training patterns (Saini, 2008). Bayesian regularisation basically adjusts the NN weights and biases to such a degree that the NN outputs to the training targets in an optimum way (Caballero et al., 2006). Because of this Bayesian neural networks are known to generalise well (Demuth et al., 2007). Bayesian reglarisation will be used for the basis of NN to generate churn index values due to their ability to generalise well.

6.2.6. NN Learning Functions

There are two learning functions available for use with Bayesian regularisation in Matlab. These learning functions are name LearnGD and LearnGDM. Details of both traning algorithms are as follows:

6.2.6.1. LearnGD

The LearnGD Matlab function represents 'gradient descent weight\bias learning function'. Learning occurs in this algorithm via two specific learning parameters - the 'learning rate' and the 'momentum constant'. These are the only two parameters needed for the LearnGD algorithm to be able to calculate a rate change. Default values for these properties are set to 0.01 for the learning rate and 0.9 for the momentum constant. These properties can be altered using the command line of Matlab using the format shown in Equation 3:

$$lp.lr \quad 0.01$$

Equation 3: Manipulating LearnGD Training Rate.

To further explain Equation 3, *lp* is the call to the NN learning parameters class while *lr* provides access to the learning rate method of the learning parameters class.

6.2.6.2. LearnGDM

LearnGDM is similar to LearnGD as discussed; however an additional factor is used for the calculation of the weight change. An example is provided in Equation 4:

$$\begin{aligned} &lp.lr \quad x \\ &lp.mc \quad x \end{aligned}$$

Equation 4: LearnGDM Learning Parameters.

As explained for Equation 3, *lp* is a call to the learning parameters class while *lr* accesses the learning rate method of the learning parameters class. *mc* provides access to the momentum constant method of the learning parameter class. The momentum constant defines amount of momentum. Momentum allows a network to ignore small features in the error surface. Without momentum a network can get stuck in a shallow local minimum. With momentum a network can slide through such a minimum (Demuth et al., 2007). The x in Equation 4 can be set to a user defined value between 0 and 1. It is suggested that faster convergence and better generalisation can be achieved by modifying the learning algorithm (Han et al., 2008), however this practice is more common with technologies such as genetic algorithms (GA) and fuzzy logic (FL) (Wang et al., 2007).

The learning function used in the construction of the churn methodology will be learnGDM as this is the default learning algorithm for TrainBR using Matlab's NN toolbox.

6.2.7. Performance Functions

As mentioned in section 6.2.7, there are three different performance functions available within Matlab for use with Bayesian regularisation. It should be noted that for the case of Bayesian regularisation early stopping should be performed when the sum of squared error and sum of squared weights values that are shown during training become static (Demuth et al., 2007).

6.2.7.1. MSE (Means Squared Error) Performance Function

The MSE performance function measures the network's performance according to its mean of squared errors. The simplest formula to represent the MSE can be seen from Equation 5:

$$MSE \quad \frac{(Actual \quad Forecast)^2}{n}$$

Equation 5: MSE Equation

Where n = the number of total data points.

6.2.7.2. MSEREG (Means Squared Error Regularisation) Performance Function

MSEREG is a modified performance function. This type of performance function makes it possible to increase generalisation through an additional term which consists of the mean of the sum of squares of the network weights and biases. The following mathematical explanation is shown in Equation 6 as provided by Demuth (2007):

$$MSEREG \quad ymse \quad (1 \quad y)msw$$

Equation 6: MSEREG Formula

Where y is the performance ratio and msw is calculated by the formula shown in Equation 7:

$$msw \quad \frac{1}{n} \quad \overset{n}{\underset{i \ 1}{}} \ w_j^2$$

Equation 7: MSW formula

6.2.7.3. SSE

SSE (sum of squared error) is the most popular method for determining the performance of Bayesian regularisation. The SSE is the sum of the squares of the residue between actual and predicted values. The SSE formula is shown in Equation 8:

$$SSE \qquad (Actual \quad forecast)^2$$

Equation 8: SSE Formula

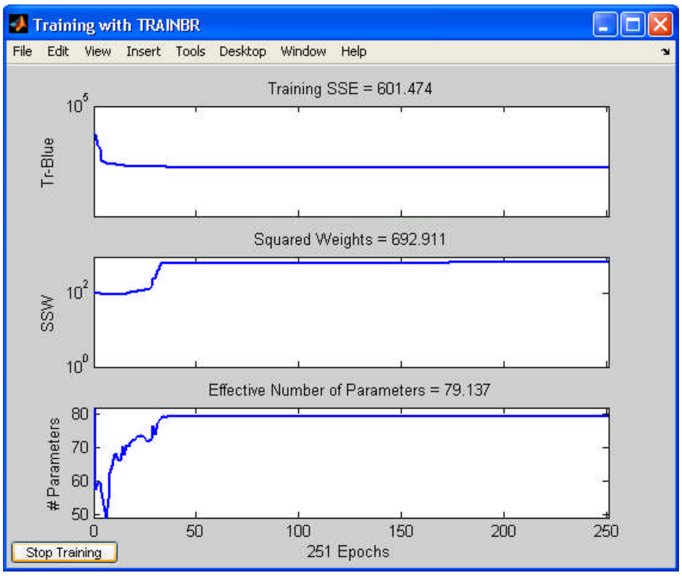


Figure 63: Matlab Training Performance Monitor.

The performance monitor shown in Figure 63 displays the SSE. Matlab allows the error for each data point to be exported. Using this feature the errors can be easily imported into MS Excel for further analysis.

6.3. Further Analysis of Matlab's NN Bayesian Regularisation Performance

With the errors generated from Matlab imported into Excel the first thing that can be checked is the sum of squared error. The sum of squared error is the value given at the top of the window shown in Figure 63.

For the analysis in this section a test neural network was created and trained. The performance of the trained network can be seen from the performance monitor in Figure 64:

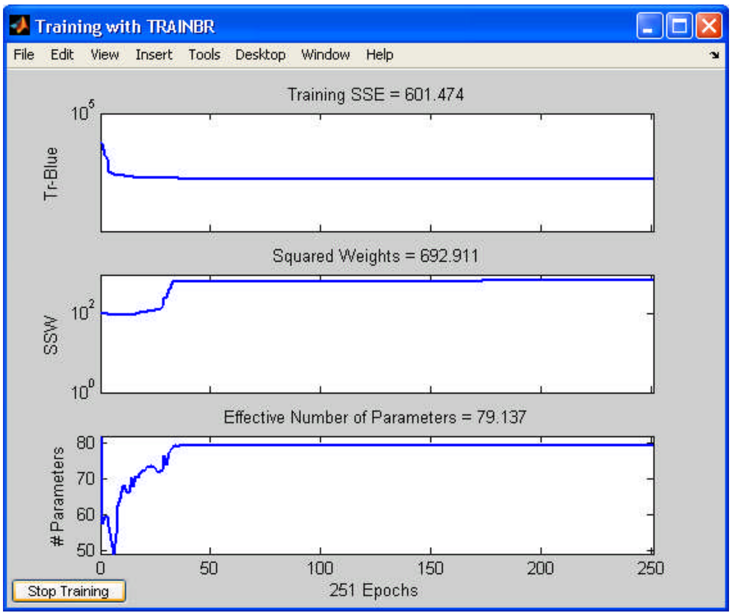


Figure 64: Data Analysis SSE

With the SSE calculation performed in Excel the returned value = 601.4741.

1	Customer ID	Error Value	> -0.05	<0.05
2	2	-0.025995461	1	1
3	3	-0.025995461	1	1
4	4	-0.025995461	1	1
5	5	-0.025995461	1	1
6	6	-0.025995461	1	1
7	7	-0.025995461	1	1
8	8	-0.040896165	1	1
9	9	0.037301919	1	1
10	10	-0.042403189	1	1
11	11	-0.025995461	1	1
12		-0.025995461	1	1

Figure 65: Additional Excel Analysis

To further analyse the error rates, all results have been filtered according to their error values to help determine if there is a structured pattern to them. To accomplish this, error values are exported out of Matlab and imported into Microsoft's. An example of the exported results is given in Figure 65. Once the values have been imported into Excel two further Columns ('>-0.05' and '<0.05') are manually added. These columns hold information regarding each error so that a simple filter can be effectively applied. The value '1' is returned if the error value is greater than abs(0.5). A filter is then applied to find all customers with loyalty values that are outside acceptable ranges of between -0.05 and +0.05. The value 0.05 was used because as can be seen from, most values fall between the range of 0.0 and 0.04. Only a small minority of customers have been graded with error values greater than 0.04 or lower than -0.04.

All of the error values that fall outside the stated conditions, so all the unusually large error values, are exported to a separate worksheet. The total number of customers in this list is 1703. The total number of customers in the entire dataset is 18453. Therefore around 9% of the dataset has a poor error value. With the SSE calculated for only these customers it is found that they account for 587.6714 of the total SSE of the original 601.4741. With these customers removed from the dataset. The SSE for the remaining 16750 customers equals a total of 13.80273.

The next step was to try to determine why the 1703 customers generated such a large error values. The ID's for these customers are queried against the full dataset to provide the data for just these customers. On verifying if these customers are churners or non-churners it is discovered that this group of customers contains almost a 50:50 split between churn and non-churn. To be precise, from the 1703 customers 866 are churners and 837 are non-churners. The entire dataset of 18453 only contains 1066 churners, therefore the 1703 customers that have been identified to hold the larges error values contain 81% of the total churners for the entire dataset.

The neural network has been successful in establishing a pattern for customers who do not churn which is evident from the very low SSE for this group. The non-churners also form the large majority in the dataset so their identification is easiest. Churning customers do not fit the conventional pattern displayed by non-churners resulting in an increased error value, and it is this increased error value that is responsible for the majority of the overall SSE of 601.474.

6.3.1. Transfer Functions

There are numerous transfer functions available for configuration through Matlab's NN toolbox. The activations functions available to the TrainBR training method are LOGSIG, PURELIN and TANSIG.

6.3.1.1. LOGSIG Activation Function

The LOGSIG (Log-Sigmoid) activation function takes any value between plus and minus infinity and squashes the output into range between 0 and 1. A sigmoid activation function is best known from the symbol shown in Figure 66

:

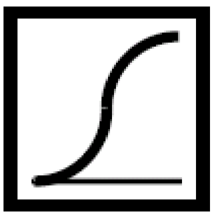


Figure 66: Sigmoid Symbol

6.3.1.2. PURELIN (Linear) Activation Function

The PURELIN activation function is a linear transfer function. Neurons that use this type of activation function are usually used as linear approximators. The PURELIN activation function produces the same output as its input. A PURELIN activation function can be recognised from the symbol in Figure 67:

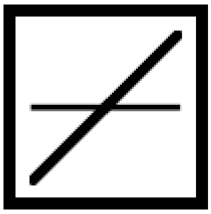


Figure 67: PURELIN Symbol

6.3.1.3. TANSIG (Tan-Sigmoid) Activation Function

The TANSIG activation function is an alternative sigmoid activation function that can be used instead of LOGSIG. The main difference between TANSIG and LOGSIG is TANSIG squashes the output value between -1 and 1 instead of 0 and 1. TANSIG was named after the hyperbolic tangent, which has the same shape. The symbol that represents a TANSIG function is shown in Figure 68:

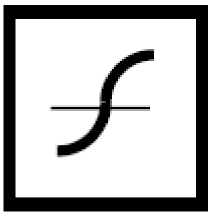


Figure 68: TANSIG Symbol

6.3.1.4. Activation Function Experiments

Experiments were performed using all three activation functions to analyse which would display the most accurate results for the prediction of customer churn. Based on the initial experiments performed during the creation of the methodology a neural network was created using a sample dataset containing 202 customers. 108 customers were churners and 94 customers were non-churners. Three NN's were created, one configured with the LOGSIG activation function, one configured with the PURELIN activation function and the final NN configured to use the TANSIG activation function. The NN's were then set to predict the churn for that dataset setting a churn threshold of 0.3. The results displayed in Figure 69 were obtained for each activation function:

True Labels	Estimated Labels 0	1	Totals
0	54	40	94
1	10	98	108
Totals	64	138	202

Figure 69: LOGSIG Activation Function Results

Figure 69 displays the results obtained for the LOGSIG activation function experiments. The LOGSIG activation function experiment has captured 90.7% churn accuracy and 57.4% non-churn accuracy with a total accuracy of 75.2%.

True Labels	Estimated Labels 0	1	Totals
0	51	43	94
1	10	98	108
Totals	61	141	202

Figure 70: PURELIN Activation Function Results

Figure 70 displays the results for the PURELIN activation function experiment. The PURELIN activation function experiment has captured 90.7% churn accuracy and 54.2% non-churn accuracy with a total accuracy of 73.7%.

```
       | Estimated Labels
 True  |
Labels |    0       1   | Totals
-------|----------------|-------
0      |   53      41   |   94
1      |   63      45   |  108
-------|----------------|-------
Totals |  116      86   |  202
```

Figure 71: TANSIG Activation Function Results

Figure 77 displays the results for the TANSIG activation function experiment. The TANSIG activation function experiment has captured 41.6% churn accuracy and 56.3% non-churn accuracy with a total accuracy of 48.5%. For easy comparison the results for all three activation function experiments have been compiled into the table shown in Table 10:

Table 10: Comparison of Activation Functions Result

Activation Function	Churn Capture Accuracy	Non-Churn Capture Accuracy	Total Accuracy
LOGSIG	90.70%	57.40%	75.20%
PURELIN	90.70%	54.20%	73.70%
TANSIG	41.60%	56.30%	48.50%

It can be identified from Table 10 that the LOGSIG activation has provided the best results with a total prediction accuracy of 75.2%. PURELIN displayed the second best results with an overall accuracy of 73.7%. The TANSIG activation was very poor in comparison to the other two activation functions only accurately capturing a total accuracy of 48.5%. Based on these experiments it has been determined that the LOGSIG activation function should be used for NN generation for the customer profiling methodology.

6.4. Validation Stages

Validation of the customer profiling methodology will be carried out on the three separate case studies as described in Table 9. The following steps will be followed for each case study:

Case study description – Describes the data that forms the case study.

Data preparation – Provides a brief description of the variables that the case stud incorporates.

Identification of the most suitable for presenting to the methodology – explains the most suitable from the case study for building the profiling methodology.

1 layer 1 neuron NN experiment – Details the NN1 experiments.

2 layer 3 neuron NN experiment – Details the NN2 experiments.

4 layer 6 neuron NN experiment – Details the NN3 experiments.

Application of customer profiling methodology – investigates the application of the customer profiling methodology and the results obtained.

Identification of the high and low risk master profiles – determines the profiles that are highly sensitive for future churn capture and those profiles less sensitive.

Determining the future churn capture – Documents how many future churners are caught by the customer profiling methodology.

Comparison of the results with two alternative predictive methodologies – Compares the customer profiling methodology with two other found in literature.

- Hu (2005) comparison – The first methodology from which a comparison is based.
- Hwang et al. (2004) methodology – The second methodology from which a comparison is based
- Customer profiling methodology – The results of the profile methodology in comparison to the other two.

Discussion of comparative results – Provides a discussion on the comparison of the results obtained from the customer profiling methodology, against the results achieved by the two other methodologies identified from the literature.

6.5. Case 1

As can be seen from Table 9, case 1 contains 35 variables focused on customer repairs and complaints. These variables include information regarding how often a customer has contacted the repairs or complaints departments, how long the last event took to resolve, promises to the customer regarding service issues, and typical types of repairs and complaints. All data from this dataset corresponds to residential mobile telephone customers. A full list of the variables contained within the case 1 dataset cannot be reported due to confidentiality reasons; however a generalised list of some of the variables with a brief description can be seen in Table 11:

Table 11: Variables Used for Case 1 Dataset

Type of Variable	Description
Repair Referred Count	Total number of times enquiry has been referred to alternative departments
Contact from customer	Communications to the customer
Communication code	Code regarding the reason for communication
Enquiry Count	Total number of times enquiries have been made
Billing Enquiry count	How many time a customer has contacted with billing issue
Billing Enquiry Code	Code regarding the type of billing enquiry
Enquiry Exception	Reason for non classified enquiry
Provision Count	Number of provisions made to the customer
Enquiry Referred	Reason code for referring an enquiry to an alternative department
Repeat enquiries	Number of repeat enquiries
Fault Count	Number of Faults
Fault enquiry	Reporting a service error
Fault Resolutions promise	Promise to resolve a fault by given date
Fault time	How long the fault has taken to resolve

6.5.1. Preparing the Data for Case 1

The data for case 1 is inspected for incompatible values that could cause the creation and training of a neural network model to fail. All the data contained within case 1 is of appropriate numerical values, however it was found that some values were missing. Missing values where replaced with '0' using the 'Is Null' condition in MS Access.

All data supplied by the sponsoring company is provided as a large text file, so one huge table is generated when this file is imported into a database application. This initial table is split by customer, by month, providing 13 months of customer data spanning from April 04 to April 05.

6.5.2. Most suitable for Creating the Neural Network Model

Each of the months for case 1 are queried against the actual churn data to see how many customers churn in each month. The month where the most customers churned was November with 290 churners, therefore it was decided that the neural network model will be based on this month as this month should contain more churn scenarios than the other months. The number of non-churners in the training dataset is reduced to a total number of 1148 in an attempt to prevent the non-churners dominating the NN model. This corresponds to a 20% churn ratio. A table is created in MS Access to hold the training set.

The training dataset is copied and pasted into MS Excel and resaved as a text file. This text file is then imported into Matlab where an input vector is compiled from all customer data and a target vector is created from the churn flag data. The inputs and target are then imported into the Matlab NN toolbox. This data is then used to create the neural network. Each neural network is validated by applying it to the full dataset of 8409 customers.

6.5.3. NN Experiment 1

A single layer/single neuron NN was tested by setting it to predict the churners for the full set of November data. This is the preferred convergence test as stated in Demuth, (2007) to determine how well, if at all, the network has trained to the desired task. The confusion matrix in Figure 72 illustrates the results: -

```
        | Estimated Labels
 True   |
Labels  |    0       1   | Totals
--------|----------------|-------
0       | 7639    480    | 8119
1       |   68    222    |  290
--------|----------------|-------
Totals  | 7707    702    | 8409
```

Figure 72: Accuracy Results For The 1 Layer, 1 Neuron NN

6.5.4. NN Experiment 2

The second neural network configuration contains 2 hidden layers, each containing 3 neurons each. This configuration can be viewed in Figure 59. The results obtained from this neural network configuration can be seen from the confusion matrix in Figure 73:

```
        | Estimated Labels
 True   |
Labels  |    0       1   | Totals
--------|----------------|-------
0       | 7997    122    | 8119
1       |  100    190    |  290
--------|----------------|-------
Totals  | 8097    312    | 8409
```

Figure 73: Accuracy Results for the 2 Layer, 3 Neuron NN

As can be seen from the confusion matrix in Figure 73, with the increased complexity of this NN architecture, the churn prediction accuracy has slightly decreased but the non-churn misclassification rate has been reduced by almost 75%.

6.5.5. NN Experiment 3

It is noticeable from the results obtained from the 1st neural network configuration, and the more complex 2nd neural network configuration, that an increase in the complexity of the neural network has also increased the churn prediction accuracy. The final NN configuration is even more complex, so it is anticipated that the churn prediction accuracy will increase further. The confusion matrix obtained from NN experiment 3 is displayed in Figure 74:

```
        | Estimated Labels
 True   |
Labels  |    0      1   | Totals
--------|---------------|-------
0       | 7798    321   | 8119
1       |   76    214   |  290
--------|---------------|-------
Totals  | 7874    535   | 8409
```

Figure 74: Accuracy Results for the 4 Layer, 6 Neuron NN

The results obtained from the final NN experiment are very similar to those results of the first single layer single neuron network shown by the confusion matrix in Figure 72. The only significant difference between the results is a decrease in the total number of non-churn misclassifications from the four layer six neuron network, however there are over 2.5 times more non-churn misclassifications generated by NN 3 over NN 2. So that an accurate comparison of the results for all three experiments can be achieved the churn and non-churn accuracies for each experiment were converted to a percentage and recorded in Table 12:

Table 12: NN Prediction Accuracy Comparison

Case 1	Churn Capture Accuracy	Non-Churn Capture Accuracy	Total Accuracy
NN1 Experiment	76.50%	94.00%	93.40%
NN2 Experiment	65.50%	98.40%	97.40%
NN3 Experiment	73.80%	96.00%	95.30%

It can be observed from Table 12, NN1 experiment achieved a total prediction accuracy of 93.4%. The total prediction accuracy has been calculated by adding the correct number of classified non-churn with the correct number of classified churn, dividing the number by the total dataset size of 8409 and multiplying by 100. NN2 achieved a 97.4% total accuracy and NN3 achieved a 95.3% total accuracy. From this information it is clear that the NN2 configuration has provided the best overall results for classifying dataset 1.

6.5.6. Generating Churn Index Values

The results obtained from NN experiment 2 shown in Figure 73 are appropriate for generating churn index values because they provide a significant number of actual churn matches with a decrease in non-churn misclassifications over those generated from NN 1 and NN 3. The 1st NN configuration has identified a similar number of churners as the 3rd NN experiment but a higher non-churn misclassification rate.

The 2nd NN configuration is used to generate churn index values for each customer over each month of the available historical dataset for analysis by the customer profiling methodology.

Once the churn index values for each month are generated they are exported out of Matlab and inserted into MS Excel. The first Excel column is the customer ID number, column 2 is the churn index values for April 04, Column 3 May 04, etc. After all the months are inserted, the churn column is added to the worksheet. This column stores as a numerical indicator depending on the month churn occurred. E.g. If the first month of data in the time sequence is April 04 then the corresponding churners for April 04 are stored as the integer '1'. May 04 churners are given the integer '2', June 4 the integer '3' etc. This is because the profiling methodology attempts to identify a pattern of how long it takes customers to churn after their loyalty index values fall below a pre-defined churn threshold. If the churn index value falls below the churn threshold in month 2 and actual churn occurs in month 4 then that particular customer took 2 months to churn. If all customers belonging to a profile take 2 months to churn we have determined that all customers being matched to that profile should also theoretically take 2 months to churn.

An example of the required customer churn index values that are used for analysis by the churn profiling software can be seen in Figure 75:

ID	Apr-04	May-04	Jun-04	Jul-04	Aug-04	Sep-04	Oct-04
1	0.0000	0.0027	0.9961	0.0156	0.0020	0.0004	0.0000

Figure 75 Churn Index Values

It is decided in advance which months will be used as inputs. E.g. months April 04 – October 04 will be used as input into the profiling methodology. The best profiles for identifying customer churn will be estimated using the churn data from Nov 04, Dec 04 and Jan 05. Allowing 3 months of future churn information for the classification of master profile clusters should be sufficient for determining any profile trends. This will then leave enough data from the case 1 dataset to validate those profiles that have been identified as the strongest predictors of future churn.

The constructed worksheet is saved as text file for importing into a database. The profiling methodology has been designed to use a MySQL database, so the customer churn index data is imported as a MySQL table.

6.5.7. Applying Customer Profiling Methodology

The churn index values such as the ones shown in Figure 75 are presented to the churn profiling software. The profiling software methodology performs its analysis in the sequence:

Convert the churn index values to loyalty index values using the formula ***Loyalty index = 1 – churn index***

Assess the loyalty index values to see if their has been fluctuating activity for the customer

Check to see if loyalty profiles already exists for determined loyalty time series pattern.

If the profile exists, assign the customer that profile.

If the profile does not exist check to see if the customers loyalty value at any point falls below a given churn threshold (The value 0.3 has been defined as a default value because this value provided best accuracy during the initial methodology experiments).

If the loyalty index value has fallen below the churn threshold check to see if the customer is recorded as an actual churner.

If the customer has been recorded as an actual churner, add the profile for that customer as a future master profile.

If the customer is not an actual churner or the customers loyalty index value has not fallen below the churn threshold then do nothing.

Three new columns are added to the results reflecting the prediction accuracy of each master profile. The column 'Total Capture' displays how many of the total cluster size churned within the time sequence being analysed. The column 'Future Capture' shows how many customers belonging to that profile cluster actually churned within the future three months between Nov 04 and Jan 05. The 'Hi/Low' column determines the profiles that are most accurate at capturing future churn. The full profile results are displayed in Table 13:

Table 13: Case 1 Full Profiling Information

ID	profile	Profile Count	Local Churn	Actual Churn	Total Capture	Future Capture	Hi/Low
11	ULD	9	2	5	77.78%	55.56%	High
86	ULLDD	2	1	1	100.00%	50.00%	High
38	ULDD	5	1	2	60.00%	40.00%	High
15	UD	78	37	29	84.62%	37.18%	High
2	D	369	189	125	85.09%	33.88%	High
84	UUULUD	3	1	1	66.67%	33.33%	High
5	DD	70	37	16	75.71%	22.86%	High
1	UUD	44	18	10	63.64%	22.73%	High
37	UULD	14	1	3	28.57%	21.43%	Low
57	UUUUD	5	1	1	40.00%	20.00%	Low
77	UUUDDD	20	1	4	25.00%	20.00%	Low
12	UDD	42	16	8	57.14%	19.05%	Low
20	UDDUDD	42	12	7	45.24%	16.67%	Low
23	UUUD	19	1	3	21.05%	15.79%	Low
7	DUD	52	19	8	51.92%	15.38%	Low
16	DUUUDD	67	4	10	20.90%	14.93%	Low
22	UDUDUD	149	42	22	42.95%	14.77%	Low
73	ULDUUD	14	1	2	21.43%	14.29%	Low
71	UUUUUD	22	2	3	22.73%	13.64%	Low
43	DUUD	50	2	6	16.00%	12.00%	Low
10	DUUULD	116	20	13	28.45%	11.21%	Low
17	DDUUUD	91	4	10	15.38%	10.99%	Low
78	DUDUDD	57	3	6	15.79%	10.53%	Low
30	UDUUUD	260	10	27	14.23%	10.38%	Low
6	UUDUUD	203	37	21	28.57%	10.34%	Low
25	UUULLD	137	4	13	12.41%	9.49%	Low
55	UUDUDD	64	2	6	12.50%	9.38%	Low
52	UULDUD	43	1	4	11.63%	9.30%	Low
26	UUDDUD	59	3	5	13.56%	8.47%	Low
96	ULLLLD	12	1	1	16.67%	8.33%	Low

97	UUUULD	12	1	1	16.67%	8.33%	Low
39	UUUDUD	102	3	8	10.78%	7.84%	Low
4	UULDD	13	3	1	30.77%	7.69%	Low
8	DUDD	52	6	4	19.23%	7.69%	Low
29	UULLLD	104	3	8	10.58%	7.69%	Low
33	UULLDD	13	1	1	15.38%	7.69%	Low
62	DUUDUDD	78	1	6	8.97%	7.69%	Low
83	UUUUDD	13	1	1	15.38%	7.69%	Low
90	DUUUDDD	66	1	5	9.09%	7.58%	Low
31	UDUUD	40	1	3	10.00%	7.50%	Low
68	UDDD	41	1	3	9.76%	7.32%	Low
53	UUDUD	47	2	3	10.64%	6.38%	Low
60	DUUUDUD	110	2	7	8.18%	6.36%	Low
56	DDUDUD	63	1	4	7.94%	6.35%	Low
42	DUUDUD	79	3	5	10.13%	6.33%	Low
3	DUUDD	49	7	3	20.41%	6.12%	Low
36	ULLDUD	17	1	1	11.76%	5.88%	Low
50	DUDDD	51	1	3	7.84%	5.88%	Low
76	DUUUULD	55	3	3	10.91%	5.45%	Low
58	DUDUUDD	75	1	4	6.67%	5.33%	Low
49	DUDUDUD	81	1	4	6.17%	4.94%	Low
67	UDUUDD	61	2	3	8.20%	4.92%	Low
28	UDUD	42	1	2	7.14%	4.76%	Low
69	DUUDDUD	63	1	3	6.35%	4.76%	Low
61	DUUUUD	73	1	3	5.48%	4.11%	Low
54	DDDUUD	49	1	2	6.12%	4.08%	Low
94	DUUDDD	49	1	2	6.12%	4.08%	Low
35	DUUUD	50	1	2	6.00%	4.00%	Low
19	DUDUUD	76	20	3	30.26%	3.95%	Low
44	DUDUD	52	1	2	5.77%	3.85%	Low
51	DUUULLD	52	1	2	5.77%	3.85%	Low
81	DUDDUUD	54	1	2	5.56%	3.70%	Low
105	DUUDUUD	29	1	1	6.90%	3.45%	Low
79	DUDUUUD	71	1	2	4.23%	2.82%	Low
48	UDUULD	40	1	1	5.00%	2.50%	Low
72	UDDUD	41	1	1	4.88%	2.44%	Low
9	UDUDDD	42	2	1	7.14%	2.38%	Low
21	UUDD	43	7	1	18.60%	2.33%	Low
85	DUULDD	44	1	1	4.55%	2.27%	Low
74	DDDD	49	1	1	4.08%	2.04%	Low
82	DDDUDD	49	1	1	4.08%	2.04%	Low
18	DDUD	50	6	1	14.00%	2.00%	Low
47	DUDDUDD	51	1	1	3.92%	1.96%	Low
91	DUDDUD	51	1	1	3.92%	1.96%	Low
41	DUDULD	52	1	1	3.85%	1.92%	Low
93	DUUUUUD	54	1	1	3.70%	1.85%	Low
13	DDUUDD	60	12	1	21.67%	1.67%	Low
14	DDUDDD	50	3	0	6.00%	0.00%	Low
24	UDDDUD	41	1	0	2.44%	0.00%	Low
27	ULDUD	5	1	0	20.00%	0.00%	Low

32	DDD	49	1	0	2.04%	0.00%	Low
34	ULUUDD	2	1	0	50.00%	0.00%	Low
40	UDUDD	43	1	0	2.33%	0.00%	Low
45	DDDUD	49	1	0	2.04%	0.00%	Low
46	ULDUDD	5	1	0	20.00%	0.00%	Low
59	UUUDD	20	1	0	5.00%	0.00%	Low
63	UDULDD	40	1	0	2.50%	0.00%	Low
64	UULDDD	13	1	0	7.69%	0.00%	Low
65	ULDDD	4	1	0	25.00%	0.00%	Low
66	DDDDDD	49	1	0	2.04%	0.00%	Low
70	ULLD	2	1	0	50.00%	0.00%	Low
75	UULLD	12	1	0	8.33%	0.00%	Low
80	DDDDUD	49	1	0	2.04%	0.00%	Low
87	UULLUD	4	1	0	25.00%	0.00%	Low
88	UUULDD	7	1	0	14.29%	0.00%	Low
89	UULUDD	1	1	0	100.00%	0.00%	Low
92	UUULD	7	1	0	14.29%	0.00%	Low
95	UUDDD	10	1	0	10.00%	0.00%	Low
98	DULUDD	44	1	0	2.27%	0.00%	Low
99	UDDUUD	42	1	0	2.38%	0.00%	Low
100	ULLLUD	2	1	0	50.00%	0.00%	Low
101	UUDLD	10	1	0	10.00%	0.00%	Low
102	DULUD	25	1	0	4.00%	0.00%	Low
103	DLUD	24	1	0	4.17%	0.00%	Low
104	DUUUUDD	27	1	0	3.70%	0.00%	Low
106	DDUDD	20	1	0	5.00%	0.00%	Low

In Table 13 the first column displays the unique profile ID number. It can be noticed that these ID numbers are not in order. This is because the profiles are sorted in order of most accurate to least accurate at capturing future churn. The table is ordered by future churn and not local churn because the main goal of the methodology is the capture of future churn. Identifying the strongest profiles for capturing future churn should result in profiles that are continuously strong at capturing future churn. The second column shows the characteristics of each profile. ***D*** represents a fall in the customer's loyalty index, ***U*** represents a rise in the customer's loyalty index, and ***L*** shows that the customer's loyalty index remained static. The third column shows how many customers' have been detected to match that profile. The fourth column displays how many customers belonging to that specific profile cluster churned within the analysis period. The fifth column displays how many customers belonging to that profile churned within a three month period future of the analysis period. The sixth column shows the total percentage of each profile cluster that actually churned. The seventh column displays the total percentage of the profile cluster who churned in the 3 month period future of the analysis period and the eighth column shows if that profile is high or low risk for future churn capture.

The methodology also attempts to asses the time it took for a customer to churn after the detection of a significant fall in customer loyalty. This calculation is updated every time a customer is assigned to a specific master profile group and is detected as being an actual churner. The reason is that it helps to build a picture of the retention potential for each specific profile group.

The first master profile in the list (profile ID 11) shows a an initial raise in the customer's loyalty index which then levels off for one month before a major drop that brings the loyalty index below the churn threshold. Only 9 customers fell into this profile, out of which a total of 7 customers actually churned and 5 of these customers churned in the three month future period. This profile can be seen in Figure 76:

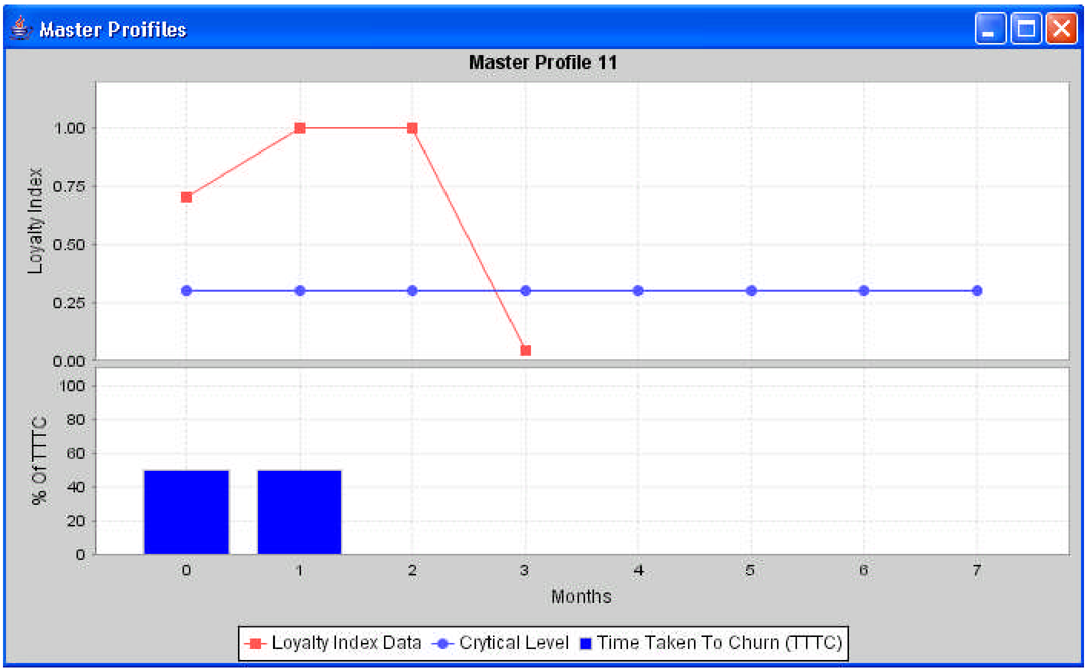


Figure 76: First Most Significant Profile for Future Churn Capture

As shown from the example master profile in Figure 76 the profile separates into two sections. A line chart at the top of the master profile display window provides a visualisation of the customer's loyalty index pattern. In the case of master profile 11ot shows an initial raise in the customer's loyalty index which remains static for 1 month before a major decline in loyalty value that brings the customer's loyalty level below the churn threshold. The bar chart at the bottom of the window illustrates how long it took the customers who churned within the analysis period to churn after their loyalty index values fell below the churn threshold. This chart shows that for the case of master profile 11 50% of the customers churned within the same month as the event that caused the major fall in customer loyalty and 50% churned 1 month after the significant fall in customer loyalty. Only 2 customers churned within the analysis period for master profile cluster 11. More information is required to really make the 'probability of churn' charts useful. This chart will become more accurate and more valuable as more customers are assigned to that profile. The second most significant master profile is shown in Figure 77:

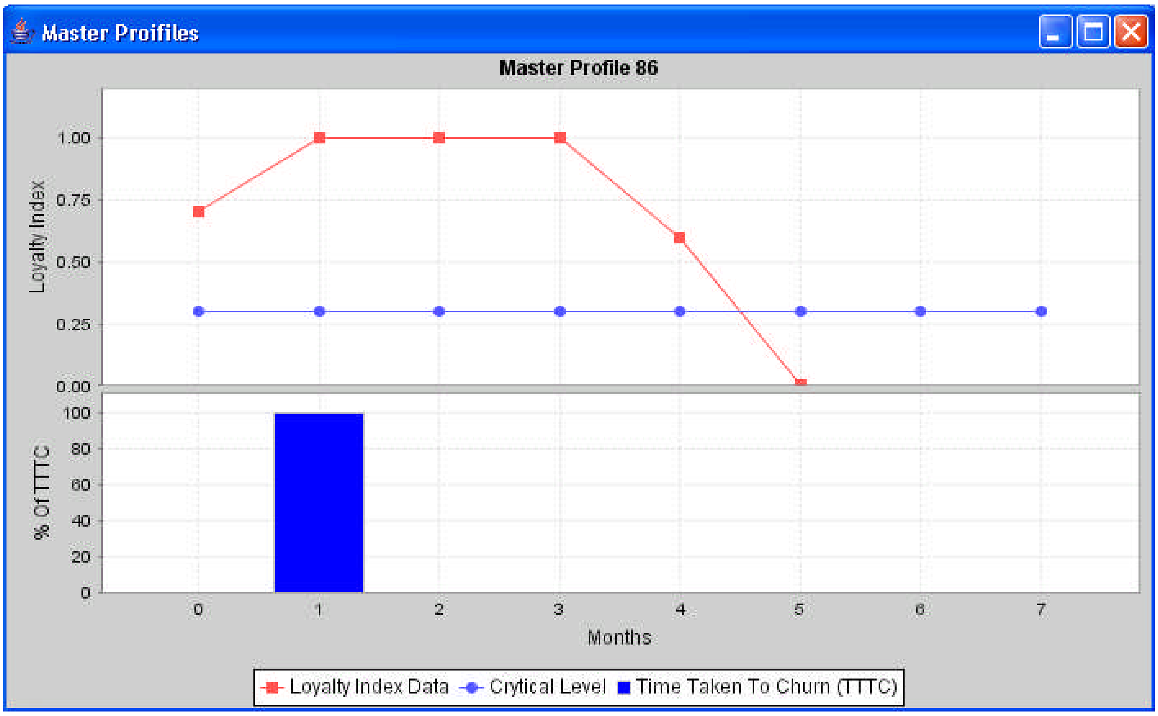


Figure 77: Second Most Significant Profile for Future Churn Capture

As can be seen from Figure 77 master profile cluster 86 begins showing a decreased loyalty value. No event is detected in month 2, no event is detected in month 3, an event in month 4 decreases loyalty; however the decrease is not significant enough to lead to churn while a final event in month 5 is of major significance bringing the customers loyalty index to below the churn threshold.

The probability of retention chart at the bottom of Figure 77 displays a 100% probability of retention within 1 month of the customer's loyalty index first falling below the churn threshold. Only 1 customer churned within the analysis period and more data is required before this chart could be considered useful. The third most significant master profile for future churn capture is displayed in Figure 78:

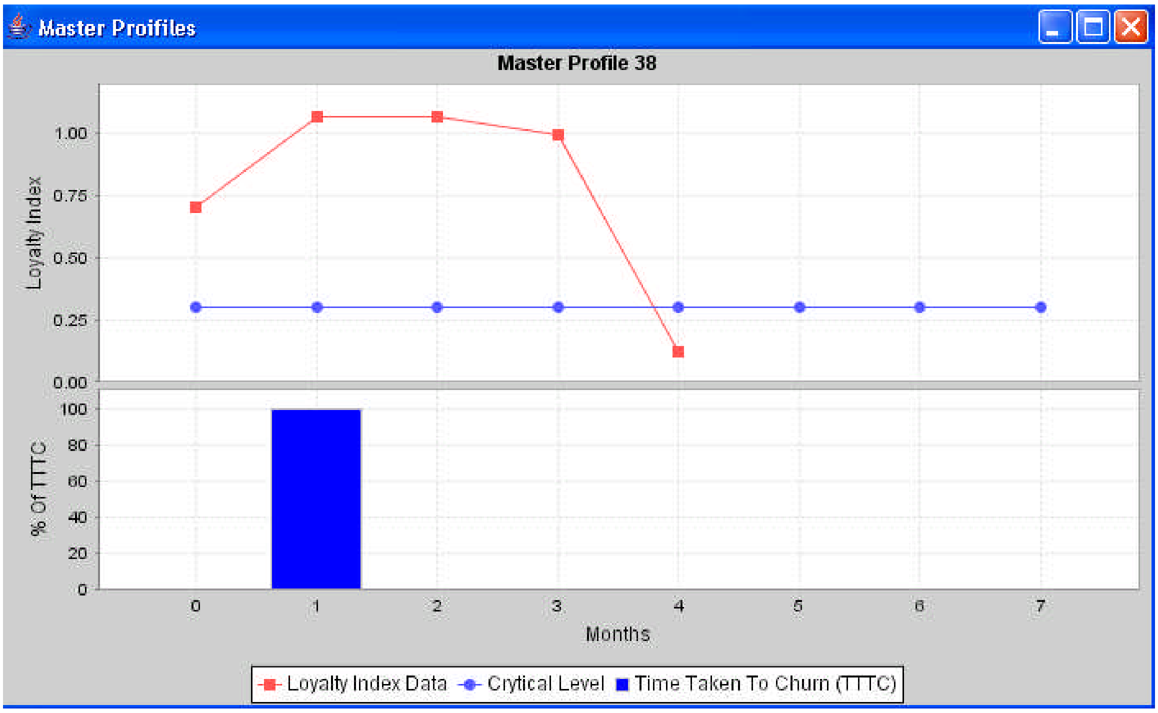


Figure 78: Third Most Significant Profile for Future Churn Capture

As can be seen from Figure 78 the third best profile for the capture of future churn is master profile 38. Master profile 38 begins with an initial decline in customer loyalty. No event is detected in month 2, a decline in loyalty is detected in month 3, although this decline is not significant enough to lead to customer churn and a final major event in month 4 is significant enough to lead to customer churn. Only 5 customers where matched to this profile; however 3 out of 5 of these customers actually churned and 2 of those customers churned future to the initial analysis period. Only 1 customer churned within the initial analysis period and that customer churned 1 month after his/her loyalty index value fell below the churn threshold. More information is required to make the probability of retention chart useful. The forth most significant master profile is displayed in Figure 79:

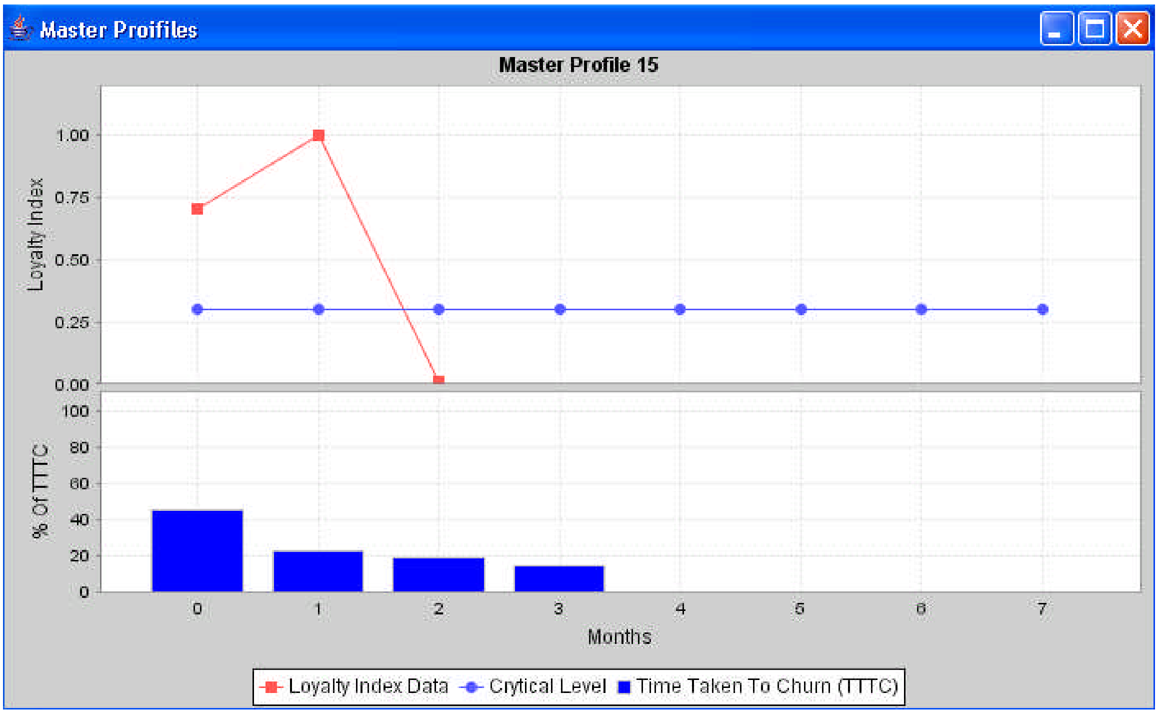


Figure 79: Forth First Most Significant Profile for Future Churn Capture

Even though the previous 3 profiles have shown greater accuracy for predicting future churn, none have had large numbers of customers classified to them. Master profile 15 has been detected as the forth best profile for future churn capture however there has been a significant increase in the number of customers who have been classified to this profile, which actually makes it more valuable. A total of 78 customers were classified to master profile cluster 12 from which 66 of those customers actually churned 37 customers churned within the analysis period and 29 customers churned in the three months future to the analysis period. From the 37 customers who churned within the analysis period 45% churned immediately so 55% of customers churned a minimum of 1 month after their loyalty index values fell below the churn threshold. The probability of retention chart is now showing greater level information. A company could be 55% confident that a customer could be retained up to 1 month after their loyalty index value had fallen bellow the churn threshold, if matched to customer profile 15. This profile also shows that 84.62% of customers matching it have actually churned. Out of the total number of customers classified to master profile cluster 15, 38.62% churned within the 3 months future of the analysis period. The fifth master profile most accurate for future churn prediction is displayed in Figure 80:

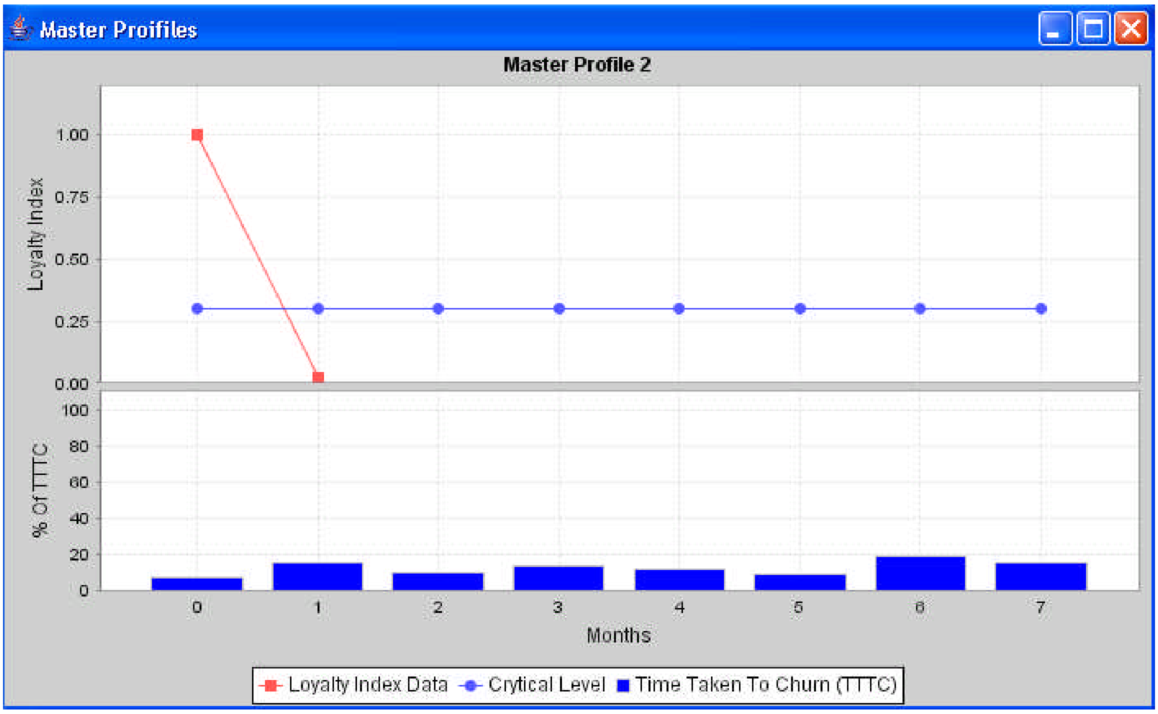


Figure 80: Fifth Most Significant Profile for Future Churn Capture

Figure 80 displays the fifth most significant master profile for future churn prediction. Master profile cluster 2 displays one major event that leads to customer churn. This is the most significant master profile in terms of cluster size. 369 customers where assigned to master profile cluster 2. Out of these 369 customers 314 customers actually churned. 189 customers churned within the analysis period and 125 customers churned in the three months future to the analysis period. From the 189 customers who churned within the analysis period only 7% churned immediately, which means 93% churned a minimum of 1 month after their loyalty index values fell below the churn threshold. For the case of master profile cluster 2 the probability of retention chart displays some very interesting and important information regarding the cluster. Immediately a company can be 85% confident that customers being assigned to the master profile cluster will inevitably churn and 93% confident that those customers will be retainable if a successful retention campaign is deployed within 1 month of the customers loyalty index initially falling below the churn threshold. Master profile cluster 2 is the most important master profile cluster because it captures such a large portion of the churning customer base. Traditionally it had been assumed that customers moving instantly from loyal to churn were spontaneous churners and retention would be unlikely. The profiling methodology shows that this is actually one of the most significant and retainable scenarios. The sixth most accurate future churn profile is displayed in Figure 81:

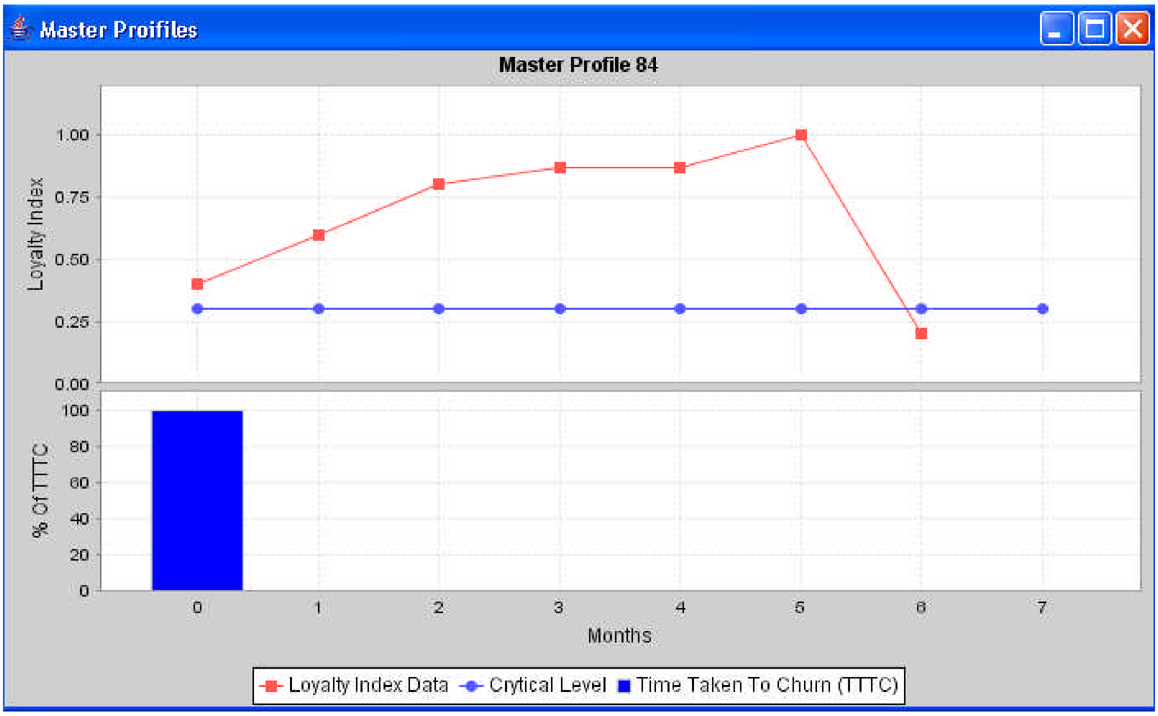


Figure 81: Sixth Most Significant Profile for Future Churn Capture

The sixth most significant begins with displaying an initial event which has not been significant enough to lead to churn. This is followed by a second event which has had less impact than the first. A third event follows that has had less impact than the first or second. A fourth event is on par with the third event. No events are detected in the 5 th month; however in the sixth month an event occurs that brings the customers loyalty index below the churn threshold. Only 3 customers where categorised to match master profile 84. From these 3 customers, one customer churned during the analysis period and that customer churned immediately. One customer churned within the three month period future of the analysis period and the final customers is not detected as a churner. More information is required to make this profile really useful. The seventh most accurate customer profile for future churn prediction is displayed in Figure 82:

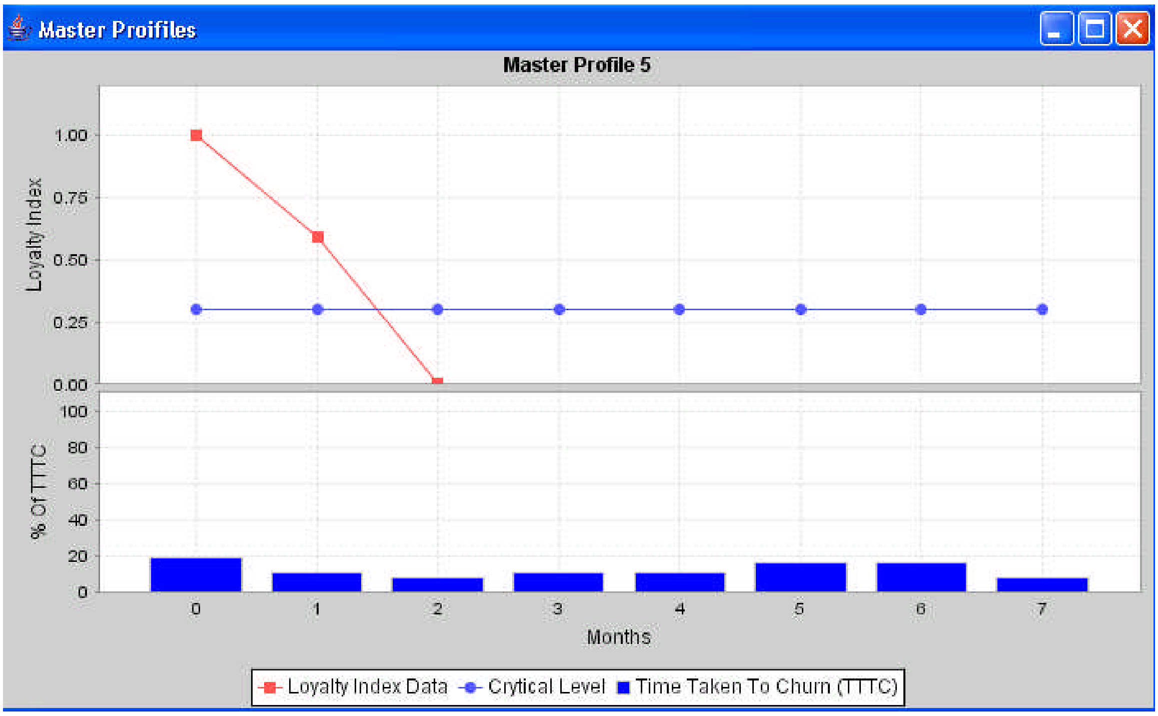


Figure 82: Seventh Most Significant Profile for Future Churn Capture

The seventh most significant master profile cluster that has been detected as being one of the strongest for future churn capture is master profile cluster 5. Master profile cluster 5 begins with no activity. There is then an event in month 2 that is not significant enough on its own to cause the customer to churn; however in the third month an event occurs that is significant to cause the customer to churn. 70 customers have been classified to the master profile 5 cluster. From these 70 customers, 37 customers churned within the analysis period and 16 customers churned within the three month future period. Therefore 53 customers out of 70 assigned to this cluster actually churned, 16 of which churned in the future. The final most significant master profile for future churn capture is displayed in Figure 83:

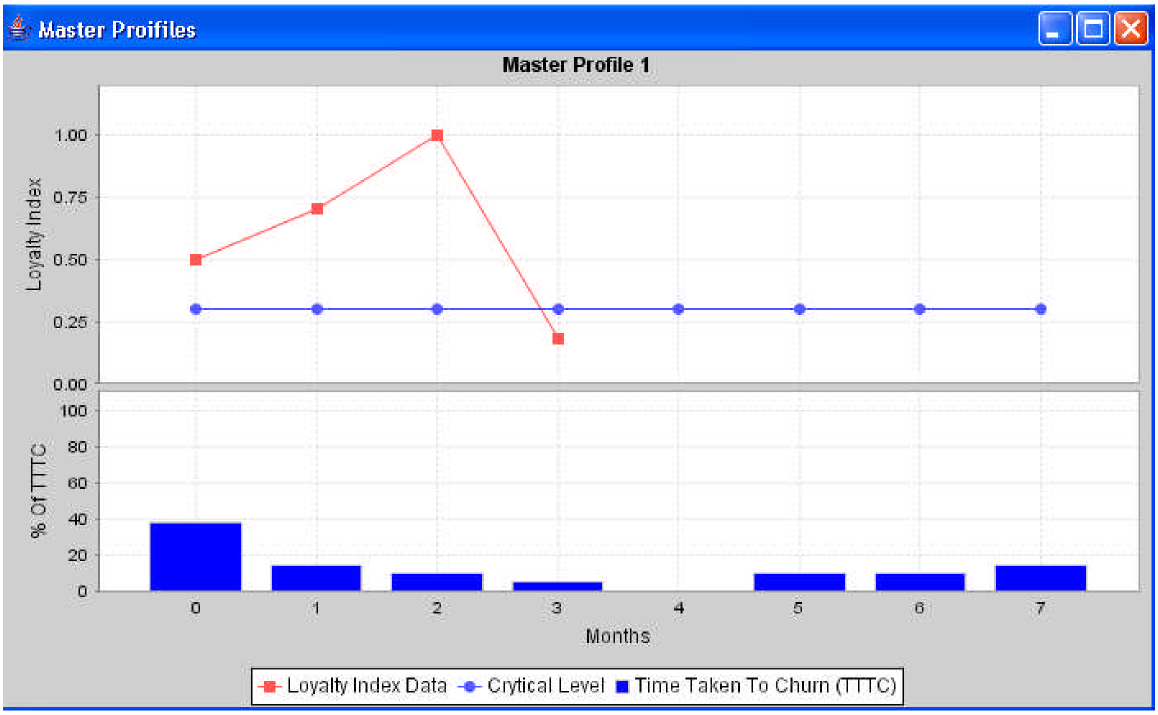


Figure 83: Eighth Most Significant Profile for Future Churn Capture

Figure 83 displays the sixth strongest master profile for future churn prediction (Master profile cluster 1). The characteristics of master profile cluster 1 are an initial event that is not in its self significant enough to lead to churn. Month 2 displays a second event that has less impact on customer loyalty than the first event. The third month shows no activity and in the forth month there is a major event that inevitably leads to customer churn. 44 customers were assigned to master profile cluster 1, from which 18 customers churned within the analysis period and 10 customers churned within the three months future to the analysis period. From this information it can be determined that a total of 63.64% of customers assigned to master profile cluster 1 actually churned.

Keeping the emphasis on future churn prediction, 758 customers churned in the three month period immediately future of the analysis period. If all profiles were used as churn predictors the following results would be obtained for churn capture three months future of the last analysis month:

True Labels	Estimated Labels 0	1	Totals
0	2860	4790	7650
1	266	492	758
Totals	3126	5282	8408

Figure 84: Using All Profiles as Future Predictors

As can be seen from Figure 84, if all master profile clusters were to be used as future churn predictors the overall results would be very poor. The actual churn capture is high; achieving a total future churn capture of 65%; however the benefits of these results are cancelled by the extremely high misclassification rate, where in total only 37% of non-churners have been accurately classified. A hit ratio for these results would equal 0.09 showing a very poor model. This is where optimisation by clustering profiles into high and low risk categories is required. After the most accurate profiles for future churn prediction have been identified and analysed, the results from Figure 84 re transformed into those displayed in Figure 85:

True Labels	Estimated Labels 0	1	Totals
0	7261	391	7652
1	567	189	756
Totals	7828	580	8408

Figure 85: Limiting Predictions to Top Profiles Only

It can be seen from Figure 85 that as a result of limiting the future predictions to those profiles that display the best accuracy for future churn capture the correct future classifications have been reduced from 492 to 189 customers, which converts to a total of 25% capture; however the misclassifications of non-churners as churners has been reduced from 4790 to 391 resulting in a total of 95% non-churn capture. The target of the profiling methodology is not necessarily to capture as much churn as possible because future churn capture is more reliant on data quality rather than the predictive model itself. The key to the best predictive model is achieving the highest possible hit ratio while at the same time maintaining as many future predictions as possible. 25% future churn capture and 95% non-churn capture demonstrates high accuracy for the data that is the source of analysis. On top of these percentages a high hit ratio helps to define a measure of the power of the methodology. The hit ratio is basically the total number of predictions divided by the total number of correct classifications. This hit ratio of 0.09 achieved from Figure 84 is greatly improved from by restricting the future churn classifiers to high risk clusters only as displayed in Figure 85. The total number of predictions has been reduced to 580 while the correct future churn classifications have been reduced to 189 (25% of all future churn over a three month period). The greatest affect has been on the hit ratio which is now the product of 189/580, equalling an improved hit ratio of 0.32. This means that for every 3 of the customers predicted as future churners, 1 customer will be as actual churner. A final stage can be added to increase these results even further. Out of the six most accurate customer profile clusters that have the strongest connection to future churn, a majority leave the company during the actual analysis period. In total 286 customers from the top profiles actually churned during the months that were being analysed to create the profiles. Therefore these customers can be removed from the overall predictions. Once these defectors have been removed from the prediction totals, the initial misclassification total of 391 is reduced to 105. The resulting impact from removing these customers is displayed from the confusion matrix in Figure 86:

```
       | Estimated Labels
 True  |
Labels |    0      1   | Totals
-------|---------------|-------
0      | 7547    105   | 7652
1      |  567    189   |  756
-------|---------------|-------
Totals | 8114    294   | 8408
```

Figure 86: Customers' Who Churn During the Analysis Period Removed

As can be seen from Figure 86 once the customers who defect during the analysis period have been removed from the future classifications the results are enhanced even further. The future prediction capture is not affected; however there is a significant impact on the misclassification rate and therefore the hit ratio. The hit ration obtained from the results displayed in Figure 85 are of good quality, as according to an industrial survey (please see appendix a) the current industrial average hit ratio is around 0.25 (1 churner in every 4 predictions). With the further misclassification reduction displayed in Figure 86, the hit ration has become the product of 189/294, equalling a hit ratio of 0.64. The correct churn classification has remained constant at 25% while the non-churn accuracy has risen to 98.6%.

6.5.8. Determine future churn capture accuracy

Case 1 contains 13 months of data. The first 7 months have provided the input to the profiling methodology, while the subsequent 3 months have been used for classification of future churn accuracy for each of the master profile cluster. Referring back to Table 13, it can be seen that master profiles 11, 86, 38, 15, 2, 84, 5 and 1 can be categorised as strong churn classifiers. The future churn captured from these profiles total 189 from 758 customers who actually churn three months future of the analysis period. This is 25% of all churn over the future period. The most important aspect of the research though is the enhancement of the hit ratio, as the hit ratio in essence defines the strength of the predictive methodology. The overall hit ratio achieved from the test phase is 0.64. This means for every 1.5 customer's contacted 1 customer will be a future churner. 10 months of data has been required to generate the customer profiling predictive model and Case 1 contains a total of 13 months of data. The analysis window is moved forward 4 months to fully test the predictive accuracy of the eight master profiles that have been identified as being the best for future prediction using months August to February as inputs for generating profiles for the prediction of the final 2 months of the dataset, March and April, the results displayed in Table 17 have been generated from the model:

Table 14: Future Churn Predictions From Strongest Master Profile Clusters

ID	Profile	Cluster Size	Local churn	Future Churn	Total Churn Accuracy	Future Churn Accuracy
11	ULD	2	1	0	50.00%	0.00%
86	ULLDD	4	1	0	25.00%	0.00%
38	ULDD	1	0	0	0.00%	0.00%
15	UD	163	119	32	92.64%	19.63%
2	D	355	188	100	81.13%	28.17%
84	UUULUD	0	0	0	0.00%	0.00%
5	DD	117	76	22	83.76%	18.80%
1	UUD	40	38	2	100.00%	5.00%

The results displayed in Table 14 have been generated by analysing the inputs of August to February for master profile classification, and then using only the strongest master profile clusters as determined from the model creation stage to classify future churn. The profiles are very strong for total churn capture with the exception of two. Profile cluster 38 has only member classified to it and that customer has not been identified as an actual churner. Master profile cluster 84 has had no members assigned to it. At the initial model creation stage master profile cluster 38 only had 5 customers classified to it and master profile cluster 84 only had 3 customers classified to it therefore the results have remained inline with those determined through via model creation. The actual future churn capture for the remainder of the dataset is displayed using the confusion matrix in Figure 87:

```
        | Estimated Labels
 True   |
Labels  |    0       1   | Totals
--------|----------------|--------
0       | 7381     526   | 7907
1       |  345     156   |  501
--------|----------------|--------
Totals  | 7726     682   | 8408
```

Figure 87: Case 1 Initial Future Predictions

As can be seen from Figure 87 the initial future predictions have caught 31% of future churn. The results as they are have achieved a hit ratio of 0.22 so basically from every four predictions, one is an actual churner. As identified from the industrial survey (Appendix A) this is about the industry average. Next the customers who actually churned during the analysis period (August to February) are removed from the future predictions (March and April). The results are updated using the confusion matrix in Figure 88:

```
       | Estimated Labels
 True  |
Labels |     0       1  | Totals
-------|----------------|-------
0      | 7796    111  | 7907
1      |  345    156  |  501
-------|----------------|-------
Totals | 8141    267  | 8408
```

Figure 88: Case 1 Future Predictions Less Churn From Analysis Period

It can be seen from the confusion matrix in Figure 88 that once the customers who churned during the analysis period are removed from the future predictions the non-churn accuracy is boosted to 98.5%, the future churn accuracy remains at 31% and the hit ratio is increased to 0.58. These results are very close to those obtained from the model creation stage showing that the strongest future churn master profile clusters have retained there accuracy. A comparison of the results obtained from the model creation stage and the pure future prediction stage is displayed in Table 15:

Table 15: Result Stability Comparison

	Model Creation Stage	Future Churn Capture
Case 1 Result Comparison		
Non-churn Accuracy	98.60%	98.50%
Churn Accuracy	25.00%	31.00%
Hit Ratio	0.64	0.58

As shown from Table 15, the results obtained from the model creation stage and the pure future churn capture stage have remained stable. The non-churn accuracy is virtually the same; the churn accuracy is similar; however the results have actually increased for future churn capture; and the hit ratio has remained reasonably stable at a level significantly higher than the current industry average.

6.5.9. Comparing Results with Two Alternative Predictive Methods

To further validate the customer profiling methodology two papers were identified as attempting to predict future churn, however the research for these paper target alternative service sectors. Both of these papers document segmenting the customer database into two groups based on loyalty values, and contacting all customers in the resulting 'high risk' group as a retention strategy to minimise churn.

The profiling methodology uses a loyalty index value itself as part of the process of predicting churn, so these index values will be used as the basis of analysis for each of the comparative studies. These two methodologies were chosen as the documented research on future churn prediction is extremely limited and the two chosen methodologies show similarities with the profiling methodology in terms of generation of a loyalty index and using that to try to determine future churn.

6.5.9.1. Hu (2005) Methodology

Hu (2005) propose a data mining approach for analysing retail bank customer attrition by use of a customer ranking based process based on a loyalty index score.

These experiments used the same method to generate the customer loyalty index as used by the customer profiling methodology. The first reason is that the customer profiling methodology uses the loyalty index value, so the resulting overall methodologies should compare accurately and regardless of how loyalty index was created. The second reason is that Hu (2005) based there research on bank customers within the financial service sector and their loyalty index scoring method is target towards this industry. Their actual methodology for classifying churn from a loyalty index value is replicated here for a comparison with other methodologies.

Hu (2005) propose a process of determining a customer loyalty index value for each customer, then splitting the customer base into two distinct groups (defectors and non-defectors). The defector group contains the top 4% of the customer base with the lowest loyalty index values. All the customers are then contacted with retention strategies.

The method of clustering customers into two groups and classifying the top 4% is applied to the loyalty index values from February 05 of the case 1 data and compared against the churn data for April 05 so that a comparison can be made with the profiling methodology. The results are presented in Figure 89:

```
        | Estimated Labels
 True   |
Labels  |     0       1   | Totals
--------|-----------------|-------
0       | 7869     324    | 8193
1       |  203      12    |  215
--------|-----------------|-------
Totals  | 8072     336    | 8408
```

Figure 89: Hu (2005) Methodology

In Figure 89 the churn capture is about 6% while the misclassification rate is high. The hit ratio for the Hu (2005) methodology is 12/336 = 0.03.

6.5.9.2. Hwang et al. (2004) Methodology

The research presented by Hwang et al. (2004) is more closely related to the authors research than the research of Hu (2005) because both the author's research and the research of Hwang et al. (2004) are targeted towards the telecommunications industry.

Hwang et al. (2004) propose a method of clustering the customer base into groups based on loyalty index value. They suggest classifying all customers with a loyalty index value lower than 0.5 as defectors. Because this research is also targeted towards the telecommunications industry the loyalty index generation is similar.

The methodology of Hwang et al. (2004) is again applied to the month of March 05 for the case 1 data, in an attempt to predict churn in April 05. The results obtained from this experiment can be viewed in Figure 90:

True Labels	Estimated Labels 0	Estimated Labels 1	Totals
0	7848	345	8193
1	203	12	215
Totals	8051	357	8408

Figure 90: Hwang et al. (2004) Methodology

As shown in Figure 90, the results obtained from the Hwang et al. (2004) methodology is similar to those obtained from Hu (2005) methodology with the only real difference being a slightly higher misclassification rate to Hu (2005) methodology. The hit ration for the Hwang et al. (2004) methodology is 12/357 = 0.03. These results are now compared with those generated from the customer profiling methodology of the author.

6.5.9.3. Customer Profiling Methodology

To compare with the experiments performed for Hu (2005) and Hwang et al. (2004), the results obtained the future results of the customer profiling methodology as shown in Figure 88 are used for the comparative analysis. In order to balance the results the average monthly churn capture will be used as the Hu (2005) methodology and the Hwang et al. (2004) methodology are based on 1 month only. The average results for the customer profiling methodology are 156/2 as the results are based on two future months of data. This gives an average monthly churn capture of 78 the remaining 78 customers from the second month is placed on the misclassification rate bringing the misclassification to 189. This gives a total churn capture of 78 customers with a total misclassification rate of 189. The hit ration for the customer profiling methodology is 78 78/189= 0.41.

6.5.10. Comparison of Methodologies

For easy analysis of the results obtained from all three methodologies have been compiled into the bar chart shown in Figure 91:

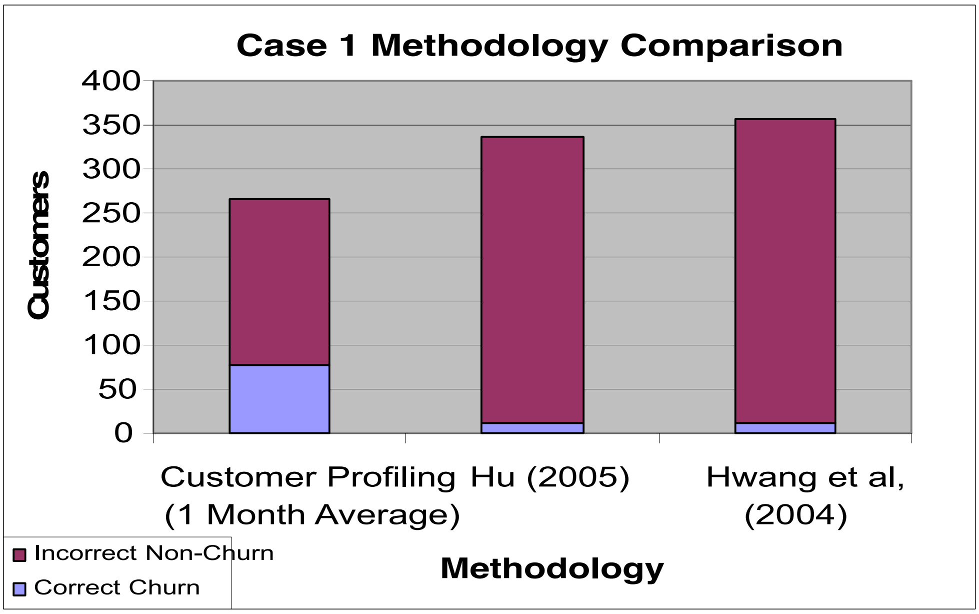


Figure 91: Comparison of Methodologies for Case 1

It is evident from the results for each methodology shown in Figure 91 that the customer profiling methodology has outperformed the one proposed by hu (2005) and the one proposed by Hwang et al. (2004). Both of these alternative methodologies have similar results for both churn prediction and corresponding non-churn misclassifications. The customer profiling methodology has captured significantly more churners. The misclassifications for the customer profiling methodology are much lower than the other two methodologies and a large proportion of total predictions did actually churn. It should be noted that the average monthly churn has been used for the customer profiling methodology and so has the average monthly non-churn misclassification.

The total churn capture for each methodology does not define the power of that methodology as prediction accuracy is directly related to the quality of the input data. The power of the model is defined by the hit ratio (the total number of predictions divided by the total number of correct classifications). The hit ratio provides an accurate measure of how well the model has determined predictions. The data used for analysis is not of very good quality which is apparent from the actual churn capture. The fact that the data is of poor quality is reflected through the Hu (2005) and Hwang et al. (2004) models. These models use general classification techniques and the results would show much stronger if the data was of better quality. The data does however; provide a suitable measure for model performance as the prediction accuracy measured by hit ratio does not require perfect data for churn prediction. Regardless of data quality, the profiling methodology has achieved a hit ration of 2.4 using monthly averages, which is over ten times better than either the Hu (2005) results or Hwang et al. (2004) results. The validation experiments performed on case 1 have determined that the customer profiling methodology has considerably outperformed the other techniques.

6.6. Case 2

This case study comprises of data from the residential broadband service sector. Churn in this area is currently a serious problem for service providers and is a major of focus of the sponsoring company's in house research.

Customer repairs and complaints data has again made up the dataset that has been used for case 2. This data comprises of 25 variables split between complaints, repairs and provisions (promises to supply service or equipment by a given date). The data contains 10 months of data between the months of January 04 to October 04. All churn for this dataset occurs in the month of October, so months April through to August are used for generation of loyalty index values with the intention of predicting churn 2 months in the future in October. Due to confidentiality reasons, it is not possible to give a full list of the variables that have been used to construct case 2; however a generalised summary with a brief description is shown in Table 19:

Table 16: An Example of Case 2 Data

Variable Name	Variable Description
LIFETIME	How long a customer has been with the company
PROMISE DURATION	Provisionally the number of days promised to the customer until resolution
REPAIR DURATION	Number of days the repair to be resolved
MU COUNT	No. of internal queues the job has come through
COMPLAINT DURATION	Duration of the complaint in days
COMPLAINT RESOLUTION OVERDUE	No. of days the promise resolution date ran over
COMPLAINT CODE	Code of complaint type
REFUND AMOUNT	Money refunded to the customer for loss of service
PROMISE COUNT	No. of provisions made
REPAIR COUNT	No. of repairs made
REPAIR OVERDUE	No. of days the promise resolution date ran over
NUMBER OF COMPLAINTS	No. of complaints made

6.6.1. Preparing the data

As discussed in case 1, the data for case 2 is analysed for incompatible or missing values. All data for case 2 is in numerical format so no conversion is necessary, however once again missing values are identified. These missing values are replaced with the value '0' as the neural network training will fail if 'NaN' values are encountered.

The data is further split into individual columns by customer, by month providing monthly time windows from which the churn index values can be generated.

6.6.2. Most suitable for Creating the Neural Network Model

An alternative approach for identifying the most suitable for NN training has had to be applied to case 2 because churn only occurrs during the month of October. For this reason 100 churners were randomly selected from the data and 400 non churners randomly selected creating a training dataset that consists of 20% churners and 80% non-churners. Also there is insufficient data to test future churn capture as there is only 1 month of churn data. For this reasons, forecasts of future predictions will be offered based on the trends apparent from the case 1 experiment.

6.6.3. NN Experiment 1

The first NN experiment was performed by constructing a NN and training it using the data described above. Once again the accuracy of the NN was established by applying it to the full month of October and constructing a confusion matrix to see how well it had identified churn. The results are displayed in Figure 92:

```
       | Estimated Labels
 True  |
Labels |    0       1   | Totals
-------|----------------|-------
0      |17237    150   |17387
1      |  559    507   | 1066
-------|----------------|-------
Totals |17796    657   |18453
```

Figure 92: NN Experiment 1 Results For Case 2

It is clear from the confusion matrix displayed in Figure 92 that the network from experiment 1 has converged well with a capture of nearly 50% of churners and a low misclassification rate.

6.6.4. NN Experiment 2

NN experiment 2 was once again constructed and trained using the set that has been created from case 2. This NN was set to predict the full churn for October to see if the performance would improve over that of experiment 1. The results are displayed by the confusion matrix in Figure 93:

```
        | Estimated Labels
 True   |
Labels  |    0       1   | Totals
-------|--------------|-------
0       |17355      32   |17387
1       |  564     502   | 1066
-------|--------------|-------
Totals  |17919     534   |18453
```

Figure 93: NN Experiment 2 Results for Case 2

The second NN architecture has outperformed the first. Despite a slightly lower churn capture the misclassification rate is significantly lower. This performance coincides with those observed in case 1.

6.6.5. NN Experiment 3

Finally NN experiment 3 was performed by creating a third NN based on the training dataset for case 2. This NN was then applied to the full month October to compare results with NN experiment 1 and NN experiment 2. The results are displayed in Figure 94:

```
        | Estimated Labels
 True   |
Labels  |    0       1   | Totals
-------|--------------|-------
0       |17370      17   |17387
1       |  602     464   | 1066
-------|--------------|-------
Totals  |17972     481   |18453
```

Figure 94: NN Experiment 3 Results for Case 2

As shown in Figure 94, the results generated from NN 3 are good. The churn capture is lower that of experiments 1 and 2, however the misclassification rate is very low. A comparison of all three NN experiments can be seen in Table 17:

Table 17: NN Experiment Comparison

	Churn Capture Accuracy	Non-Churn Capture Accuracy	Total Accuracy
Case 2			
NN1 Experiment	47.60%	93.40%	96.10%
NN2 Experiment	47.10%	94.00%	96.70%
NN3 Experiment	43.50%	94.10%	96.60%

The NN experiment for case 2 shown in Table 17 conforms with the experiments performed for case 1, with NN experiment 2 displaying the highest level of overall accuracy. NN configuration 2 will be used for generating churn index values for case 2.

6.6.6. Generating Churn Index Values

Comparing the results generated from NN experiments 1, 2 and 3 it can be determined that NN experiment 2 has provided the best performance. The actual churn capture is slightly lower that that of experiment; however the misclassification rate is significantly lower. Even thought the misclassification rate for NN experiment 3 is lower than that of NN experiment 2 the churn capture is also significantly lower. Therefore it has been determined that NN architecture 2 will be used for the generation of customer churn index values for case 2.

Each month contained within the case 2 data is analysed by NN architecture 2 for generation of customer churn index values. Once these index values are generated they are exported into excel with customer number and churn data, saved as a text file and imported into MySQL. With this dataset all churn occurs in October. Therefore the churn month for this dataset is set to 10 for all churners.

6.6.7. Apply customer profiling methodology

Due to limitations of the amount of data available for analysis with case 2, the entire dataset is split in half. Each half contains 50% of the churn and non-churn from the full dataset. This is a method of cross-fold validation as discussed in section 2.7. This creates two separate datasets. The training dataset contains 9227 customers and the validation dataset contains 9226 customers. Each dataset contains 533 churners.

The months January through to July are set as the inputs into the customer profiling methodology to allow for prediction of churn in October, providing a 3 month future prediction. The initial profiles generated by the customer profiling methodology are shown in Table 18:

Table 18: Determined Profiles for Case 2

Profile Number	Profile Characteristics	Cluster Size
1	UDUDD	16
2	DUD	823
3	UUDUUD	64
4	DUUD	823
5	UUUUD	30
6	UUDUD	9
7	DD	822
8	D	825
9	UUUD	9
10	DUUUD	822
11	UUD	9
12	UDD	5
13	DDUUD	822
14	UUUUUD	172
15	UD	5
16	DUDUDD	822
17	UUUDD	9
18	UDUUD	5

It can be seen from Table 18 that the profiling methodology has been applied effectively to case 2. A number of profiles were identified from this case study and a significant number of customers assigned to each one.

For case 2 a total of 6092 customers out of a total number of 9227 have been clustered into profile categories. The creation of high risk and low risk clusters should allow us to determine the churners and drastically reduce the number of misclassifications.

6.6.8. Identify the high and low risk profile clusters

Applying the results shown in Table 18 directly as churn predictions would have a high misclassification rate for non-churners. It has already been shown from the case 1 experiments how the customer profiles can be used to drastically improve these results by determining high and low risk clusters targeted at future churn capture. Each master profile cluster is assessed individually to determine which are the most effective for capturing future churn. Table 19 displays the churn results for case 2 based on future churn accuracy inclusion for each cluster:

Table 19: Individual Profile Accuracy for Case 2

ID	Profile	Cluster Size	Members Who Actually Churned	Future Churn Accuracy	Future Churn Capture
1	UDUDD	16	1/16	6.25%	Low
2	DUD	823	3/823	0.36%	Low
3	UUDUUD	64	8/64	12.50%	Low
4	DUUD	823	1/823	0.00%	Low
5	UUUUD	30	2/30	6.67%	Low
6	UUDUD	9	2/9	22.22%	High
7	DD	822	8/822	0.97%	Low
8	D	825	347/825	42.06%	High
9	UUUD	9	1/9	11.11%	Low
10	DUUUD	822	1/822	0.00%	Low
11	UUD	9	1/9	11.11%	Low
12	UDD	5	1/5	20.00%	Low
13	DDUUD	822	1/822	0.00%	Low
14	UUUUUD	172	7/172	4.07%	Low
15	UD	5	1/5	20.00%	Low
16	DUDUDD	822	1/822	0.00%	Low
17	UUUDD	9	1/9	11.11%	Low
18	UDUUD	5	1/5	20.00%	Low

As shown by Table 19, case 2 reinforces the methodology displaying that not all profile clusters are good for identifying churners. In fact only 2 clusters for case 2 have displayed suitability for future churn capture, while the remaining 16 clusters have had little significance on the overall churn results meaning the majority should be classified as low risk clusters. Although the customers belonging to the weaker profiles presently show a small probability of churning, this does not mean that these profiles will continuously remain low risk. In fact it is anticipated that these lower risk profiles could easily shift to high risk in response to environmental changes. It should be noted that in the past customers with similar profiles have actually churned. All customers belonging to profile clusters have been detected to have suffered satisfaction damaging experiences. Therefore all customers classified into profile clusters do have the potential to churn. Should competing service providers suddenly release attractive service deals, it is believed that the customers belonging to the lower risk profile groups will be the first ones to sway to these offers and inevitably churn, causing the lower risk profiles to suddenly shift to high risk categories.

Referring to Table 19 it can be identify that 42% of the customers belonging to profile cluster group 8 did actually churn and 22% of the customers belonging to profile cluster group 6 did actually churn. These groups are therefore considered as churn sensitive. All other groups contain an actual churn capture of between 0 and 20% churn. Therefore the customers belonging to these groups are classified as low risk churners. The reason up to 20% is regarded as low risk is because it has been identified from expert opinion that the industry average hit ratio is about 0.25. This converts to 1 in 4 customers predicted is an actual churner. 20% capture or below converts to 1 in 5 correct classification which is regarded as significantly lower than the industry average. Above 20% is around the industry average and can be included as high risk.

Filtering out the low risk churn categories it can be determined that 349 of a total of 533 churners have been caught by the customer profiling methodology while 485 customers have been misclassified as churners. From a total of 9227 customers this misclassification rate is well within acceptable limits as we can determine a total 94% Non-Churn capture. The capture accuracy of 65% total churn 3 months after classification is a high percentage, so application of the profiling methodology to case 2 has been successful. The filtered results for case 2 can be seen from the confusion matrix in Figure 95:

```
       | Estimated Labels
 True  |
Labels |    0       1  | Totals
-------|---------------|-------
0      | 8209    485   | 8694
1      |  184    349   |  533
-------|---------------|-------
Totals | 8393    834   | 9227
```

Figure 95: Filtered Results for Case 2

The results are clearly good from the confusion matrix in Figure 95. A churn capture accuracy of 65% is identified while a non-churn accuracy of 94% is sustained. The hit ratio is 0.41, which is significantly higher than the industry average; as identified from expert opinion (please see appendix A). Therefore the methodology has been applied effectively to case 2.

6.6.9. Determining Future Churn Capture

This case study contains 10 months of data. The first 7 months provide the input for the profiling methodology while the subsequent 3 months are used for classification of future churn accuracy for each high predictor master profile cluster. The subsequent 3 month period contains a total of 533 churn points. Referring back to Table 19, it can be seen that master profiles 8 and 6 can be categorised as strong churn classifiers. The future churn captured from these two profiles total 349 churners out of the 533 who actually churn 3 months past the input data. This is 65% accurate.

The customer profiling methodology is applied to the second dataset, using just the two master profile templates that are identified as suitable future churn predictors, identified from the model creation stage. The Future churn prediction results are displayed in Table 20:

Table 20: Full Future Prediction Results

ID	Profile	Cluster Size	Local churn	Future Churn	Future Churn Accuracy	Future Churn Capture
6	UUDUD	12	0	0	0.00%	High
8	D	704	0	368	52.27%	High

As can be seen from Table 20, the future churn capture remains almost identical to the initial model creation results. Master profile cluster 8 has classified 704 customers as churners and 368 of those customers actually churned 3 months in the future. Master profile cluster 6 was also identified as a possible future predictor from the model creation stage; however no future churn was caught from this cluster when applied to the validation dataset. From the model creation stage only 9 customers where assigned this profile cluster and only 2 actually churned. 12 customers were assigned this profile template from the validation dataset but none actually churned. With such a small classification at both model creation and prediction stages this profile can be considered to be stable. Master profile cluster 6 was recognised as a suitable future predictor at model creation so those customers belonging to that cluster are classified as churners. The future classification results from master profiles 8 and 6 are displayed in Figure 96:

```
        | Estimated Labels
 True   |
Labels  |    0       1    | Totals
--------|-----------------|-------
0       | 8345     348    | 8693
1       |  165     368    |  533
--------|-----------------|-------
Totals  | 8510     716    | 9226
```

Figure 96: Case 2 Future Prediction Results

In total 533 customers churned in 3 months future of the input data. From the strong churn predictor profiles 368 of the churners for October were captured.

It can be identified from the confusion matrix in Figure 96 that 69% of future churn was captured by the high risk profile clusters and 96% of non-churn was sustained. The hit ratio for the future prediction results is 0.51. A comparison between the results obtained at the model creation stage and the results from the future prediction stage can be seen from Table 21:

Table 21: Result Comparison

	Model Creation Stage	Future Churn Capture
Case 2 Result Comparison		
Non-churn Accuracy	94.00%	96.00%
Churn Accuracy	65.00%	69.00%
Hit Ratio	0.41	0.51

It can be identified from Table 21 that the results have remained reasonably constant when applying the clusters identified at the model creation stage to determine future churn. The main observation is that the results actually improve slightly when applied to the validation dataset, and similarly to case 1, the hit ratio is over double that of current industry standards as identified from expert opinion (please see appendix A).

6.6.10. Comparing Results with Two Alternative Predictive Methods

The research of Hu, (2005) and Hwang et al. (2004) has still been applied to the case 2 data to establish the churn capture that would be achieved from other methodologies.

6.6.10.1. Research of Hu (2005)

In order to recreate the experiments by Hu (2005) the author had to first select just one month of loyalty index data. For the purpose of the experiment August data from case 2 was selected to attempt prediction of churn in October, allowing for a 2 month future churn capture.

The data was originally stored in the database in order of customer number. The data was rearranged and ordered by loyalty index value in ascending order. The top the top 4% of customers with the lowest loyalty index values are categorised as churners, corresponding to the methodology proposed by Hu (2005). This data was used to create a new table for the purpose of the experiment. A further field was added to the table which contained the value '1' for every customer in the top 4% with lowest loyalty index values. This is because all these customers are theoretically churners. The resulting dataset contains records for all 9226 customers from the case 2 validation dataset and predictions are compared with the churn records for October. Predictions are made from July loyalty results to establish a comparison with the predictions 3 months in the future as obtained from the customer profiling methodology. The confusion matrix in Figure 97 demonstrates how well the experiment predicted churn for October:

```
       | Estimated Labels
 True  |
Labels |    0       1  | Totals
-------|---------------|-------
0      | 8330    363   | 8693
1      |  527      6   |  533
-------|---------------|-------
Totals | 8857    369   | 9226
```

Figure 97: HU (2005) Churn Prediction

As can be seen from Figure 97, 6 of the customers who churned in October were correctly identified, from a total of 533 when attempting to match the time scale of 3 months as achieved by the customer profiling methodology. 363 non-churn customers were misclassified as churners and 527 customers who actual churned where misclassified as non churners. The churn capture accuracy for the Hu (2005) methodology is 1%. The hit ratio achieved by the Hu (2005) methodology is 0.01.

6.6.10.2. Research of Hwang et al. (2004)

In order to recreate the experiments of Hwang et al. (2004) the customer loyalty values have to be segmented into two groups. Those customers with a loyalty value of over 0.5 and those customers with a lower loyalty value.

All customers that have been segmented into the group with a loyalty index value lower than 0.5 are classified as churners. All customers belonging to the group with a loyalty index value of higher than 0.5 are classified as non-churners. These results can be viewed in Figure 98:-

```
       | Estimated Labels
 True  |
Labels |    0      1   | Totals
-------|---------------|-------
0      | 5357   3336   | 8693
1      |  380    153   |  533
-------|---------------|-------
Totals | 5737   3489   | 9226
```

Figure 98: Hwang et al. (2004) Churn Prediction Accuracy

As can be seen from Figure 98, the Hwang et al. (2004) methodology has identified 153 out 533 churners providing 29% accuracy. The hit ratio achieved from the Hwang et al. (2004) methodology however is 0.4.

When comparing the two benchmark experiments it can be observed that the actual churn capture from Hu (2005) methodology is not as high as that obtained from the Hwang et al. (2004) methodology. The misclassifications rate from Hwang et al. (2004) appears significantly higher than the misclassification rate from Hu (2005) methodology; however on inspection of the hit ratio of both methodologies it can be noticed that Hwang et al. (2004) achieved a hit ratio of 0.4 while Hu (2005) achieved a hit ratio of 0.01. This means that Hwang et al. (2004) outperformed Hu (2005) for both churn capture and hit ratio.

6.6.10.3. Comparison of Methodologies

Hwang et al. (2004) methodology has been regarded as better than the Hu (2005) methodology because the hit ratio and churn capture for Hwang et al. (2004) is a better than both of those obtained from the Hu (2005). The customer profiling methodology has outperformed both Hu (2005) and Hwang et al. (2004) with churn capture of 368 customers and hit ratio of 0.51. Figure 99 provides the results for all three methodologies

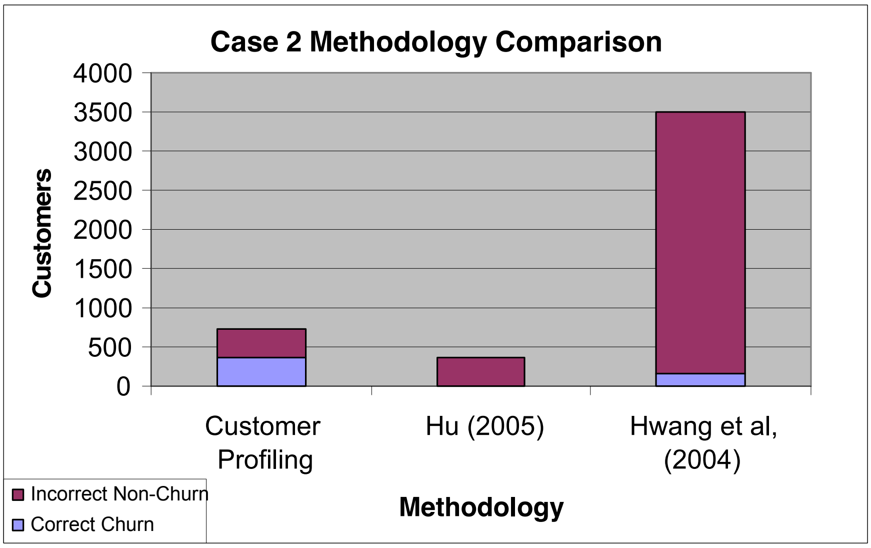


Figure 99: Case 2 Methodology Comparison

As can be seen by the chart in Figure 99, the customer profiling methodology has performed significantly better than the other two methodologies for both churn capture and misclassifications when predicting churn 3 months in advance.

6.7. Case 3

As stated in Table 9, case study 3 is a relatively small dataset containing information regarding business user's landline telephone accounts. The data contains only four variables per customer. These variables however are powerful in that they represent counts of customer repairs, complaints and promises that have been made to the customer by the service provider. This data is in the form of variable counts. There is no detailed information regarding complaints, repairs or provisions, only the count values.

6.7.1. Preparing the Data

As performed on the previous case studies, the data for case 3 is evaluated for incompatible formats and missing or 'NaN' values. Because this data is all counts of complaints, repairs and provisions all data is numerical in format, however missing values were identified and replaced with the value '0'. The data contains 9 months of customer data spanning from January 04 to September 04.

6.7.2. Most suitable for Creating the Neural Network Model

Because there is only 9 months of data for this dataset months January to July will be used as the inputs into the profiling methodology to allow the final 2 months to be used for profile accuracy grading. Therefore months August and September will be excluded from the analysis for determining the most suitable month for creating the NN model.

On analysis of the data it is determined that the month of May should be used for NN construction because although each month has a similar number of churners, May contains slightly more than others. 124 non churners are randomly selected from May and all 31 churners are included to create a training dataset of 155. This corresponds to the 20:80 Churn/Non-churn ration used with the previous two studies. The NN experiments are created from this training dataset.

6.7.3. NN Experiment 1

The first NN experiment was performed by constructing a NN and training it using the training data described above. The accuracy of the NN was determined by applying it to the full month of May and constructing a confusion matrix to see how well it had identified churn. The results are displayed in Figure 100:

```
        | Estimated Labels
 True   |
Labels  |    0      1   | Totals
--------|---------------|-------
0       | 7159    168   | 7327
1       |   16     15   |   31
--------|---------------|-------
Totals  | 7175    183   | 7358
```

Figure 100: NN Experiment 1 Results For Case 3

Case 3 is realistic in terms of churn/non-churn ratio when compared to actual business figures; however the limited churn figures make it difficult to obtain good predictions. Nearly 50% of churners for May have been captured by the NN model.

6.7.4. NN Experiment 2

A NN was once again constructed and trained using the set that has been created from case 3. This NN was set to predict the full churn for May to see if the performance would improve over that of experiment 1. The results are displayed in the confusion matrix in Figure 101:

```
        | Estimated Labels
 True   |
Labels  |    0       1  | Totals
--------|---------------|-------
0       | 7160     167  | 7327
1       |    7      24  |   31
--------|---------------|-------
Totals  | 7167     191  | 7358
```

Figure 101: NN Experiment 2 Results for Case 3

It can be seen from the confusion matrix in Figure 101 that churn capture has improved significantly over that obtained from experiment 1 suggesting that the increased complexity of the NN model has had a positive effect on the NN performance. The percentage of misclassifications has remained unchanged.

6.7.5. NN Experiment 3

The third NN experiment is created to determine if even greater NN complexity would further churn capture and reduce the misclassification rate. The results from NN experiment 3 are displayed in Figure 102:

```
        | Estimated Labels
 True   |
Labels  |    0       1   | Totals
-------|--------------|-------
0       | 7160    167   | 7327
1       |   10     21   |   31
-------|--------------|-------
Totals  | 7170    188   | 7358
```

Figure 102: NN Experiment 3 Results For Case 3

As shown from the confusion matrix in Figure 102, greater complexity of the NN model has had little effect on the prediction results for case 3. The misclassifications remain the same; however there has been a slight decrease in the number of correctly identified churners.

6.7.6. Generating Churn Index Values

Comparing the results generated from NN experiments 1, 2 and 3 it can be determined that NN experiment 2 has again provided the best performance. This can be seen clearly from the comparison of all results presented in Table 22:

Table 22: Comparison of NN Results for Case 3

Case 2	Churn Capture Accuracy	Non-Churn Capture Accuracy	Total Accuracy
NN1 Experiment	48.40%	97.70%	97.90%
NN2 Experiment	77.40%	97.70%	98.50%
NN3 Experiment	67.70%	97.70%	98.00%

The actual churn capture accuracy of NN experiment 2 is highest out of the three experiments, while the misclassification rate is on par with NN experiment 3. Also the total accuracy is highest for the NN 2 experiment. Therefore it has been determined that NN architecture 2 will be used for generation of the customer churn index values for case 3.

Each month contained within the case 3 data is analysed by NN architecture 2 for generation of customer churn index values. Once these index values are generated they are exported into excel with customer number and churn data, saved as a text file and imported into MySQL.

6.7.7. Apply Customer Profiling Methodology

Months January to July will be used as inputs into the customer profiling methodology to predict churn in months August and September. Due to a shortage of data with this case study a cross-fold validation method is applied, similar to the method applied to case 2. Therefore the dataset is split 50:50 to create a dataset of 3679 for building the model and a separate dataset of 3679 for validating the model. The churn for each month was also split 50:50 so each separate dataset could receive exactly half the churn on a monthly basis each. The total churn for the training dataset totalled 218 with 31 customers churning in the two month to be used for future classification, and the total churn for the validation dataset totalled 217 with 30 customers churning in the two months to be used to determine future churn. The profiles shown in Table 26 were detected at the model creation stage:

Table 23: Generated Profiles for Case 3

Profile Number	Profile Characteristics	Cluster Size
Profile 1:	Profile DDD	Members: 16
Profile 2:	Profile DLUDD	Members: 14
Profile 3:	Profile DLDD	Members: 23
Profile 4:	Profile D	Members: 30
Profile 5:	Profile DUD	Members: 21
Profile 6:	Profile DLD	Members: 19
Profile 7:	Profile DLDUD	Members: 34
Profile 8:	Profile DD	Members: 91
Profile 9:	Profile DLLLLD	Members: 19
Profile 10:	Profile DLLD	Members: 18
Profile 11:	Profile DUUDULD	Members: 20
Profile 12:	Profile DLLLD	Members: 21
Profile 13:	Profile DUDUD	Members: 19
Profile 14:	Profile DDULDUD	Members: 24
Profile 15:	Profile DDULLD	Members: 21
Profile 16:	Profile DUDD	Members: 25

It can be seen from Table 23 that the profiling methodology has been applied effectively to case 3. For case 3 a total of 415 customers out of a total number of 3679 have been clustered into profile categories. The results for classifying all customers in all profiles as churners to predict the next two months of churn are displayed in Figure 103:

True Labels	Estimated Labels 0	1	Totals
0	3257	391	3648
1	7	24	31
Totals	3264	415	3679

Figure 103: All Profiles Used to Predict Future Churn for Case 3

It can be seen from Figure 103 that 77% of the future churn has been captured for case 3 by classifying all profiles as churn sensitive; however for such a small number of churn the misclassification rate is high, resulting in a hit ratio of 0.05. Clustering the profiles into high and low risk churn categories should provide a solution to this problem as seen for cases 1 and 2.

6.7.8. Identify the High and Low Risk Clusters

As seen from the previous two case studies, the detection of high and low risk churn sensitive groups significantly reduce the total number of misclassifications while maintaining most of the correctly identified churn classifications. A total of 61 future churners defected in the final 2 months future to the input data. The results can be seen from Table 24:

Table 24: High and Low Risk Categories for Case 3

Profile Number	Profile Characteristics	Cluster Size	Local Churn	Future Churn	Total Accuracy	Future Churn Accuracy	Churn Capture
4	D	30	9	14	76.67%	46.67%	High
1	DDD	16	7	4	68.75%	25.00%	High
6	DLD	19	8	4	63.16%	21.05%	High
10	DLLD	18	8	1	50.00%	5.56%	Low
8	DD	91	59	1	65.93%	1.10%	Low
16	DUDD	25	13	0	52.00%	0.00%	Low
13	DUDUD	19	12	0	63.16%	0.00%	Low
11	DUUDULD	20	16	0	80.00%	0.00%	Low
2	DLUDD	14	9	0	64.29%	0.00%	Low
3	DLDD	23	17	0	73.91%	0.00%	Low
5	DUD	21	13	0	61.90%	0.00%	Low
7	DLDUD	34	16	0	47.06%	0.00%	Low
9	DLLLLD	19	8	0	42.11%	0.00%	Low
12	DLLLD	21	6	0	28.57%	0.00%	Low
15	DDULLD	21	4	0	19.05%	0.00%	Low
14	DDULDUD	24	7	0	29.17%	0.00%	Low

As can be seen by Table 24, three master profile clusters show strongest accuracy for future churn capture. Master profile cluster 4 shows one major event that leads to customer churn, Master profile 1 has three consecutive events, each impacting customer loyalty more than its predecessor until the third one, which has enough impact to lead to churn. The final master profile cluster which is most suitable for future churn capture is master profile 6. Master profile cluster 6 displays an initial event which does not have enough impact to cause churn, a second event on par with the first event in terms of loyalty value impact, and a third event that does impact loyalty enough to cause churn. The confusion matrix in Figure 104 shows the results for case 3 using only the high risk profile clusters to classify customer churn:

```
        | Estimated Labels
 True   |
Labels  |    0       1   | Totals
--------|----------------|-------
0       | 3605      43   | 3648
1       |    9      22   |   31
--------|----------------|-------
Totals  | 3614      65   | 3679
```

Figure 104: Capture Accuracy Using Only High Risk Clusters as Churn Predictors

It is clear from Figure 104 that the high risk future churn master profile clusters are 4, 1 and 6. Once future churn predictions are limited to these clusters the results displayed in Figure 104 are obtained. The results show that the actual churn capture has not been greatly impacted however the misclassification rate has been significantly reduced to 43. The hit ration for case 3 has become 22/65 which is equal to 0.33. This in itself is higher than the average hit ratio as discovered from the industrial survey (Please see appendix a). The final stage is to remove all customers from these high risk profile clusters who actually churned within the analysis period. These customers are shown in Figure 105. Once these customers are removed the results displayed by Figure 111 are achieved:

True Labels	Estimated Labels 0	1	Totals
0	3629	19	3648
1	9	22	31
Totals	3638	41	3679

Figure 105: Case 3 Profile Classification Excluding Local Churn

For case 3, churn data is available regarding the customers who churn during the analysis period. Once these customers are excluded from the future prediction results the results obtained are those displayed by Figure 105. The churn capture accuracy is at 71%, non-churn accuracy is 99% and the hit ratio has been enhanced to 22/41 = 0.53, which is similar to the accuracy achieved from cases 1 and 2.

6.7.9. Determine Future Churn Capture Accuracy

With the model created it can be applied to the validation dataset as to simulate future predictions. The high risk profile clusters as identified from the model creation stage classified the numbers of predicted churn shown in Table 25:

Table 25: Raw High Risk Churn Capture

Profile Number	Profile Characteristics	Cluster Size	Local Churn	Future Churn
4	D	45	12	18
1	DDD	21	9	6
6	DLD	16	6	2

The results displayed in Table 25 have been generated by applying the customer profiling methodology to the full validation dataset and using only the high risk profile clusters identified from the model creation stage as future churn predictors. Master profile 4 has had 45 customers classified to it; master profile 1 has had 21 customers classified to it and master profile 6 has had 16 customers classified to it. Therefore, initially 82 customers have been classified as churners. These results can be seen from the confusion matrix in Figure 106:

```
        | Estimated Labels
 True   |
Labels  |    0       1   | Totals
--------|----------------|-------
0       | 3593      56   | 3649
1       |    4      26   |   30
--------|----------------|-------
Totals  | 3597      82   | 3679
```

Figure 106: Case 3 Initial Future Churn Results

As can be seen from Figure 106, the initial future churn results have caught 86% of all future churn. Due to the initial misclassification numbers, the hit ration from this stage of the model is equal to 0.31. This is close to the hit ratio obtained from these three clusters in the model creation stage. The customers who actually churned during the analysis period can be removed to determine the final future prediction results and model accuracy. 27 customers actually churned during the analysis period, so with these customers excluded from the results the final accuracy is displayed in Figure 107:

```
        | Estimated Labels
 True   |
Labels  |    0       1   | Totals
--------|----------------|-------
0       | 3620      29   | 3649
1       |    4      26   |   30
--------|----------------|-------
Totals  | 3624      55   | 3679
```

Figure 107: Case 3 Final Future Prediction Accuracy

It can be seen from the confusion matrix in Figure 107 that once the local churn has been removed from the misclassification numbers the hit ratio becomes 26/55 = 0.47. The non-churn prediction accuracy equals 98% and the churn capture accuracy equals 87%. This shows that for case3 the identified high risk master profile clusters identified from the model creation stage have retained their accuracy when applied to the task of capturing future churn. A comparison of the results between model creation and future churn capture can be seen from Table 26:

Table 26: Case 3 Result Stability Comparison

Case 3 Result Comparison	Model Creation Stage	Future Churn Capture
Non-churn Accuracy	99%	98%
Churn Accuracy	71%	87%
Hit Ratio	0.53	0.47

As shown from Table 26 the results obtained at the model creation stage have remained stable when applying the model to the validation dataset. Non-churn accuracy has fallen by 1%, however churn accuracy has increased. The fact that there is such a small amount of churn and high number of non-churn has resulted in an overall decrease in hit ratio. The hit ratio is 0.47 which is still almost double current industrial standards as identified through the industrial survey (please see appendix A), determining a strong predictive model.

6.7.10. Comparing Results With Two Alternative Predictive Methodologies

The results are obtained from Hu (2005) and Hwang et al. (2004) to illustrate the prediction accuracy of other methodologies.

6.7.10.1. Research of Hu (2005)

To recreate the research by Hu (2005) the month of August was selected for analysis as this will enable a determination of how well Hu (2005) has performed at capturing churn a month in the future.

The loyalty index data for August was reordered in ascending order. The top 4% of customers with the lowest churn index values were then extracted and classified as churners. The remaining 6974 customers were classified as non churners. The churn data was then compared with the churn flag for September for identification of future churn capture. The results have been displayed in the confusion matrix shown in Figure 108:

```
       | Estimated Labels
 True  |
Labels |    0       1   | Totals
-------|----------------|-------
0      | 6952    377    | 7329
1      |   22      7    |   29
-------|----------------|-------
Totals | 6974    384    | 7358
```

Figure 108: Hu (2005) Churn Capture

As can be seen from the results obtained from the Hu, (2005) methodology shown in Figure 108, the results are poor. Only 7 out of a total of 29 churners have been captured from the methodology while the misclassification rate is very high with 377 non-churners wrongly classified as churners. 24% of churn is captured by this methodology. The hit ratio achieved by the Hu (2005) methodology is 7/384 = 0.01

6.7.10.2. Research of Hwang et al. (2004)

The methodology proposed by Hwang et al. (2004) involves classifying every customer with a churn index value of lower than 0.5 as churner. This has been performed on the case 3 data and the results are displayed in Figure 109:

True Labels	Estimated Labels 0	1	Totals
0	7237	92	7329
1	23	6	29
Totals	7260	98	7358

Figure 109: Hwang et al. (2004) Methodology for Case 3

As shown from the confusion matrix in Figure 109, Hwang et al. (2004) methodology does not produce good results for Case 3. Churn capture is lower than that caught by Hu (2005) methodology; however there has been a significant decrease in misclassifications. The hit ratio achieved by the Hwang et al. (2004) methodology is 6/98 = 0.06.

6.7.10.3. Comparison of Churn Prediction Methodologies

The Hwang et al. (2004) methodology has provided better results than Hu (2005) methodology with a higher churn captured while misclassification is much lower. The Hu (2005) methodology shows a small churn capture similar to that of the Hwang et al. (2004) methodology; however the misclassification rate is large in comparison. The Hwang et al. (2004) methodology has achieved a hit ratio of 0.06 while the hu (2005) methodology has achieved a hit ratio of 0.01, again showing that Hwang et al. (2004) has been the strongest methodology. Both methodologies have failed in comparison with the customer churn methodology. A comparison of the results obtained from all three methodologies is presented in Figure 110:

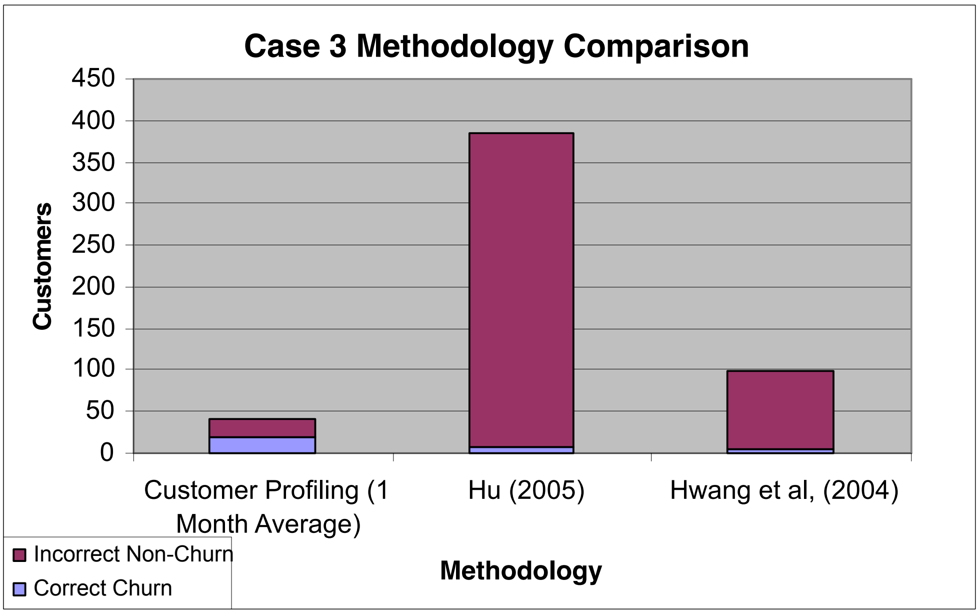


Figure 110: Case 3 Methodology Comparison

It is evident from the methodology comparison presented in Figure 110 that the customer profiling methodology has again outperformed the other two methodologies. Churn capture is of greater accuracy while the misclassifications have been controlled and limited.

6.8. Summary

The aim of this chapter has been to validate the research and resulting customer profiling methodology by use of three separate case studies which cover three separate areas of the telecommunications industry. The methodology has also been compared against two other methodologies from literature that aim to predict customer churn.

Case 1 provided enough data to test a 'moving window' concept. The results were used to identify trends between customers. From this case study it could be detected that a similar number of the churners caught when determining the best profiles were also caught when detecting future churn. The initial model development stage for case 1 established churn model that provided a hit ratio of Limitations with data prevent an analysis of future churn capture for cases 2 and 3; however it can be assumed that a similar to pattern to that observed from case 1 should be anticipated. Based on the results it can be assumed that the profiling methodology could provide a powerful tool for churn prediction.

Based on the experiments in this section, it is determined that the customer profiling methodology is an accurate method for capturing future churn while keeping the non-churn misclassification at a low level. The ‘moving window’ concept is very difficult to fully test without the methodology being successfully implemented and monitored in a real working environment, however from the experiments performed for case 1 this concept shows the potential to radically boost the customer profiling results further contributing to an extremely powerful methodology. This ‘moving window’ concept and the fact that the customer profiling methodology uses a time sequence of data also makes it possible for the real-time monitoring of customer loyalty.

6.9. Expert Opinion

This section applies a qualitative validation approach to the proposed methodology. A questionnaire/interview technique has been used to capture professional knowledge for determination of the performance of the proposed customer profiling methodology.

6.9.1. Methodology

Three customer relationship management professionals from different areas were identified and interviewed, providing feedback on developed customer profiling methodology from three separate perspectives. Respondent 1 works in the field of developing churn prediction strategies, respondent 2 works in a CRM based field and respondent 3 works in a data mining/data analysis field. The expertise of the respondents is shown in Table 27:

Table 27: Respondent Skill Levels

Expertise	Respondent 1	Respondent 2	Respondent 3
Churn Prediction	5	4	4
Programming	5	3	4
Data Mining	5	4	5
CRM Products	3	5	2
Customer Relationship Management	5	5	4
Interviewee graded on an experience scale of 1-5 (5 =Maximum, 1 = Minimum)			

The skill strengths shown in Table 27 have been self assigned by each of the respondents. Respondent 1 is responsible for experimenting with, and developing, churn prediction strategies. He/she has stated average knowledge of CRM products and maximum knowledge on customer management.

The strength of respondent 2 is clearly in CRM and customer management techniques where he/she has stated maximum knowledge. This is because respondent 2 works in a marketing/customer management role.

The strength of respondent 3 is clearly in data mining where he/she has indicated maximum knowledge. He/she has indicated poor knowledge in CRM technologies but good knowledge in customer management techniques. Respondent 3 is a senior data miner and data analyst and his/her skills assessment reflect such a role.

The customer profiling methodology was presented to each of the three experts to provide them with a thorough understanding of the developed methodology, the results that it achieves and how it could be implemented in industry. The presentation in each case complimented with a question and answer session to ensure that the respondent had gained complete understanding of the proposed methodology. Each interviewee was then required to complete a semi-structured questionnaire (please see Appendix A).

6.9.2. Questionnaire

A semi-structured questionnaire was developed to capture expert opinion regarding the proposed methodology and current practice in industry. The questionnaire comprised of the following four sections:

Industrial churn management practices – This section has been designed to capture the practices that are used and developed in industry.

Customer Profiling Methodology – This section has been designed to capture professional opinion regarding the proposed methodology.

Churn prediction results – This section has been designed to capture churn prediction achievements in industry, and how these achievements compare to those accomplished by the proposed methodology.

Respondents' background – This section has been included to capture the professional experience of the respondents, to achieve confidence in the captured information.

The questionnaire has been designed and structured to capture expert opinion about the proposed churn prediction methodology and the current accuracy that is being achieved in industry. This allows an analysis of results between current standards and the proposed churn prediction methodology. Information regarding the suitability of each of the respondents, for commenting both on the proposed methodology and current industrial standards has been captured to ensure a level of confidence in the captured information.

The proposed methodology was delivered to the respondents by a detailed presentation session. A comprehensive question and answer session followed the presentation to ensure that all respondents fully understood the proposed methodology. Each respondent was individually interviewed so that their opinion could be accurately recorded. The respondents were asked semi-structured questions from each of the four questionnaire sections and all responses recorded by the interviewer. The respondents were asked to check the recorded responses to ensure that the captured information accurately reflected the respondent's opinion.

6.9.3. Analysis

The results from the questionnaire were examined to determine an understanding of current churn management practices, and identification of how current techniques perform in the capture of customer churn. Conclusions have been made by identifying common practices and typical techniques. The responses have been carefully examined for any discrepancies between the respondents. The results for each section are discussed below.

6.9.4. Section 1 - Industrial Churn Management Practices

All respondents have stated that the companies they work for develop in-house solutions for customer churn capture. They identified neural networks, classification trees and regression analysis as predictive techniques used in their companies. One respondent has signified investigations using Markov models and one respondent indicated that other solutions had been investigated. This respondent was questioned on his/her investigation into 'other' techniques. It was identified that experiments had been made into support vector machines (SVM). This respondent stated that investigations into SVM were terminated as it consumed too much of the businesses computing resources and it took a long time to complete an analysis. It has been claimed by respondent 1 that he/she prefers to use classification trees whenever possible because once the classification rules have been extracted from the initial analysis they can be applied directly for data analysis using SQL, providing extremely fast generation of results. This respondent identified that NN may provide more powerful prediction than classification trees; however speed of analysis is regarded as a critical factor.

Next the use of 'off the shelf' CRM products was investigated. All respondents have used 'off the shelf' CRM software and have commonly identified Oracle CRM as the preferred product respondent 1 has stated that he/she has found 'off the shelf' products to be expensive while the churn prediction results obtained by these products are generally poorer than those that they achieve themselves through in-house development. Respondent 2 has claimed that from his experience, 'off-the-shelf' CRM products do not generally include or commit to churn capture and generally concentrate more on business strategy and churn prevention through aiding the identification of competitive marketing campaigns. Respondent 3 has stated that current churn capture methods, either 'off the shelf' or developed 'in-house', commonly achieve only a short churn window. A table of questions and answers for the multiple choice questions in section 1 is presented in Table 28:

Table 28: Section 1 Multiple Choice Questions and Answers

		Respondent 1	Respondent 2	Respondent 3
Technology Used				
	Neural Networks			
	Classification Trees			
	Regression Analysis			
	Evolutionary Computing			
	Markov Model			
	Other			

It can be seen from Table 28 that all respondents have used neural networks, classification trees and regression analysis for analysis. Evolutionary computing has not techniques have not been investigated by any of the respondents. Respondent 1 has used Markov models and other techniques for analysis; however respondents 2 and 3 have not performed analysis using these technologies. Through interviewing each of the respondents it has been understood that respondent 1 has investigated many technologies while attempting to develop the most accurate churn prediction model possible. Respondents 2 and 3 have not been required to perform such a comprehensive analysis on predictive technologies as it is not required for their areas of occupation. The most popular techniques suffice for these respondents.

6.9.5. Section 2 - Customer Profiling Methodology

From the understanding that each of the respondents have gained about the customer profiling methodology, all three have indicated that it is a sound and valid step in the evolution of customer churn management practices. Respondent 1 has stated that the customer profiling methodology analyses customer data in several ways. The presented results are unique and novel for customer churn analysis and provide good value to business. Respondents 2 and 3 have both stated that the methodology is novel, logical, and robust. Respondent 3 has stated that from a data mining perspective, the methodology is logical; however to his knowledge it has never been identified and explored before in industry.

Respondent 2 has indicated that the methodology is practical and strong due to its ability to generate results from multiple data sources. Respondent 3 has indicated the practicality of the customer profiling methodology as strong; but has stated that he/she feels it may be resource hungry. Respondent 1 has scored practicality as average, stating that if applied in industry, the customer profiling methodology would require a user friendly interface and the reports would need to be displayed in a clear fashion. For a similar reason respondent 3 has also graded implementation possibilities as average, stating that he/she is unsure of how a software tool would connect directly to the complex data warehouse they have in place. Respondent 2 has indicated that implementation possibilities are strong as it should not be difficult to create a tool based on the proposed methodology that is capable of accessing and analysing data directly from the data source. Respondent 3 has stated that he does not know enough about CRM to comment on implementation; however from the perspective of his company, he/she states that implementation issues would come from within the business not the methodology.

The respondents were asked their opinion regarding maintenance issues, if the customer profiling methodology was to be implemented as a software tool. None of the respondents could foresee any issues with maintaining this system. A table of questions and answers for the multiple choice questions in section 1 is presented in Table 29:

Table 29: Section 2 Multiple Choice Question Responses

	Respondent 1	Respondent 2	Respondent 3
The proposed methodology is a sound and valid step in the evolution of churn prediction techniques	4	4	4
The proposed methodology could strengthen the companies current CRM practices	4	4	4
The proposed methodology could be implemented by the company	4	4	4
1 = strongly disagree, 5 = strongly agree			

It can be seen from Table 29 that all respondent have agreed that the proposed methodology is a valid step in the evolution of churn prediction techniques. They also agree that the proposed methodology would strengthen their current CRM practices and it should not be very difficult to implement into current business strategies.

6.9.6. Section 3 - Churn Prediction Results

The third section of the questionnaire has been included to capture information regarding the churn capture that is currently being achieved by industry. All respondents have stated that they have seen predictions a maximum of 2 months future of initial churn identification; however they state that these results were weak in terms of actual numbers caught, while displaying large misclassification numbers. All respondents agree that current churn capture practices commonly achieve churn prediction accuracy of between 5% and 15%; and hit ratio is on average 0.25. They further agree that the customer profiling methodology is generating a much higher level of accuracy compared with current industry standards, and detecting churn further in the future than they have currently been able to achieve. Respondent 3 has developed churn capture strategies but within a much smaller time window.

When presented with the results from the customer profiling methodology all respondents graded the accuracy as either good or excellent. One of respondents is familiar with the data that has been used for the generation of the results and has stated that the data is not of the highest quality for churn prediction. He feels that if the data was of a higher standard the customer profiling methodology would show an even higher level of churn capture. He also states that the Hwang et al. (2004) methodology used for a comparison against the customer profiling methodology is a popular classification technique and is therefore a good choice for benchmarking. All respondents agree that from their industrial experience, churn prediction results currently in industry are generally very poor.

All respondents agree that the customer profiling methodology has captured churn at a high level of accuracy, sufficiently in advance of actual churn. They further agree that the strength of the methodology should be measured more on the hit ratio, and less by the actual churn capture, as they all state that churn capture has a high dependency on the quality of the data.

Comparing the customer profiling methodology with other solutions, all respondents favour the customer profiling methodology. The respondents have stated:

> Churn capture is not generally supported by 'off the shelf' products and they feel that the research is novel, and beneficial to industry.
>
> The customer profiling methodology generates information that current practices are unable to achieve.
>
> The customer profiling incorporates a time element that other techniques do not achieve.
>
> It would be interesting to monitor profile activity to see if the customers who display continuous loyalty have higher monetary values than other customers.
>
> The customer profiles could lead to further investigations that have previously been unavailable to the industry.

Table 30: Answers for Questionnaire Section 3

	Respondent 1	Respondent 2	Respondent 3
How do the results from the proposed methodology compare to those currently being achieved in industry	3	4	4
Please provide a rating for the churn capture that has been achieved from the profiling methodology	3	5	4
Please provide a rating for the non-churn capture that has been achieved from the profiling methodology	4	4	4
1 = Poor, 2 Average, 3 Good, 4 Very Good, 5 = Excellent			

It can be seen from the responses in Table 30 that there has been some discrepancy between the answers from each of the respondents. Respondent 1 develops and tests churn prediction techniques on a day to day basis, so has been meticulous when scoring the proposed methodology. Respondent 2 recognises that industrial CRM products do not sufficiently predict or identify customer churn, and has regarded the proposed methodology as excellent. From a data mining perspective, respondent 3 has also graded the proposed methodology as very good.

6.9.7. Conclusions

Capturing expert opinion through the use of semi-structured questionnaire has reinforced the benefits of the customer profiling methodology. It has been identified that industry commonly achieves between 5%-15% churn capture with an average hit ratio of 0.25. This means that the customer profiling methodology is achieving over double the industrial average churn capture, while minimising misclassification rates to fewer than half the industrial average. All experts agree that the customer profiling methodology is a robust and logical approach to the capture of customer churn.

7. Discussion and Conclusions

This chapter concludes this thesis with a discussion on the findings of this research. It also identifies the limitations of this work and the corresponding future research activities. This chapter aims to achieve the following.

To summarise key observations of this research

To identify the main contributions of this research

To discuss the limitations of this research

To frame future research activities

7.1. Discussion

This section discusses the key observations of this research.

7.1.1. Key Observations of This Research

Customer churn is a complex problem. The commercial CRM products provide little support for churn analysis and concentrate mainly on the areas of cross-selling and up-selling. The main reason for this is that the identification of cross-selling and up-selling opportunities is easier compared to the correct identification of customer churn.

Customer churn is a complex and difficult problem to predict. The main reasons for the complexity are:

Multiple causes of churn

Not every individual thinks alike

Not everyone has the same reasons for being with their service provider

Not everyone holds the same opinions regarding the service provider

Due to the many reasons that customers may churn, predictive models have traditionally been unpredictable and provide poor results. If a CRM software development company approached a business stating that their product would solve all the companies churn problems, it is doubtful that they would deliver on that promise and possibly damage their company name. This research is novel in that it specifically targets future churn prediction. It is important to capture churn accurately so that those customers who intend on defecting can be targeted with retention strategies. It is also important that the generated results do not misclassify a high portion of non-churners as churners because contacting high numbers is difficult and costly. To address the problem of future churn capture the customer profiling methodology creates advanced clusters of customer groups based on varying customer satisfaction levels, and identifies which clusters are most suitable for future churn capture by assessing each cluster for the future churn each has caught. All profiles that do not contain significant numbers of future churn are disregarded as churn predictors. This methodology actively minimises misclassification rates while the future reference in the methodology determines a method that is tailored for capturing future churn.

7.1.2. Literature Review and Gap Analysis

This research has looked into CRM, and the areas of customer retention and predictive techniques that could allow for the prediction of customer churn. It has analysed the drawbacks of current techniques to identify the main gaps in current research discussed. This research has progressed keeping these gaps in focus and as a result the churn prediction methodology has been developed.

The first research gap is that the current churn prediction techniques focus on the identification of immediate churn; future churn has found little focus in the literature. The proposed customer profiling methodology addresses this gap. Customer profiles are classified as high and low risk based on their suitability to capture future churn. The validation process documented in chapter 6 has shown that the profiles that have high churn capture at the model creation stage have continued to capture a high level of future churn. This has resulted in a methodology that is tailored towards capturing churn at a future point in time as opposed to the traditional methods.

The second gap identified by this research focuses on the data that is traditionally used for churn capture. It has been seen from the literature survey that usage and demographics data (such as RFM variables) are popular choices for basing churn models. In the UK, telecommunications and media companies are governed by Ofcom regulations. The purpose of Ofcom is to ensure that services are not monopolised by single companies. It achieves this by preventing certain types of analysis and encouraging fare competition. The major providers of telecommunications infrastructure, such as the company that has sponsored and supported this research, are therefore reluctant to use the usage and demographics data, In this research, a focus has been made on repairs and complaints data to ensure that capture of churn is determined for only those customers who are considering defecting because of service problems. This approaches the problem of capturing churn from an entirely different angle from traditional methods. The research has shown that repairs and complaints data can be used as a suitable addition to usage and demographics data, widening the scope of churn analysis.

The third gap is that the existing approaches do not support real time monitoring of the customer base. The customer profiling methodology has been developed in such a way that once the model is created it can be applied on a regular basis in the form of a moving time window. Therefore once the model has been created it can be reused on multiple time slots; every so often the high risk churn clusters should be monitored and assessed to ensure that environmental changes in the service sector have not affected the churn capture levels.

The fourth gap has been identified as an issue regarding misclassification levels. Misclassifying non-churning customers as churners can be costly to a business and can counteract the benefits of successfully capturing churn. The proposed customer profiling methodology largely reduces the misclassification of customers through the elimination of the profiles that are identified as low risk of containing future churn. It has been demonstrated through validation that a typical hit ration for churn prediction using the customer profiling methodology is between 0.5 and 0.6. This is over twice the accuracy that is typically achieved currently in industry (please see appendix A), suggesting a significant decrease in misclassification levels.

7.1.3. Methodology Development

A novel customer profiling methodology has been created from this research to address the limitations of existing techniques. Development of this methodology is carried in stages, adding new features as research progressed. As discussed in the previous section, the features targeted were those that have not been adequately addressed by current research.

Three predictive models for churn capture (neural networks, regression tress and linear regression) were investigated and the NN approach was identified as the most suitable. To be able to successfully develop the methodology, various stages were identified and followed. The first stage was the identification of the challenges in creating a churn prediction methodology. Investigations into this stage directed the research by uncovering typical problems regarding data and solutions to overcome them. This stage also investigated regulation restrictions to ensure that the developed solution would not breach any set of regulations which would result in an unusable methodology from a business context.

Development of the customer profiling methodology was not a straight forward process. Initial experiments suggested that it was relatively easy to identify churn from the same month as the source data being analysed but to extend that prediction into the future was much more challenging. This was achieved by developing customer profiles, and analysing them using a clustering method that uses future churn as a target during the model development stage. Because the best clusters for future churn prediction are generated from the analysed data, they are suitable for continuously targeting and capturing future churn.

7.1.4. Prototype Development

Prototype software was constructed for the customer profiling methodology by first performing an analysis of five software development methodologies. These methodologies were assessed for their applicability to the development of a customer profiling prototype software application. The analysis suggested that the XP software development methodology provided the closest fit and its stages were applied throughout the building of the software prototype. Incorporating a software development methodology aided in keeping the prototype in line with the customer profiling methodology, and helped to ensure that a stable prototype was constructed. The XP methodology proved to be a good framework for development as software releases and documentation were fast and of manageable size.

Each software release that had been determined from the XP methodology was presented in chapter 5 as a flow chart. These flow charts represent how the software releases were constructed. The final stage of development involves combining all staged releases into a single prototype. The flow charts help to visualise the final product.

The development of the customer profiling prototype was a large project with some challenging implementations. Advanced programming techniques had to be investigated for the prototype to conform to the research. Features such as MySQL database connectivity and chart generation are not directly supported by Borlands JBuilder which meant determining resources to support these features.

7.1.5. Validation

This research has analysed the performance of the customer profiling methodology in two ways. First it has been applied to three case studies. These case studies were selected from different service sectors. Second the results from the customer profiling methodology have been compared with two state-of-the-art techniques from literature.

When using the customer profiling methodology to determine future churn for Case 1, 31% of churn over a two month future period was successfully identified. This churn capture was reinforced by a low misclassification rate, achieving a hit ration of 0.51. When comparing this with the methodologies proposed by two other classification techniques that also attempt to identify future churn (Hu (2005) and Hwang et al. (2006)), the customer profiling methodology displays a significant increase in churn capture and a corresponding considerable decrease in misclassifications rates. The average hit ratios for both the above mentioned technologies is about 0.03 and the average churn capture for the analysed dataset is around 5%. The goal of the proposed methodology is to classify future churn with minimum misclassifications; the hit ratio is a good measure of this. Therefore, the comparison between all three methodologies is more accurately determined by the hit ratios of each of the methodologies rather than the actual future churn capture.

The pie chart presented in the literature survey chapter, based on Chu et al. (2007), provides another perspective for analysing the results from the proposed customer profiling methodology. This chart states that for the telecommunications industry, 11% of customers churn for reasons related to coverage, 11% of customers churn for reasons related to quality, 13% of customers churn because of issues with customer service, 47% of customers churn due to price issues and 18% of customer churn for other reasons. This information suggests that 47% of churn occurs because the customers have found a cheaper service. In this case the only real way to retain those customers would be to slash the price of the service to match that of the competitor. The pricing issue is a challenge because it is difficult to acquire the data to indicate the customers' intention to defect. The customers in this category could be completely happy with the service they are receiving however still churn because of price. From the rest 53% it can be assumed that not all customers would take the time to complain, so for a complete capture of this churn, multiple data types would require simultaneous analysis. It was shown in chapter 6 that case 1 captured 31% of churn for a future two month period. Taking into account that only the faults and complaints data was analysed, this represents a high accuracy of churn capture.

7.1.6. Business Impact Analysis

This section discusses the impact that the proposed methodology could have on the business. The business impact analysis is structured as follows:

Process Analysis – Identification of the contributions provided by the methodology for the fulfilment of the target problem.

Resource Requirements – Identification of required resources that should be available to the methodology

7.1.6.1. Process Analysis

The customer profiling methodology proposed by this research has addressed the problem of achieving an advanced prediction of customer churn while maintaining minimum error. This has been achieved through several processes. Each process has been developed to address specific objectives of the research. These processes are discussed below in context of the objectives they target.

The first objective was to develop a quantitative model for determining a churn index based on repairs and complaints data. This objective has been achieved through the identification of the most accurate predictive technique for predicting customer churn using repairs and complaints data. It has been demonstrated throughout this research that the churn index generated from the neural network provides an accurate churn index for presentation to the customer profiling methodology.

The second objective was the incorporation of a time series into the prediction of customer churn. This has been achieved by the process of basing a churn prediction on the generation of customer profiles. The customer profiles aid in maximising the time between churn prediction and churn occurrence in several ways. It is possible to part match customers to full profile clusters to determine a likelihood of the customer's lifetime with the company. Secondly the profiles are graded by their accuracy at predicting future churn. Finally, the profiles incorporate the average time it takes for members of that cluster to actually churn, based on historical data. All these features lead to an extension of timeframe between the point of classifying the customer as a churner and the actual churn occurrence. This also addresses the third objective.

The fourth objective was based on the requirement to control prediction error. The customer profiling methodology accomplishes this by pre-determining the most accurate master profile clusters for capturing future churn. All other clusters are eliminated, reducing misclassification. This is proven to be an accurate and effective method of reducing misclassification of non-churners as churners and significantly boosts churn prediction accuracy.

The fifth objective involved the validation of the proposed methodology using three case studies. This was included as an objective so that the predictive accuracy of the model could be determined and to illustrate the capability of applying the proposed methodology to various formats of data could be illustrated. The results obtained from all three cases are accurate. Misclassification rates are minimised resulting in a high hit ratio. This objective demonstrates that a large portion of customer churn can be adequately captured through the use of the customer profiling methodology while misclassification levels can be reduced to a small number.

The sixth objective was to identify and carry out three real-life case studies for validating the proposed methodology. This objective has been completed successfully. Three real-life case studies from specific telecommunications service sectors were secured and applied to the customer profiling methodology. In all three cases the customer profiling methodology determined high levels of churn capture, low misclassifications rates, hence high hit ratios. This objective has helped to demonstrate the strength of the developed methodology.

The business impact of the proposed methodology is viewed holistically. By identifying a large portion of churn typically 2-3 months before it actually occurs, a business has time to contact and retain those customers who plan to defect. This enables a business to minimise revenue loss and potentially increase revenue if combined with a cross sell, up sell campaign tailored to the customer's preference. The fact that misclassification rates are reduced to a small number ensures that the customers being contacted for retention have a high chance of being actual churners. This maximises the effectiveness of retention campaigns.

7.1.6.2. Resource Analysis

The proposed methodology requires the availability of several resources. Assuming that necessary equipment and resources such as computers and data storage are in place, the most important resource is the availability of data. To effectively identify churn using the proposed methodology, good quality data is required about the customers interactions with the company, preferably comprising of several months.

The proposed methodology also requires a neural network for the generation of churn index values. There are several ways that a company can gain access to neural network architecture. First, the business could code a neural network into the software platform and generate churn index values as the data is fed into the tool. The fastest and easiest way to generate a neural network would be through a package such as the neural network toolbox that is available for Matlab. Once a software platform based on the methodology is in place, the system should be light on it resources. Existing architectures will support the methodology and no additional upgrades would be required.

7.1.6.3. Maintaining the System

Once developed, maintenance of the system would be minimal. Profiles are added to the profile template database as and when identified in real time. This prevents the need to have to continuously run a full analysis frequently to update the master profile database. It would be advisable however to frequently check the future churn capture accuracy of master profile clusters. It is anticipated that shifts in cluster accuracy could occur through the fluctuations in business environment. For these reasons it is advised to analyse cluster churn accuracy levels every few months.

7.2. Main Contributions

The main contribution of this research is the new customer profiling methodology. This methodology contributes to research in the following ways.

The first contribution of this methodology is the inclusion of the probability of retention measures. Each customer profile is assessed to determine how long customers generally take to churn for that profile after their loyalty index values fall below the churn threshold. The results obtained by this feature will grow in accuracy as more and more customers are assigned to the profile cluster. This feature has not been monitored by previous research.

The second contribution of the profiling methodology is the fact that it incorporates future churn. The customer profiles are clustered by their ability to capture churn at a future point in time; these high risk profiles are then used to capture churn in the future.

The third contribution comes from the development of customer profiles. These profiles could provide a great deal of information regarding customer behaviour if combined with other analysis. For example combining profiles with an assessment of customer monetary value could show that customers who never display any increase in customer loyalty are actually the companies' most profitable customers, subscribing to multiple services with the highest volumes of usage.

The fourth contribution of the customer profiling methodology is its ability to minimise misclassification errors. This has been a goal of the research as identified by the initial research gaps and the profiling methodology accomplishes this to a high degree through the elimination of profile clusters that do not significantly contribute to future churn capture.

The fifth contribution of the profiling methodology is the potential to develop it into a real-time monitoring system. This can be achieved by enabling the software prototype to automatically read from the data warehouse in a continually moving time window.

As mentioned throughout this thesis, difficulties can arise if a predictive model is built from demographics and usage data. The methodology proposed in this thesis takes a novel approach of using customer repair and complaint data for churn prediction. This ensures of a model that can be implemented across service sectors without the risk of breaching any service regulations.

The proposed customer profiling methodology can be regarded as a new methodology for churn prediction unlike any other that have been developed in the past. The methodology could significantly widen the scope of churn analysis by providing the possibility of exploring new opportunities that have previously been unavailable.

7.3. Generality of Research

An attempt has been made in this research to keep it as general as possible. However, as with any other research, this work also has some limitations. Here, some of these limitations are identified.

7.3.1. Data Limitations

The success of any churn prediction methodology is hugely dependent on the data that is presented to it. Data limitations were encountered as follows:

It was originally anticipated that the methodology could be built through the analysis of RFM data. The Ofcom regulations were identified later in the research and a change in direction was required.

Access to case data has been very limited and restricted. Due to the fact that the successful prediction of customer churn is dependent on the data, the task of creating a suitable methodology has been very challenging.

The proposed methodology requires large volumes of data for development, testing and validation. Data limitations have had a significant impact on the time required for validation and the level that the methodology could be thoroughly validated.

A full analysis into the characteristics of the different profile clusters would have added value to the research. Unfortunately the data was not available to perform this analysis.

The data provided by the industrial sponsor had to be pre-processed to remove quality issues (such as invalid entries).

Data was provided for three separate services supplied by the sponsoring company. It would have been interesting to apply a dataset from an alternative organisation. However, because of the strategic nature of customer churn information it was not possible to source data from any other company.

7.3.2. Limitations of the Customer Profiling Methodology

The customer profiling methodology is limited by the way that the loyalty index value is generated. The most reliable customer loyalty measure should account for all aspects of customers and their environment. The proposed methodology was however limited to the faults and complaints data only. The proposed methodology should display even greater accuracy if combined with a more exhaustive measure of customer loyalty.

7.3.3. Limitations of the Customer Profiling Software Prototype

The prototype was not installed into a real industrial environment. It was only used as a tool for obtaining and validating the results from the proposed customer profiling methodology. Releasing software into industry is a long and complex process which was deemed beyond the scope of the research.

7.4. Future Research

An important area for future research is to use the proposed customer profiling methodology for developing a real-time monitoring system for churn prediction.

Research dedicated to the development of an exhaustive customer loyalty value would have significant benefits to industry. It would also improve the results generated from the customer profiling research.

Greater analysis of the customer profiles should be a focus for future research. It is anticipated that the profiling methodology could provide an insight into customer behaviour, spending patterns, cross-selling and up-selling opportunities. Seasonal trends could be apparent if the same data was studied over a period of several years. It is predicted that low risk clusters could display spontaneous movement to high risk as the business environment changes. The customers that have been matched to low risk clusters are stable customers at the time the profiles are generated but it is assumed that these customers will be the most sensitive to competing offers. To validate this claim a high level of data access and industrial implementation would be required.

7.5. Conclusions

This research has met its aim of developing a customer profiling methodology for predicting churn in advance, while keeping misclassification rates to a minimum. This section compares the achievements of this research with the objectives stated in Chapter 3. The following points analyse the key areas of this research:

> *Objective 1: To develop a quantitative model for churn index based on repairs and complaints data.*
>
> The literature survey has included a thorough analysis of existing techniques for churn prediction. Using the literature findings, a quantitative model for the generation of a customer churn index has been developed, based on complaints

and repairs data, to ensure that the final methodology does not breach governing regulations.

Objective 2: To develop a customer profiling methodology that incorporates time element in the prediction of customer churn.

Objective 3: To develop a customer profiling based technique for maximising future churn capture by identifying a potential loss of customer at the earliest possible point.

The development of a novel customer profiling methodology (that can enhance future churn capture while keeping misclassification rates to a minimum) has been documented. The developed methodology incorporates a time-element for maximising the time between the churn prediction and churn occurrence. The stages incorporated into the development of the proposed methodology could be followed and implemented by another researcher or business organisation.

Objective 4: To develop a customer profiling based technique for reducing misclassified customers, so that errors can be controlled.

The developed methodology is able to minimise misclassification levels by the identification of the most accurate customer profile clusters for future churn capture.

Objective 5: To identify and carry out three case studies for validating the proposed methodology using repairs and complaints data.

Objective 6: To compare the results from the proposed methodology against popular churn prediction techniques reported in literature.

A software prototype has been developed based on the proposed methodology to aid in testing and validation. The development of this prototype tool is clearly laid out, so the implementation of the proposed methodology can be replicated. The validation stage of this research has been carried out using the software prototype. It has demonstrated the strengths and limitations of the developed methodology. Three case studies from three different services within the telecommunications industry have been developed for validating the proposed methodology. The results from the proposed methodology have been compared against two state-of-the-art churn prediction techniques reported in literature. It has been demonstrated that the proposed methodology provides a greater accuracy of churn capture and better hit ratios compared with the other techniques. The research has demonstrated that a customer profiling approach to churn prediction performs better than current state-of-the-art techniques.

The main achievement of this research can be briefly and precisely stated as follows:

Development of a customer profiling methodology for predicting churn in advance, while keeping the misclassification rates to a minimum.

8.References

AHN, J., HAN, S. & LEE, Y. (2006) Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry. *Telecommunications Policy,* 10, 552-568.

ANG, L. & BUTTLE, F. A. (2002) ROI on CRM: a customer journey approach. *18th IMP-conference*. Australia, IMP Group in Asia.

APENYO, K. (1999) Using the entity-relationship model to teach the relational model. *ACM SIGCSE Bulletin,* 31, 78-80.

ATHANASSOPOULOS, A. D. (2000) Customer Satisfaction Cues to Support Market Segmentation and Explain Switching Behaviour. *Journal of Business Research,* 47, 191-207.

AU, W., CHAN, C. C. & YAO, X. (2003) A novel evolutionary data mining algorithm with applications to churn prediction. *IEEE transactions on evolutionary computation,* 7, 532-545.

AUH, S. & JOHNSON, M. D. (2004) Compatibility Effects in Evaluations of Satisfaction and Loyalty. *Journal of Economic Psychology,* 26, 35-57.

AYTEK, A. & KISI, Ö. (2008) A genetic programming approach to suspended sediment modelling. *Journal of Hydrology,* 351, 288-298.

BAESENS, B., VERSTRAETEN, G., VAN DEN POEL, D., EGMONT-PETERSON, M., VAN KENHOVE, P. & 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.

BAIRD, S. (2003) *Teach Yourself Extreme Programming In 24 Hours 2003,* Indianapolis, USA, Sams Publishing.

BEHARA, R. S., FISHER, W. W. & LEMMINK, J. (2002) Modelling and evaluating service quality measurement using neural networks. *International journal of operations and production management,* 22, 1162-1185.

BELLAZZI, R. & ZUPAN, B. (2008) Predictive data mining in clinical medicine: Current issues and guidelines. *International journal of medical informatics,* 77, 81-97.

BERNE, C., MUGICA, J. M. & YAGUE, M. J. (2001) The Effect of Variety Seeking on Customer Retention in Services. *Journal Of Retailing and Consumer Services,* 8, 335-345.

BLOEMER, J., BRIJIS, T., VANHOOF, K. & SWINNEN, G. (2002) Comparing complete and partial classification for identifying customers at risk. *International journal of research in marketing,* 20, 117-131.

BOONE, D. S. & ROEHM, M. (2002) Retail Segmentation Using Artificial Neural Networks. *International journal of research in marketing,* 19, 287-301.

BRUHN, M. & GRUND, M. (2000) Theory, development and implementation of national customer satisfaction indices: The Swiss index of customer satisfaction. *Total Quality Management,* 11, 1017-1028.

BUCKINX, W. & VAN DEN POEL, D. (2004) Customer base analysis: Partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting. *European journal of operational research,* 164, 252-268.

BUREZ, J. & VAN DEN POEL, D. (2008) Separating financial from commercial customer churn: A modeling step towards resolving the conflict between the sales and credit department. *Expert Systems with Applications,* xxx, In Press, Corrected Proof.

CABALLERO, J., GARRIGA, M. & FERNÁNDEZ, M. (2006) 2D Autocorrelation modeling of the negative inotropic activity of calcium entry blockers using Bayesian-regularized genetic neural networks. *Bioorganic & Medicinal Chemistry,* 14, 3330-3340.

CANNING, G. (1982) Do A Value Analysis of Your Customer Base. *Industrial Marketing Management,* 11, 89-93.

CEVIK, A. & GUZELBEY, I. (2008) Neural network modeling of strength enhancement for CFRP confined concrete cylinders. *Building and Environment,* 43, 751-763.

CHEN, I. J. & POPOVICH, K. (2003) Understanding Customer Relationship Management (CRM). *Business Process Management Journal,* 9, 672-688.

CHEN, W., TSENG, S. & HONG, T. (2008) An efficient bit-based feature selection method. *Expert Systems with Applications,* 34, 2858-2869.

CHIANG, D., WANG, Y., LEE, S. & LIN, C. (2003) Goal-oriented sequential pattern for network banking and churn analysis. *Expert systems with applications,* 25, 293-302.

CHU, B., TSAI, M. & HO, C. (2007) Toward a hybrid data mining model for customer retention. *Knowledge-Based Systems,* 20, 703-718.

CONKLIN, M., POWAGA, K. & LIPOVETSKY, S. (2004) Customer Satisfaction Analysis: Identification of Key Drivers. *European Journal of operational Research,* 154, 819-827.

CORTIÑAS, M., ELORZ, M. & MÚGICA, J. (2008) The use of loyalty-cards databases: Differences in regular price and discount sensitivity in the brand choice decision between card and non-card holders. *Journal of Retailing and Consumer Services,* 15, 52-62.

COUSSEMENT, K. & VAN DEN POEL, D. (2008) Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. *Expert Systems with Applications,* 34, 313-327.

CRESPO, F. & WEBER, R. (2004) A Methodology for Dynamic Data Mining Based on Fuzzy Clustering. *Fuzzy Sets and Systems,* 150, 267-284.

DATTA, P., MASAND, B., MANI, D. R. & LI, B. (2001) Automated cellular modelling and prediction on a large scale. *Issues on the application of data mining*, 485-502.

DEMUTH, H., BEALE, M. & HAGAN, M. (2007) Neural Network Toolbox 5 User Guide. IN INC, T. M. (Ed.) 2007b ed., The Mathworks Inc.

DO, D. Q., ROWE, R. C. & YORK, P. (2008) Modelling drug dissolution from controlled release products using genetic programming. *International Journal of Pharmaceutics,* 351, 194-200.

DUDOIT, S. & VAN DER LAAN, M. (2005) Asymptotics of cross-validated risk estimation in estimator selection and performance assessment. *Statistical Methodology,* 2, 131-154.

EPSRC (2008) Mission and Vision. IN EPSRC (Ed.) *About EPSRC*. Swindon, UK, EPSRC.

ERIKSSON, K. & FMARCK VAGHULT, A. L. (2000) Customer Retention, Purchasing Behaviour and Relationship Substance in Professional Services. *Industrial Marketing Management,* 29, 363-372.

ERIKSSON, K. & VAGHULT, A. M. (2000) Customer Retention, Purchasing Behavior and Relationship Substance in Professional Services. *Industrial Marketing Management,* 29, 363-372.

ESHGHI, A., HAUGHTON, D. & TOPI, H. (2007) Determinants of customer loyalty in the wireless telecommunications industry. *Telecommunications Policy,* 31, 93-106.

FARQUHAR, J. D. & PANTHER, T. (2008) Acquiring and retaining customers in UK banks: An exploratory study. *Journal of Retailing and Consumer Services,* 15, 9-21.

GANS, J. S. (2000) Network competition and consumer churn. *Information economics and policy,* 12, 97-109.

GERPOTT, T. J., RAMS, W. & SCHINDLER, A. (2000) Customer retention, loyalty, and satisfaction in the German mobile cellular telecommunications market. *Telecommunications Policy,* 25, 249-269.

GRAY, J. B. & FAN, G. (2008) Classification tree analysis using TARGET. *Computational Statistics & Data Analysis,* 52, 1362-1372.

HADDEN, J., TIWARI, A., ROY, R. & RUTA, D. (2006a) Churn Prediction Using Complaints Data. *Transactions on Engineering, Computing and Technology.* Budapest.

HADDEN, J., TIWARI, A., ROY, R. & RUTA, D. (2006b) Churn Prediction: Does Technology Matter? *International Journal of Intelligent Technology,* 1, 104 - 110.

HADDEN, J., TIWARI, A., ROY, R. & RUTA, D. (2006c) Computer assisted customer churn management: State-of-the-art and future trends. *Computers & Operations Research,* 10, 2902-2917.

HAHM, J., CHU, W. & YOON, J. W. (1997) A Strategic Approach to Customer Satisfaction in the Telecommunications Service Market. *Computers & Industrial Engineering,* 33, 825-828.

HAN, F., LING, Q. & HUANG, D. (2008) Modified constrained learning algorithms incorporating additional functional constraints into neural networks. *Information Sciences,* 3, 907-919.

HIDALGO, P., MANZUR, E., OLAVARRIETA, S. & FARÍAS, P. (2008) Customer retention and price matching: The AFPs case. *Journal of Business Research,* xxx, In Press, Corrected Proof.

HSIEH, N. (2004) An Integrated Data Mining and Behavioural Scoring Model for Analysing Bank Customers. *Expert systems with applications,* 27, 623-633.

HSU, S. & CHIEN, C. (2007) Hybrid data mining approach for pattern extraction from wafer bin map to improve yield in semiconductor manufacturing. *International Journal of Production Economics,* 107, 88-103.

HSU, S. H. (2008) Developing an index for online customer satisfaction: Adaptation of American Customer Satisfaction Index. *Expert Systems with Applications,* 34, 3033-3042.

HU, X. (2005) A Data Mining Approach for Retailing Bank Customer Attrition Analysis. *Applied Intelligence,* 22, 47–60.

HUNG, S., YEN, D. C. & WANG, H. (2006) Applying data mining to telecom churn management. *Expert Systems with Applications,* 31, 515-524.

HWANG, H., JUNG, T. & SUH, E. (2004) An LTV Model and Customer Segmentation Based on Customer Value: A Case Study on the Wireless Telecommunications Industry. *Expert systems with applications,* 26**,** 181-188.

JENAMANI, M., MOHAPATRA, P. K. J. & GHOSE, S. (2003) A Stochastic model of e-customer behaviour. *Electronic commerce research and applications,* 2**,** 81-94.

JONKER, J., PIERSMA, N. & 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.

KAVZOGLU, T. & MATHER, P. M. (2001) The role of feature selection in artificial neural network applications. *International Journal of Remote Sensing,* 23**,** 2919-2937.

KENDALL, K. E. (2005) *Systems analysis and design* New Jersey Prentice Hall.

KEREN, D. (2003) Recognizing image "style" and activities in video using local features and naive Bayes. *Pattern Recognition Letters,* 24**,** 2913-2922.

KIM, H. & YOON, C. (2004) Determinants of Subscriber Churn and Customer Loyalty in the Korean Mobile Telephony Market. *Telecommunications Policy,* 28**,** 751-765.

KIM, K., JEONG, I., PARK, J., PARK, Y., KIM, C. & KIM, T. (2007) The impact of network service performance on customer satisfaction and loyalty: High-speed internet service case in Korea. *Expert Systems with Applications,* 32**,** 822-831.

KIM, M., PARK, M. & JEONG, D. (2004) The Effects of Customer Satisfaction and Switching Barrier on Customer Loyalty in Korean Mobile Telecommunications Services. *Telecommunications Policy,* 28**,** 145-159.

KITAYAMA, M., MATSUBARA, R. & IZUI, Y. (2002a) Application of data mining to customer profile analysis in the power electric industry. *Power Engineering Society Winter Meeting*.

KITAYAMA, M., MATSUBARA, R. & IZUI, Y. (2002b) Data mining for customer load profile analysis. *Transmission and Distribution Conference and Exhibition 2002*. Asia Pacific, IEEE/PES.

KNOX, S. (1998) Loyalty-Based Segmentation and the Customer Development Process. *European Management Journal,* 16**,** 729-737.

KOZA, J. R. (1992) *Genetic Programming: On the Programming of Computers by Means of Natural Selection,* Cambridge, MA., Massachusetts Institute of Technology.

LARIVIERE, B. & VAN DEN POEL, D. (2004) Investigating The Role Of Product Features in Preventing Customer Churn, by Using Survival Analysis and

Choice Modeling: The Case of Financial Services. *Expert systems with applications,* 27, 277-285.

LEE, H. (1999) Semantics of recursive relationships in entity-relationship model. *Information and Software Technology,* 41, 877-886.

LEE, T., CHIU, C., CHOU, Y. & LU, C. (2006) Mining the customer credit using classification and regression tree and multivariate adaptive regression splines. *Computational Statistics & Data Analysis,* 50, 1113-1130.

LIAO, S. & CHEN, Y. (2004) Mining Customer Knowledge for Electronic Catalog Marketing. *Expert systems with applications,* 27, 521-532.

LIAO, Y., FANG, S. & NUTTLE, H. (2004) A neural network model with bounded-weights for pattern classification. *A neural network model with bounded-weights for pattern classification,* 31, 1411-1426.

LIU, D. & SHIH, Y. (2004a) Hybrid Approaches to Product Recommendation Based on Customer Lifetime Value and Purchase Preferences. *The journal of Systems and Software,* 77, 181-191.

LIU, D. & SHIH, Y. (2004b) Integrating AHP and Data mining for Product Recommendation Based on Customer Lifetime Value. *Information & Management,* 42, 387-400.

LIU, X., ZENG, X., XU, Y. & KOEHL, L. (2008) A fuzzy model of customer satisfaction index in e-commerce. *Mathematics and Computers in Simulation*, In Press, Corrected Proof.

LOZANO, F. J., SUÁREZ-SEOANE, S., KELLY, M. & LUIS, E. (2008) A multi-scale approach for modeling fire occurrence probability using satellite data and classification trees: A case study in a mountainous Mediterranean region. *Remote Sensing of Environment,* 112, 708-719.

MA, M., LI, Z. & CHEN, J. (2008) Phase-type distribution of customer relationship with Markovian response and marketing expenditure decision on the customer lifetime value. *European Journal of Operational Research,* 187, 313-326.

MADDEN, G., SAVAGE, G. & COBLE-NEAL, G. (1999) Subscriber churn in the Australian ISP market. *Information economics and policy,* 11, 195-207.

MADNICK, S. & ZHU, H. (2006) Improving data quality through effective use of data semantics. *Data & Knowledge Engineering,* 59, 460-475.

MANKILA, M. (2004) Retaining students in retail banking through price bundling: Evidence from the Swedish market. *European Journal of operational Research,* 155, 299-316.

MARCZYK, G., DE MATTEO, D. & FESTINGER, D. (2005) IN JOHN WILEY & SONS, I. (Ed.) *Essentials of Research Design and Methodology.* New Jersey, USA John Wiley & Sons, Inc.

MEYER-BASE, A. & WATZEL, R. (1998) Transformation Radial Basis Neural Network for Relevant Feature Selection. *Pattern Recognition Letters,* 19, 1301-1306.

MEYER-WAARDEN, L. (2006) The effects of loyalty programs on customer lifetime duration and share of wallet. *Journal of Retailing* 83, 223-236.

MIHELIS, G., GRIGOROUDIS, E., SISKOS, Y., POLITIS, Y. & MALANDRAKIS, Y. (2001) Customer Satisfaction Measurement in the Private Bank Sector. *European Journal of operational Research,* 130, 347-360.

MINAMI, C. & DAWSON, J. (2008) The CRM process in retail and service sector firms in Japan: Loyalty development and financial return. *Journal of Retailing and Consumer Services,* xxx, In Press, Corrected Proof.

MOZER, M. C., WOLNIEWICZ, R., GRIMES, D. B., JOHNSON, E. & KAUSHANSKY, H. (2000) Predicting Subscriber Dissatisfaction and Improving Retention in the Wireless Telecommunications Industry. *IEEE Transactions on Neural Networks,* 11, 690-696.

MUATA, K. & BRYSON, O. (2004a) Evaluation of Decision Trees, A Multi Criteria Approach. *Computers and Operational Research,* 31, 1933-1945.

MUATA, K. & BRYSON, O. (2004b) Evaluation of decision trees: a multi-criteria approach. *Computers & Operations Research,* 31, 759-767.

NANNINGS, B., ABU-HANNA, A. & DE JONGE, E. (2008) Applying PRIM (Patient Rule Induction Method) and logistic regression for selecting high-risk subgroups in very elderly ICU patients. *International Journal of Medical Informatics,* 77, 272-279.

NEFESLIOGLU, H. A., GOKCEOGLU, C. & SONMEZ, H. (2008) An assessment on the use of logistic regression and artificial neural networks with different sampling strategies for the preparation of landslide susceptibility maps. *Engineering Geology,* 97, 171-191.

NG, K. & LIU, H. (2001) Customer retention via data mining. *Issues on the application of data mining*, 569-590.

PAL, S. & GHOSH, A. (2004) Soft Computing Data Mining. *Information Sciences,* 163, 1-3.

PRINZIE, A. & VAN DEN POEL, D. (2004) Investigating purchasing-sequence patterns for financial services using Markov, MTG and MTGg models. *European journal of operational research,* 170, 710-734.

RAM, S. (1995) Intelligent database design using the unifying semantic model. *Information and Management,* 5, 191-206.

RAM, S. & KHATRI, V. (2005) A comprehensive framework for modeling set-based business rules during conceptual database design. *Information Systems* 30, 89-118

RANAWEERA, C. & NEELEY, A. (2003) Some Moderate Effects on The Service Quality-Customer Retention Link. *International journal of operations and production management,* 23, 230-248.

ROSSET, S., NEUMANN, E., EICK, U. & VATNIK, N. (2003) Customer lifetime value models for decision support. *Data mining and knowledge discovery,* 7, 321-339.

RUNDLE-THIELE, S. (2005) Elaborating customer loyalty: exploring loyalty to wine retailers. *Journal of Retailing and Consumer Services,* 12, 333-344.

RUST, R. T. & ZAHORIK, A. J. (1993) Customer Satisfaction, Customer Retention, and Market Share. *Journal of retailing,* 69, 193-215.

RYALS, L. (2002) Are your Customers Worth More Than Money? *Journal Of Retailing and Consumer Services,* 9, 241-251.

RYGIELSKI, C., WANG, J. & YEN, D. C. (2002) Data Mining Techniques for Customer Relationship Management. *Technology in Society,* 24, 483-502.

SAINI, L. M. (2008) Peak load forecasting using Bayesian regularization, Resilient and adaptive backpropagation learning Based artificial neural networks. *Electric Power Systems Research*, In Press, Corrected Proof.

SEO, D., RANGANATHAN, C. & BABAD, Y. (2008) Two-level model of customer retention in the US mobile telecommunications service market. *Telecommunications Policy*, In Press, Corrected Proof.

SHETH, J. N. & PARVATIYAR, A. (1995) The Evolution of Relationship Marketing. *International Business Review,* 4, 397-418.

SHIN, H. W. & SOHN, S. Y. (2004) Segmentation of Stock Trading Customers According to Potential Value. *Expert systems with applications,* 27, 27-33.

SISKOS, Y., GRIGOROUDIS, E., ZOPOUNIDIS, C. & SAURAIS, O. (1998) Measuring customer satisfaction using a collective preference disaggregation model. *Journal of global optimization,* 12, 175-195.

SLOTNICK, S. A. & SOBEL, M. J. (2005) Manufacturing lead-time rules: Customer retention versus tardiness costs. *European journal of operational research,* 163, 825-856.

SOUITARIS, V. & BALABANIS, G. (2007) Tailoring Online Retail Strategies to Increase Customer Satisfaction and Loyalty. *Long Range Planning,* 40, 244-261.

SPITERI, J. M. & DION, P. A. (2004) Customer Value, Overall Satisfaction, End-User Loyalty, and Market Performance in Detail Intensive Industries. *Industrial Marketing Management,* 33, 675-687.

STAHL, H. K., MATZLER, K. & HINTERHUBER, H. H. (2003) Linking Customer Lifetime Value With Shareholder Value. *Industrial Marketing Management,* 32, 267-279.

SUN, Z., BEBIS, G. & MILLER, R. (2004) Object detection using feature subset selection. *Pattern Recognition,* 37, 2165-2176.

SWEENEY, J. & SWAIT, J. (2008) The effects of brand credibility on customer loyalty. *Journal of Retailing and Consumer Services,* xxx, In Press, Corrected Proof.

TAN, K., YU, Q., HENG, C. M. & LEE, T. H. (2005) Evolutionary computing for knowledge discovery in medical diagnosis. *Artificial Intelligence in Medicine,* 27, 129-154.

TAN, S. (2006) An effective refinement strategy for KNN text classifier. *Expert Systems with Applications,* 30, 290-298.

TAO, Y. H. & YEH, C. R. (2003) Simple database marketing tools in customer analysis and retention. *International journal of information management,* 23, 291-301.

TINDALL, C. I., MOORE, R. V., BOSLEY, J. D., SWETNAM, R. D., BOWIE, R. & DE RUDDER, A. (2006) Creating and using the urgent metadata catalogue and thesaurus. *Science of The Total Environment,* 360, 223-232.

TUREL , O. & SERENKO, A. (2006) Satisfaction with mobile services in Canada: An empirical investigation. *Telecommunications Policy,* 30, 314-331.

VAN DEN POEL, D. (2003) Predicting mail-order repeat buying: Which variables matter? Hoveniersberg, Ghent University.

VAN DEN POEL, D. & LARIVIERE, B. (2003) Customer attrition analysis for financial services using proportional hazard models. *European journal of operational research,* 157, 196-217.

VELLIDO, A., LISBOA, P. J. G. & MEEHAN, K. (1999) Segmentation of the Online Shopping Market Using Neural Networks. *Expert systems with applications,* 17, 303-314.

VERHOEF, P. C. & DONKERS, B. (2001) Predicting Customer Potential Value an Application in the Insurance Industry. *Decision Support Systems,* 32, 189-199.

VOLZ, R., HANDSCHUH, S., STAAB, S., STOJANOVIC, L. & STOJANOVIC, N. (2004) Unveiling the hidden bride: deep annotation for mapping and migrating legacy data to the Semantic Web. *Web Semantics: Science, Services and Agents on the World Wide Web,* 1**,** 187-206.

WANG, W., KAO, C. & LEE, W. (2007) A new interactive model for improving the learning performance of back propagation neural network. *Automation in Construction,* 16**,** 745-758.

WEISBERG, S. (2005) *Applied Linear Regression,* Minnesota, USA, John Wiley & Sons, Inc.

WIKIPEDIA (2008a) British Telecom. IN WIKIPEDIA (Ed.) *Wikipedia The Free Encyclopaedia*. St. Petersburg, Wikimedia Foundation, Inc.

WIKIPEDIA (2008b) Ofcom. IN WIKIPEDIA (Ed.) *Wikipedia The Free Encyclopaedia*. St. Petersburg, Wikimedia Foundation, Inc.

WYTHOFF, B. (1993) Backpropagation neural networks: A tutorial. *Chemometrics and Intelligent Laboratory Systems,* 18**,** 115-155.

YAN, L., WOLNIEWICZ, R. & DODIER, R. (2004) Predicting customer behaviour in telecommunications. *IEE Intelligent Systems,* 19**,** 50-58.

YANG, J. & PENG, S. (2008) Development of a customer satisfaction evaluation model for construction project management. *Building and Environment,* 43**,** 458-468.

YI, Y. (1989) A Critical Review of Customer Satisfaction. *Review of Marketing*. Chicago, US, American Marketing Association.

APPENDIX A

Capturing Expert Opinion Questionnaire

Expert Opinion Regarding
Customer Profiling Methodology

Interviewer: John Hadden

Interviewee:________________

Section 1 – Churn Management Practice

Q1) Does your company develop 'in-house' solutions for customer churn management?

(If yes please state your development role)

.

Q2) What technologies are mainly used in development of churn management software? (Please tick all that apply)

Technology	Yes	No
Neural Networks	[ ]	[ ]
Classification Trees	[ ]	[ ]
Regression Analysis	[ ]	[ ]
Evolutionary Computing	[ ]	[ ]
Markov Model	[ ]	[ ]
Other	[ ]	[ ]

Q3) If answered 'other' to question 4 – Please specify?

Q4) Does your company use any 'off the shelf' churn management solutions? (Please state)

Q5) If answered yes to question 6 – how well do these products compare to the solution that are developed in-house?

Q6) Please provide any additional information or views regarding the capabilities of current churn management (including pros and cons) that you feel maybe important to this research?

Section 2 – Customer Churn Management Methodology

The questions in this section are structured to capture your views on the methodology that is outlined in section 1. Please study section 1 before answering these questions to ensure that an accurate response is provided. Please encircle the appropriate box for each question, 1 = strongly disagree, 5 = strongly agree.

Q1) The proposed methodology is a sound and valid step in the evolution of customer churn management techniques?

Strongly Agree				Strongly Disagree
1	2	3	4	5

Please give reason for answer

Q2) From a business point of view, the proposed methodology is practical

Strongly Agree				Strongly Disagree
1	2	3	4	5

Please give reason for answer

Q3) The proposed methodology could be implemented into your current CRM process

Strongly Agree				Strongly Disagree
1	2	3	4	5

Please give reason for answer

Q4) What do you think would be the challenges in implementing the proposed methodology?

Q5) What do you think would be the challenges in maintaining the proposed system?

Q6) Please provide any other comments you have regarding the methodology

Section 3 – Churn Prediction

This is the final section of the questionnaire and it is designed to capture any opinion regarding the results of the methodology. Please encircle 1 answer for each question; 1 = poor, 5 = exceptional.

Q1) How far into the future does your current system predict churn?

Q2) The results displayed in section 1 are churn predictions 2 months after the customer had originally been identified and classified as a defector. How does this measure to current standards?

Poor	Average	Good	Excellent	Exceptional
1	2	3	4	5

Please give reason for answer

Q3) From the three case studies, please provide a rating for the churn accuracy based on your knowledge of current industry standards:

Poor				Exceptional
1	2	3	4	5

Please give reason for answer:

Q4) From the three case studies, please provide a rating for the non-churn accuracy based on your knowledge of current industry standards:

Strongly Agree				Strongly Disagree
1	2	3	4	5

Please give reason for answer:

Q5) Please state how well you think the methodology compares to current churn management practices

Q6) Please state any other comments you have regarding the proposed methodology here?

Section 4 – Respondent Background

This section of the questionnaire is optional. The sole purpose of this section is to gather some information about you so that your comments may be evaluated in the right perspective. In case you do not feel comfortable in answering any part of this section, please ignore it. If you are happy to answer a question, please tick the appropriate box.

Q1) Churn prediction

	Best				Worst
Knowledge	1	2	3	4	5
Experience	1	2	3	4	5

Q2) Programming

	Best				Worst
Knowledge	1	2	3	4	5
Experience	1	2	3	4	5

Q3) Data mining

	Best				Worst
Knowledge	1	2	3	4	5
Experience	1	2	3	4	5

Q4) Off the shelf CRM or churn management software

	Best				Worst
Knowledge	1	2	3	4	5
Experience	1	2	3	4	5

Q5) Working in the field of customer management

	Best				Worst
Knowledge	1	2	3	4	5
Experience	1	2	3	4	5

Thank you for your time in completing this questionnaire. Your expert knowledge
is very valuable to this research and greatly appreciated

APPENDIX B

CRM Software Analysis

Cranfield University

CRM Analysis

CRM Software and Vendor Analysis

John Hadden

Academic Supervisors:

Dr Ashutosh Tiwari and Dr Rajkumar Roy

CRM Vendor Analysis

This report has been prepared as part of PhD research entitled 'Analysis of Soft Computing Based Software Platform for Customer Churn Management'.

Aim and Objectives

The aim of the report is to provide a comparative analysis of the most popular CRM vendors in the industry, identifying key trends and areas of particular focus. Various objectives have had to be carried out. These objectives include: -

Identification of Industry Leaders

Vendor Investigation

Software Analysis

Comparative Study

Identification of Published Work (if any)

Survey Methodology

The following section provides a brief description of the methodology used in the preparation of this report.

Industry Leaders

There are dozens of vendors offering CRM software, therefore the top nine leaders in the market have been identified, and these are the ones that have been researched. Each vendor will be examined in its own section of this paper, concluding with a comparison of all. The vendors have been identified through the process of initial research, and discussions with industry professionals.

The list of vendors that have been identified as the most popular and more successful in the industry are as follows: -

Siebel Systems CRM on demand – A CRM solution for unifying, tracking and managing customer information and processes across sales, marketing and customer service. The software runs over the internet.

Amdocs ClarifyCRM – Amdocs has tried to cater for every aspect of the customer life-time cycle. They offer a broad selection of products which are intended to link all customer facing business processes

PeopleSoft CRM – As with Siebel, PeopleSoft CRM is a pure internet application. It claims to be the only vendor to provide a solution for changing customer demand to adapt planning and production in real-time.

MySAP CRM – MySap enables complete integration and easy collaboration over the internet. It claims to offer complete life-cycle management. Again, MySap is a web-service based product.

Oracle CRM – Oracle's solution offers a means of identify customer requirements and providing a broad range of implementation resources. Customers have been involved in the design and development stages in order to ensure a product which meets the customers needs.

Microsoft – Microsoft CRM has been built from the ground up with the Microsoft .NET framework. Microsoft have made there CRM solution compatible with Outlook, Word and other products so creation and deployment

Zero Attrition eCVM – eCVM is a term that has been created by Zero Attrition. It stands for Enterprise Customer Value Management and combines various fundamentals of CRM, pCRM (predictive CRM) and Risk Management.

SPSS Base – SPSS Base boasts a huge variety of services including service and market research, academic support, planning and forecasting and quality improvement etc. SPSS is the full product including everything,, but it can also be purchased in chunks giving the customer the choice to only buy what is required.

CustomerSat – CustomerSat is another solution that is hosted online. They claim that their CRM product offers support for all survey types, such as transactional, periodic, etc, and provides expert services for design, analysis, reporting and recommendations.

All information in this report has been acquired directly from each of the vendor's websites with an additional search of available journal databases in the attempt of identifying published work related to each of the vendor's products. An extensive research phase has lead to a collection of product brochures, specifications, white papers for each of the products. Some journal papers have also been identified for certain vendors but it has been assumed after a thorough investigation that not all software has related publications associated with them.

Criteria for Analysis

The above products will be analysed based on certain criteria. These criteria have been identified as follows: -

Phases of the customer life-cycle covered – This criteria should make it possible to identify trends (i.e. the area(s) where most focus has been given). With an analysis of these trends amongst the various products, a clear picture should emerge as to where the weaknesses can be found in current CRM products. The stages of the customer life-cycle are as follows: -

- **Prospects –** The people that are in the target market, but have not actually become customers.

- **Responders –** These are the people that have made a strong enquiry but are not actual customers. The process of converting responders to actual customers differs with each industry.

- **Established Customers –** These are actual customers using the company's products or services. When first entering this phase the customer is known to be a new customer. The behaviour of a new customer is often highly predictive of future behaviour.

- **Former Customers –** These are the customers that have churned, either voluntarily or involuntarily.

Churn management tool included – If the product includes a churn management tool it suggests that the vendor realises the importance of customer retention. If a tool of this kind is included an analysis will be carried out in order to determine exactly what it does.

Churn prediction – This document is part of a PhD which has a lot of interest in churn prediction. If it is found that any of the software includes capabilities for churn prediction, an analysis will be carried out surrounding the techniques that have been used.

Technologies used for development – This criteria is intended to identify if the software uses any kind of soft computing/artificial intelligence techniques, or if development has been focused around statistics or any other technologies.

Forecasting – Forecasting across various levels is very advantageous to a business. An investigation will be carried out on all products to see if forecasting is supported, at what level, and the types of predictive techniques used.

Can software be customised? – CRM can be very specific to various service sectors, these criteria will provide information about whether the product can be customised in order to meet a specific company's needs.

After sales and service support – After sales service and support is an important consideration for most products. CRM is no acceptation so an analysis of the company's after sales service will be undertaken.

Training – Training is another important factor that has to be taken into consideration. It would be useful to know if any training programs can be arranged and the complexity of the software etc.

Examples of some current clients (reference sites) – the final criteria for analysis will identify the service industries and some companies using the software. This will provide an idea about what size company the software can be successful implemented with, and how it is regarded within industry.

The above criteria have been selected because using these headings as a guide a full and complete analysis of each of the products could be achieved.

Software Analysis

The above section briefly described each of the vendors and their software. The following section will provide a more detailed analysis of each of the products, based on the mentioned criteria.

Siebel Systems

Siebel CRM offers full analytical capabilities providing a company with the ability to gain a deep insight into their business. The software has been designed to: -

Identify opportunities
Determine which customers produce the highest margins
Understand the root causes of service problems
Gain insight into sales, service and customer trends
Generate reports
Conduct comparative and historical trend analysis

Siebel has been in the CRM business for over ten years. They claim to of designed their product to identify, acquire, serve, and retain profitable customers. Their product brochure also states that the entire customer life-cycle has been catered for (Siebel, 2004c).

Customer Life-Cycle

Stage 1 - Prospects

The first stage of the customer life-cycle is addressed in Seibel CRM OnDemand's marketing paper. Their marketing paper, Siebel (2004f) states that "Siebel CRM OnDemand provides a complete closed-loop solution, so that you can track leads through each stage of the lead management process – from lead qualification to closing revenue". The paper adds that leads can be associated with a company, a contact, a campaign, or any combination of all three.

Stage 2 – Responders

Siebel's sales paper addresses stage two of the customer life-cycle. Tools are included with Siebel for auto-forecasting, embedded analytics, and closed loop lead management. Information is provided about decision makers, partners, competitors, history and revenue potential (Siebel, 2004g).

Stage 3 – Established Customers

Once a company has established their customers, serious efforts have to be made to ensure that they keep them. Siebel has tried to cater for established customers by offering customer service and support, and analytical tools. Their software includes assignment rules, so incoming service requests automatically get routed to the right departments, in an attempt to meet the customer's needs. The customer service representatives are provided with instant access to complete customer records and a knowledge base providing solutions to the most common problems (Siebel, 2004c).

Stage 4 – Former Customers

Information on the final stage of the customer life-cycle doesn't seem to be very well documented. References are made to reducing churn in the product description section of Siebel's website (2004a), but no information has been found about churn tools, prediction, analysis of churn rates, analysis of causes of churn etc. This information is regarded by the author as important, as it would provide a company with an insight about were they are going wrong, and how improvements could be made in the efforts of retaining more customers. It appears at this point as though Siebel has focused most of its attention to customer acquisition and satisfaction. This does not mean in any way that Siebel does not provide churn management or prediction tools, only that the author has failed to identify any in his research.

Technologies Used

Siebel CRM OnDemand is a hosted CRM product, using IBM e-business hosting. The advantage of this type of product is that it eliminates software installations, hardware, upgrade and support cists, and maintenance costs.

A disadvantage to this type of strategy could again be cost. The CRM product costs $70 per user, per month. If the company wishes to register ten users, they are looking at a cost of $700 per month. Over a period of one y ear this totals $8400 per year, and the cost is ongoing. A strategy hosted in-house so to speak, could be more cost effective in the long run.

Other technologies that Siebel have used to their advantage include SSL encryption, multi-level firewalls, and statistics (mentioned in more detail in the next section) (Siebel, 2004a).

Forecasting

Siebel offers some forecasting capabilities, but they are focused around predicting quarterly revenue targets rather than customer churn. They have named their forecasting modal, triangulated forecasting. An example can be seen in fig 1.

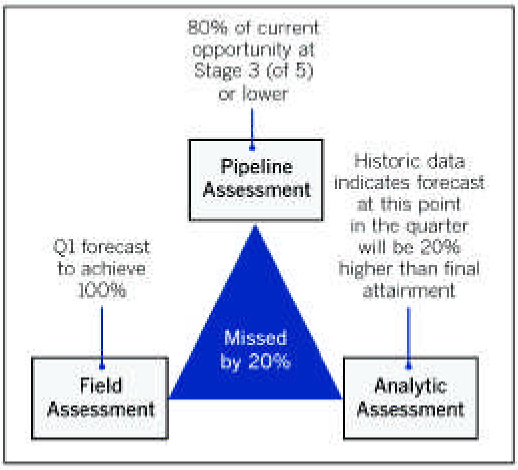


Fig 1 - Siebel's triangulated Forecasting

According to Siebel, organisations should assess forecasts from three perspectives, which is why Siebel likes to refer to the approach as triangulated forecasting. The three perspectives as defined by Siebel are as follows: -

> **The Field Perspective** – A rollup of individual forecasts in order to provide a view of the current market conditions.
>
> **The Pipeline Perspective** – Achieved by analysing opportunities at each stage of the pipeline to provide a sales analysis.
>
> **The historical Perspective** – Achieved by comparing current pipeline data with known historical trends.

It is not clear what type of predictive modal is used in conjunction with this triangulated forecasting system, however it is assumed that the process has been designed from statistics (Siebel, 2004b).

Customisation

The documentation available from Seibel and the fact that the service is hosted on-line suggests that the product isn't customisable. However, Siebel's paper, entitles 'Consulting Services', (Siebel, 2004d) does state that one of the plans available has been called TurnKey packages. TurnKey packages have been designed to meet an organisations need, by the deployment of specific Siebel CRM OnDemand processes.

After Sales Service and Support

Siebel claims to offer world class service and support. The have identified the importance of a fast and reliable response, and offer support in the form of two packages. The standard package promises response within 1 business day, while the Gold package promises response within 4 hours. Full details of both packages can be seen in fig 2 (Siebel, 2004e).

SIEBEL CRM OnDEMAND CUSTOMER CARE OFFERINGS		
	Siebel CRM OnDemand Standard Customer Care	Siebel CRM OnDemand Gold Customer Care
Response Time	1 Business Day	4 Hours
24x7 Web Access	Yes	Yes
Toll-Free Phone Access	12x5* Unlimited Calls	24x7 Unlimited Calls
Service Request Limit	Unlimited	Unlimited
Siebel CRM OnDemand Usage Success Reviews	Not Included	Yes

*Available 6 a.m.–6 p.m. PT, Monday–Friday (excluding holidays)

Fig 2 - Table of Siebel's customer care packages

Training

Siebel offers various training options at no extra cost, which include a range of online tutorials and interactive web courses. If the company wishes to pay for training Siebel can offer a service, customised to an organisations specific processes and terminology, which can be delivered on-site or over the internet (Siebel, 2004c).

Amdocs ClarifyCRM

Amdocs CRM has been designed with large companies in mind, with millions of customers, offering a wide and complex variety of products. Amdocs offers a large range of services focused around customer loyalty and retention.

Amdocs (2004b) states "In order to achieve an integrated approach to customer management companies need CRM solutions that manage the complete customer life-cycle, that enable you to identify, service, retain and grow profitable customer relationships". {Amdocs 2004 #100}

Customer Life-Cycle

Stage 1 – Prospects

Amdocs provides a sales and ordering solution in an attempt to provide the sales team with the strategies needed to make opportunities, however, according to information located in the sales section of Amdocs website, the sales and ordering tools can be applied across every stage of the customer life-cycle. The sales tools provide support for tracking customer information such as opportunities, accounts, contacts, tasks, sales processes and quote information (Amdocs, 2003).

Stage 2 – Responders

This stage of the customer life-cycle is again dealt with by the sales and ordering solution provided by Amdocs.

Stage 3 Established Customers

Although there is evidence that Amdocs provides tools to gain new customers. The literature provided by their website strongly suggests that most emphasis for their product has been placed on customer retention and relationship building. Some of the tools available for this stage of the customer life-cycle are as follows, however this should not be regarded as a complete list: -

> **Opportunities Advisor** – Described by Amdocs (2004c) as "A real-time recommendation and offer optimization engine". The service has been created in an attempt to identify opportunities for cross-selling and up-selling. The opportunities advisor has been designed to keep track of its history, so that customers are targeted with only the offers that are relevant to them and they do not receive repeat offers (Amdocs, 2004c).
>
> **Predictive Analytics** – An application provided by Amdocs to predict when customers might behave in certain ways and why. Using data mining and statistical techniques, the application can make predictions on such things as credit and churn risks. More specifically the predictive analytics application has been developed using PMML (Predictive Modelling Mark-up Language) and XML (Extensible Mark-up Language), which allows the application to share models with other applications that use these technologies.{Amdocs 2004 #190}

Stage 4 Former Customers

Again, it is difficult to identify a great deal of literature associated with this stage of the customer life-cycle. The predictive analytics section of the website, Amdocs

(2004e) states that " Predict which high-value customers are likely to discontinue use of products or services, and gain insight into the root causes to enable proactive customer retention". Taking this into consideration, along with the importance that Amdocs places on churn management, it would be fair to assume that they would have provided a means to keep track of all reasons for churn (Amdocs, 2004d).

Churn Management Tool

As mentioned above Amdocs has developed predictive analytics, which is a predictive modelling and customer analysis application. It allows companies to evaluate the customer behaviour that has an impact on the customer lifetime-cycle.

Churn Prediction

Churn prediction can be done using the predictive analytics application. Amdocs claim that their product is based on award winning data mining and knowledge discovery technology. In an attempt to confirm this the author has identified a paper regarding these technologies (Rosset S and Neumann E, 2003), which was published as part of the Third IEEE International Conference on Data Mining.

The paper discuses the importance of identifying high valued customers, it states that the prediction of a customer with a low value to the company should not bother the business at all. This would mean that calculating a customer's value to the company is an important step in Amdocs predictive analytics. According to Amdocs this calculation is a straight forward process (in context of the telecommunications industry) based on the customers price plan, usage, payments, call centre contacts etc. The value can be calculated from the customers received payments minus the cost of supplying services and products. This value is regarded as an important piece of information for the performance of the churn indicator. The paper than applies its formulas using linear regression (Rosset S and Neumann E, 2003).

Technologies Used

Amdocs have combined various technologies including data warehousing, XML, PMML, statistics and OLAP. To compile a complete list of technologies would be virtually impossible without inside knowledge from Amdocs so the mentioned technologies are those which have been mentioned within the various sources of literature.

Forecasting

At this point is already clear that Amdocs provide forecasting capabilities. These forecasting capabilities can be applied to four main areas: -

> **Customer Retention –** Predicting the likeliness of a high-valued customer leaving the company.
> **Cross Sell and Up Sell Analysis –** Predict which customers are likely to have an interest in products and services.

Credit Risk and Bad Debt Management – Prediction of potential payment hazards.

Campaign Response Analysis – Predict the customers which are most likely to respond to marketing campaigns.

Amdocs have used flexible sampling techniques to select the data for analysis. Statistical algorithms are then applied to identify relationships the customers attribute. Once created, models can be evaluated, published, and used for scoring campaigns (Amdocs, 2004d).

Software Customisation

Amdocs is willing to partner with its customers in order to tailor their CRM solution to the customers specific needs and ensuring that the right CRM solution has been applied.

After Sales and Service and Support

It has been difficult to locate any substantial information about Amdocs after sales service and support. It is believed that the reason for this is because Amdocs offers customer support services as part of their solution for CRM. All searches regarding customer support result in matches regarding the product they provide.

Training

Amdocs offers a variety of different training courses, related and unrelated to their CRM software. This section is concerned with training related to the ClarifyCRM software. The courses offered are as follows (Amdocs, 2004a): -

Product Training
End-User Training and Implementation
User Documentation and Performance Support
Training Consulting and Training Program Management

PeopleSoft CRM

PeopleSoft supplies their product on a world-wide level, claiming to have more than 12,200 customers in 150 countries. The software is again applied via the internet. PeopleSoft regard their product as most suitable for manufacturing, distribution, life science, natural resources, chemicals, construction and real estate. PeopleSoft provides applications tailored for CRM, Human Capital Management, Financial Management and Supply Chain Management. Products have been designed specifically for use with relational database management systems (RDBMS) (PeopleSoft, 2003a).

Customer Life Cycle

The following information has been taken from the media section of PeopleSoft's website, which can be found at

Stage 1 Prospects

PeopleSoft recognise that an understanding of the needs of prospects is important to the success of the sales force. In response to this view, PeopleSoft have tried to develop easy to use tools to ensure that the sales cycle isn't prolonged and information does not get lost.

Another feature offered by PeopleSoft is web based functionality. EnterpriseOne Sales also has a feature, which has been designed in an attempt to capture all relevant information on prospects to support sales staff by providing management with pipeline and revenue information (PeopleSoft, 2004g).

Stage 2 Responders

According to a white paper provided by PeopleSoft (2003b), once a hot lead has been identified, the sales force have to assess the opportunity by quantifying the potential revenue, determining prospects need, and identifying competitive threats. Once this has been done a proposal can be tailored to the prospects requirements. Access to critical data is required for this process in order to successfully close the sale.

Sales people spend a lot of time away from the office. To ensure that the sales force has constant access to up to date, and critical information, PeopleSoft have developed a mobile sales application. This application has been designed to allow the sales force to complete sales cycle tasks without having to be connected to head office (PeopleSoft, 2003c).

Stage 3 Established Customers

In order to help build customer relationships, PeopleSoft have integrated an e-ordering system. This allows customers 24 hour access to purchasing and order status information. The application has been named Customer Self Service. It can recognise each customer providing only information that is specific and relevant to each individual, providing a personalised customer experience.

PeopleSoft have adopted the view that making it easy for a customer to do business with an enterprise allows service agents to focus on more complex issues and build stronger customer relationships (PeopleSoft, 2004c).

Stage 4 Former Customers

Again there is not very much documentation available, which supports this final stage of the customer life-cycle. PeopleSoft offer Customer Behaviour Modelling which combines customer data with behavioural metrics and demographic information from third party providers. This data can be used to predict the customers that are more likely to leave the company. Based on this information it would be fair to assume that PeopleSoft provide support of analysing route causes of churn in the attempt to remedy problem areas (PeopleSoft, 2004b).

Churn Management tool

This section will take a greater look at the Customer Behavioural Modelling feature offered by PeopleSoft. It has been stated that with the information gained from this process it is possible to build predictive models and score customers based on relevant criteria (PeopleSoft, 2004b). The customer behavioural modelling process can be used to: -

Predict a customers response to marketing campaigns
Identify cross-sell and up-sell opportunities
Manage customer attrition
Perform customer valuations

Churn Prediction

Churn prediction can be done using the Customer Behavioural Modelling feature. It is performed by the use of data mining techniques and predictive analytics. Unfortunately the author has failed to identify any in-depth information regarding this feature, only that it supports predicting the likelihood of a customer leaving the company (PeopleSoft, 2004b).

Technologies Used

PeopleSoft uses an open-standard based, pre-integrated software infrastructure. There are two versions of the software available: -

A Browser Based Version
Windows Client Based Version

PeopleSoft have allowed for interoperability by offering solutions, which include EDI, MQseries and MSMQ (PeopleSoft, 2004a).

Forecasting

The predictive analytics, as mentioned in the churn management section, can be used to create a variety of predictive models. These models include: -

The customer that are most likely to purchase certain products
Customer segmentation

Customer churn

Rather than a tool to replace statisticians, PeopleSoft have created this feature with the intention to free a statisticians time so that it can be focused on more important tasks. Certain templates are provided for such tasks as predicting churn and campaign responses (PeopleSoft, 2004f).

Software Customisation

PeopleSoft can provide an application designer, which is an integrated development environment that allows the user to customise all of PeopleSoft Enterprise's Applications. It is in fact the same tool that PeopleSoft developers use to create the PeopleSoft Enterprise applications. The tools allow for interoperability with, and reuse of business logic, written in Java, VB, or C++ (PeopleSoft, 2004e).

After Sales Service and Support

PeopleSoft offer three levels of support depending on the company's requirements. The options available are as follows: -

Standard Support – Provides rights to future software releases and product features, 24/7 real time support, indefinite technical support for as long as a company holds a valid software licence.

Premium Support – In addition to the standard support features, this option provides priority product and technical support and priority response.

Platinum Support – designed for complex multinational companies that whish to join forces with PeopleSoft. This level of support offers a dedicated PeopleSoft account, team led by a service director. Priority problem resolution and Opportunities to influence future product enhancements.

With a companies annual paid support subscription, PeopleSoft's technical support includes access to problem solving assistance, case management, existing fixes and updates and tips and techniques (PeopleSoft, 2004h).

Training

PeopleSoft offer a variety of different training programs. The levels of training available are as follows: -

Management education – Designed to provide management with a better understanding of technical, analytic, planning and strategic implementations of PeopleSoft.

Project team education – Designed to provide the project team with technical skills for more effective implementation and upgrades.

End-user education – Designed to provide end-users with the skills they need to increase productivity, accuracy and build more confidence.

There are also web-based courses that are available to anyone. They have been designed to provide an overview of the most popular PeopleSoft Enterprise products (PeopleSoft, 2004d).

MySap CRM

MySap Customer relationship management belongs to a larger suite of products provide by SAP called MySap Business suite. SAP claim that MySap CRM provides industry specific features and best practices, based on 30 years of knowledge (SAP, 2003c)(.

Customer Life-Cycle

MySap CRM claim to provide insight into the entire customer life-cycle including developing strategic sales plans for new opportunities and identifying the strongest marketing leads (SAP, 2003e).

Stage 1 Prospects

MySap Sales has been designed to increase the productivity of the sales reps. SAP claim that it guarantees consistency, effectiveness and predictability in order to make the most promising leads MySap offers two solutions to the sales team, Enterprise Sales is a web based portal for sales reps and managers, working within the enterprise. Mobile Sales has been designed to work via laptops, tablet PC's and mobile phones. It provides sales reps, working in non-connected environments with instant access to customer data. MySap Sales has been designed with the following sales channels in mind.(SAP, 2003e): -

Direct Sales
Telesales
Channel Sales
E-selling

Stage 2 Responders

MySap Sales has been developed to identify the strongest market leads and generating customised quotes and offerings to try to ensure that the most valuable customers are identified and the sale finalised. Some of the key features that support this stage of the customer life-cycle are as follows (SAP, 2003e): -

Organisational and territory management – Allows managers to define territories based on product lines, products, geography allowing them to identify prospects and products associated with each area.

Accounts and contacts manager – Helps monitor and store information about customers, prospects and partners.

Opportunities manager – Providing sales reps with information about each sales opportunity.

Stage 3 – Established Customers

MySap CRM focuses on customer satisfaction. SAP understands that industries are competitive, and as a result customer's expectations are rising. In order to maintain good relationships with the customer SAP have developed their product to achieve the following customer centric objectives (SAP, 2003f): -

Customer service and support – Helps to manage and fulfil customer commitments, by providing support for quickly resolving customer issues.

Service planning and forecasting – Helps create service plans to maintain products for optimum performance and provides forecasting of planned services to distribute service resources to help meet customer requirements.

Resource Planning – Helps to plan long term resource strategies and optimise short term scheduling tactics.

Field service – Allows the field service reps to access the enterprise remotely.

E-Service – A web-based self service solution that allows customers to access the enterprise in order to research and diagnose their own problems.

Stage 4 – Former Customers

MySap CRM uses analytics to provide information on why sales and marketing efforts failed to work in the past, the reasons behind poor retention rates and analyse unpredictable customers. Having the answers to these problems should provide the company with the knowledge on how to improve them in the future (SAP, 2003d).

Churn Management Tool

The author has failed to find any information regarding a specific churn management tool, however customer retention efforts can be done using the analytics software as mentioned in the customer life-cycle section, stage 4.

Churn Prediction

Churn prediction can again be accomplished using the analytics. Analytic scenarios are pre-built into MySap CRM these scenarios include Customer analytics. The analytical scenario built, has been designed to gain a better understanding of customer preferences. Value analysis can be performed to gain the profitability and lifetime value of each customer (SAP, 2003d).

Technologies Used

MySap have used Microsoft .NET and Java technologies along with web based technologies such as HTML and XML. These technologies are referenced in MySap documentation for compatibility with customer technologies, however it is not known which technologies were implemented as part of the main development of MySap CRM (SAP, 2003g).

Forecasting

Using MySap, forecasting can be performed in various different areas.

Customisation

Customisation can be performed using a technology built by SAP which they call SAP NetWeaver. SAP NetWeaver is an open integration and application platform that can be used to design build and execute new business strategies and processes. It is compatible with internet standards such as HTML, XML and Web Services, and is interoperable with Microsoft .NET and Java 2 Enterprise Edition environments (SAP, 2003g).

After Sales Service

SAP offer a variety of customer support packages available from the initial stages of implementing SAP to an industry, through to customer support once SAP is up and running. The two main support packages available are as follows (SAP, 2003a): -

> **SAP Standard Support** – Reduces total cost of ownership. This level of support provides all the knowledge, tools and functions needed to implement, manage and enhance SAP.
>
> **SAP MaxAttention –** A premium level support service which includes on-site assistance. All features of the standard support package are included with additional features such as 24/7 support for high priority use. SAP claim to respond to 95% of messages within 2 hours.

Training

SAP offers a variety of educational services that have been designed to maximise a company's return on their investment in SAP solutions. Training is offered across the board to ensure that the IT professionals, und-users and decision makers possess the knowledge required to effectively carry out there roles using the SAP product. Training facilities have been set up in over 50 countries worldwide offering multilingual services (SAP , 2003b).

Oracle CRM

Oracle offers their CRM in the form of Oracle e-business suite. The focus of e-business suite is lowering the cost of ownership and understanding, and responding to customers issuers. Oracle claims that e-business suite offers strategies for identifying and fulfilling customer requirements, and offers a wide range of implementation resources, in an attempt to enrich the customer experience (Oracle, 2003).

Customer life-Cycle

Stage 1 – Prospects

This stage of the customer life-cycle is focussed on making leads in order to try to gain more customers. In order to achieve this, Oracle provides marketing tools. Oracle says that their marketing tools drive profit, not just responses, and the strategy has been designed to target the most profitable customers. Marketing campaigns can be refined in real-time, using analytical tools. Oracle marketing can be integrated with other Oracle E-Business tools such as sales, finance and the rest of the organisation (Oracle, 2004c).

Stage 2 – Responders

In order to turn responders into customers, oracle provides Sales tools. According to Oracle, the sales tools go further than just administrating transactions. They have been designed to aid the sales team by providing more information about the opportunities, by incorporating analytical functionality. Oracle Sales has been designed to allow the sales team to meet challenging revenue goals, using fewer resources. The key benefits of Oracle Sales are as follows (Oracle, 2004f): -

Providing comprehensive information about customers and prospects.
Focus on the opportunities that provide the best potential for profit.
Lower costs for software implementation and maintenance.

Stage 3 – Established Customers

Oracle E-business has been created in an attempt to enrich the customer experience. It could be argued that every aspect of this product has some focus on improving business to customer relationships. However established customers would most likely be more interested in the service they receive when doing business with the organisation. Oracle includes ordering tools, which have been designed to: -

Provide accurate order promise dates.
Respond quickly to changes in the customers needs.
Execute customer shipment accurately and on time.

The above key features are essential in building customer relationships, however from the business point of view, Oracle Sales improve visibility into customer demand, across all channels (Oracle, 2004d).

Stage 4 – Former Customers

Oracle can offer analytical capabilities but it is not clear if these are provided as part of the E-Business suite, or part of other Oracle products such as databases. Again, if Oracle does provide support for reasons associated with churn, it is not very well documented. The authors personal view is, if this support is available using Oracle, it would be included in Oracles analytic tools. The only documentation found for analytics is located in the section of the web site on business intelligence, which appears to be separate from Oracles E-business and require an additional purchase (Oracle, 2004a).

Churn Management Tool

As mentioned in the previous section (3.5.1 Customer Life-Cycle, Stage 4) the author has found no real evidence supporting that a churn management tool has been included with Oracle's E-Business suite, however churn management functionality may be possible with Oracle's analytical tools. Oracle's analytical tools include data mining products and service's that can be used to find hidden trends and patterns associated with customer data ((Oracle, 2004a).

Churn Prediction

If churn prediction is possible with Oracle, again it will be from using the analytic components.

Technologies Used

Oracle is one of the leaders in Java standards so a fair assumption would be that java is the core technology used for development. Also many oracle products are available on various platforms, such as Microsoft Windows, Linux and UNIX, to achieve this Java would be the most likely language. For Oracle's analytics, technologies such as OLAP and data mining are used (Oracle, 2004e).{Oracle 2004 #520}

Forecasting

Forecasting would again be part o Oracle's analytic services.

Customisation

Oracle has designed their product based on constant customer feedback. For implementing Oracle E-Business suite, Oracle has provided a set of tools to automate the configuration. One such tool has been named *i*Setup which allows the user to quickly set up critical business processes using web based questionnaires (Oracle, 2003).

After Sales Service and Support

Oracle claim to provide award winning after sales service and support and provide a team of on-site consultants, on-line assistance, 24/7 phone help and IT Mangement on demand. The key components of Oracle's sales and support services are as follows (Oracle, 2004g): -

Metalink – 24/7 access to online technical support.
OTN – Online services and resources to help build, test and deploy Oracle applications and resources.
Ask Tom – Ask questions to Oracle expert Tom Kyte
Live Support Seminar – For working with support and advanced metalink.

Training

Oracle offers a huge list of training services for all of there products. The key topics that have training associated with them, and are relevant to Oracle's E-Business suite are as follows (Oracle, 2004b): -

Technology
Financials
Procurement
Order Fulfilment
Projects
Advanced Panning
Manufacturing
Human Resources
Sales
Contracts
Services
E-commerce
Business Intelligence

Microsoft CRM

Microsoft has developed their CRM solution as a means to meet a precise organisations needs and help build more profitable customer relationships. Microsoft claim that their product is easy to use, enabling management and other employees to make informed decisions, increase sales, and provide exceptional customer service (Microsoft, 2003).

Customer Life-Cycle

Microsoft CRM claims to help an organisation increase sales success, deliver superior customer service and make informed, responsive business decisions. It is reported as being easy to use with the ability to customise and grow with a business.

Stage 1 - Prospects

For this initial stage of the customer life-cycle, Microsoft CRM offers a leads and opportunities manager. The leads and sales manager is reported to shorten the sales cycle and provide detailed reports, displaying information on sales and support activity and history, attempting to identify the opportunities, trends and problems that can affect decision making. Information can be tracked on prospective customers. (Microsoft, 2003).

Stage 2 – Responders

Microsoft's sales process management is a feature that has been included to help the sales force initiate, track and close sales opportunities. It includes work flow rules that automate the stages of the selling process. Another feature that can help with this stage of the customer life-cycle is an opportunities manager. The opportunities manager can help to convert qualified leads into opportunities without the need for data re-entry. The leads can then be tracked throughout the sales cycle (Microsoft, 2002b).

Stage 3 - Established Customers

As with the other all of the other CRM vendors that have been analysed in this report, the main focus behind Microsoft's CRM is customer relationship building. One of the main features of the product that has been provided to help with customer relationship building is customer services. Customer services offer increased capacity to help handle requests. This feature helps the service representatives track customer requests, manage support issues from start to finish, and provide customers with an efficient, consistent service, which Microsoft believes is the key to ensuring customer satisfaction (Microsoft, 2002a).{Microsoft 2002 #570}

Stage 4 – Former Customers

The author has found no evidence supporting any tools that can aid the business in identifying key issues that lead to customer attrition. The focus on Microsoft's CRM appears to be customer satisfaction rather than churn management. At this stage it appears as though this is an unsupported issue.

Churn Management Tool

The author has found no evidence of a churn management tool, provided with the Microsoft CRM product.

Churn Prediction

The author has been unable to locate any churn prediction or analytical support for Microsoft CRM.

Technologies Used

Microsoft CRM has been developed using .NET technology, and offers integration with Microsoft Outlook and Microsoft Office. The .NET framework enables users to access Microsoft CRM through a web client running in Microsoft Internet Explorer and through a Microsoft Outlook client. It also supports other open standards, such as XML and WSDL (web service description language) (Microsoft, 2003).

Forecasting

A certain amount of forecasting can be achieved using the reporting tools available from Microsoft CRM such as forecasting future sales performance. Microsoft CRM reports can also be imported into Microsoft's Office Excel, suggesting that some manually created forecasting algorithms could be applied to the data (Microsoft, 2003).

Customisation

Microsoft claims that their CRM product offers sufficient customisation processes, which reduce cost, and offer high ROI. Some of the customisation features available are as follows: -

- **Configurable workflow rules** – designed to automate existing business processes.
- **Tailored forms** – to capture the data needed to close sales
- **In-house functionality enhancements** – Allowing developers to build customised vertical solutions, create integration with third party applications, and extend the solution to web service platforms.
- **Easy customisation implementation** – Because Microsoft CRM is browser based, customisation can be published throughout the organisation in one upload.

After Sales Service and Support

Microsoft claim to offer an award winning support service. Support professional are said to handle most technical support requests immediately, and for the ones that are unable to be resolved immediately, Microsoft guarantees a response time. Self support options are available 24/7 online. There are several support options as follows (Microsoft, 2004b): -

- **Premium Services** – Microsoft's most proactive and personalised business plan. Premium services are offered in two forms: -
  - **Premium Enterprise**
    - 200 hours of service from assigned technical account manager
    - Unlimited telephone and electronic support
    - 1 hour guarantee response time on support requests
    - Emergency access 24/7, 7 days a week, 365 days a year
  - **Premium 100**
    - 100 hours of service from assigned technical account manager
    - 100 telephone or electronic support incidents
    - 1 hour guarantee response time
    - Emergency access 24/7, 7 days a week, 365 days a year
    - Option to purchase additional support incidents
- **Software Assurance –** Automatic enrolment if Microsoft's CRM solution is purchased through Microsoft Volume Licensing. Benefits for this plan scale depending on the product purchased, such as Open, Open Value, Select, or Enterprise Agreement licensing programs.

FPP Maintenance – offers all product updates, the option to move the product suite without having to repurchase.

- Searchable technical database
- eLearning options

Flex 5-Pack Support – incident support provided on an as and when basis

Training

Microsoft offer several training options some of which are available to non CRM customers. The following training options are available (Microsoft, 2004a): -

E-learning – Several courses on offer
Classroom training – Learning form an expert in an interactive environment
Onsite Training – Certified instructor comes to the customers site
Training Materials – designed to compliment classroom and online training
Training From Microsoft Certified Partner – Local partner may offer additional, specific training.

Zero Attrition

Zero attrition call their CRM product eCVM. This is a term that was created by them meaning Enterprise Customer Value Management. eCVM combines elements of customer relationship management, Predictive customer relationship management and risk management (Zero Attrition, 2002a).

Customer Life-Cycle

Zero Attritions product is focused around ensuring that customers are treated consistently and in accordance with their current value throughout the customer life-cycle (Zero Attrition, 2002a).

Stage 1 – Prospects

Zero Attrition claim to cater for every stage of the customer life-cycle, however no documentation can be located related to this initial stage. It has been assumed that Zero Attrition's CRM software is applicable from the responder stage of the customer life-cycle, and offer no marketing product to help identify prospective customers.

Stage 2 - Responders

Zero Attrition has three products in there eCVM suite. Each product has a different job. They should be used at different stages of the customer life-cycle. The product associated with this stage of the customer life-cycle is called cRisk. According to Zero Attritions documentation, cRisk has been designed for use at the beginning of the customer life-cycle when a company is thinking about acquiring a customer. Using multiple data sources, segmentation, predictive modelling and multiple decision features, cRisk can assess the credit risk, profit potential and attrition potential for each customer (Zero Attrition, 2002e).

Stage 3 – Established Customers

Zero Attrition state that a lot of companies often assess customer value at the beginning of their business relationship with a customer, but don't change the customer's value as the relationship between them and the business progresses. In response to this problem, Zero Attrition supply a product named cValue, which continually monitors, tracks and predicts a customers behaviour to keep a precise record of the value of the customer, to ensure that they always receive a service that is fitting to their value. cValue offers benefits in either real-time or batch modes, these benefits are as follows (Zero Attrition, 2002f): -

Real-Time
- Individual Customer Migration Management
- Manages Individual Customer Credit
- Optimises Individual Cross-Selling

Batch Mode
- Manages Portfolio Migration
- Manages Portfolio Credit
- Optimises Portfolio Cross-Selling

Stage 4 – Former Customers

For this final stage of the customer life-cycle, Zero Attrition offers a product called cRecovery. cRecovery can predict when a customer is on the verge of churning either voluntarily or involuntarily. cRecovery then assess the future profit and attrition potential of that customer and the likelihood of the company receiving payment to establish if the customer is worth keeping or not. cRecovery can also provide steps for improving the companies chances of getting paid. The key points for Zero Attritions cRecovery tool are as follows (Zero Attrition, 2002c): -

Real-time
- Manages Individual Customer Migration
- Analyses Customer Propensity to Pay
- Individual Cross Selling Analysis

Batch Mode
- Manages Portfolio Customer Migration
- Analyses Portfolio's Propensity to Pay
- Optimises Portfolio Cross-Selling
- Manages Attrition

Churn Management Tool

Churn management is dealt with by Zero Attritions cRecovery tool. As mentioned above (3.7.1, Customer Life-Cycle, Stage 4) when the cRecovery tool is running in batch mode it manages attrition. Unprofitable customers are identified and decisions are delivered regarding the value of the customer, hence the efforts that should be made regarding persuading the customer to stay with the company (Zero Attrition, 2002c).

Churn Prediction

Again churn prediction is offered by the cRecovery tool. The tool makes an analysis of the likelihood of a particular customer leaving the company and provides financial information on that customer so that the company can make a decision on weather or not the customer is worth fighting for (Zero Attrition, 2002c).

Technologies Used

There is little documentation about the technologies used by Zero Attrition, however it is stated that all products are delivered via ASP, which would suggest that development has been done by use of either a web based or .NET technology. Zero Attrition believes that ASP delivery is beneficial to a company for the following reasons (Zero Attrition, 2002d): -

Fast and Easy to Implement in Any Environment
Low Cost of Use
No Technical Maintenance
Secure and Reliable
Products Are Always Up-To-Date
Products Are Fitted to the Companies Needs

Forecasting

It would appear as though Zero Attrition offers certain levels of forecasting because as mentioned already, the cRecovery tool offers churn prediction capabilities and a section from the Zero Attrition web-site states that a key feature of the Zero Attrition Suite is the ability to perform 'What-If' analysis, however no more information than this can be located (Zero Attrition, 2002b).

Customisation

No documentation can be found on customisation of the product suggesting that the tool is provided with a standard interface.

After Sales Service and Support

No documentation can be founds regarding after sales service and support, however in the case of this feature, it is assumed that after sales service and support is offered, but not documented. It is possible that this is a is an issue that would be discussed with a company by the sales representative.

Training

The author has also had difficulty in locating training about the services and products offered by Zero Attrition. Again, it is assumed that a certain level of training would be available, however it appears as though Zero Attrition has failed to document it.

SPSS

According to the Two Crows Corporation, the purpose of a CRM product is to aid a company in improving the profitability of interactions with its customers, while at the same time making the interaction appear friendly through individualisation. It also states that companies need to match products and marketing campaigns to prospects and customers, to intelligently manage the customer life-cycle. This information has been taken from a white paper, written by the Two Crows Corporation in conjunction with SPSS (Two Crows Corporation, 2000).

Customer Life-Cycle

Stage 1 – Prospects

SPSS say that marketing campaigns that use to take days to create and execute can be done within a few hours using their Predictive Marketing tools. The product is reported to support each step of the campaign process, enabling the marketing department to better target campaigns to achieve better results in less time (SPSS, 2004d). Little information can be located on the SPSS regarding sales and marketing, however this maybe because they have such an extensive website, containing a huge amount of information, it could be that there is information regarding these processes, but it is lost amongst the masses of data available about the other services on offer.

Stage 2 – Prospects

As mentioned above, little information can be located that is specifically aimed towards the first two stages of the customer life-cycle. An assumption has had to be made that this second stage of the cycle would be performed using the same tool that is used for stage 1. The Predictive Marketing tool could be used to perform an analysis of prospect data, so that they could be targeted with the most suitable product in order to achieve full possibilities of that customer doing business with the company.

Stage 3 – Established Customers

At this stage of the customer life-cycle, SPSS offer a whole range of products that can aid in customer satisfaction. A brief summary of these products is as follows: -

> **Text Mining –** According to SPSS a large amount of valuable data can be found in text associated with e-mails, call centre transcripts and other customer communication. SPSS's Text Mining solution is aimed at retrieving the valuable information from these text files to provide a better overall view of the company (SPSS, 2004c).

Answer Tree – Another data mining technique aimed at empowering a company to target the right groups of customers with specific promotions and offers, to gain maximum results (SPSS, 2004g).

Data Distiller – An application that is again targeted at offering the right promotions to the right customers. Works by combining customer data with campaign information and business rules (SPSS, 2004e).

Web Analytics – Web measurements or e-metrics to provide information about web activities and customer behaviour (SPSS, 2004b).

Predictive Call Center – SPSS say that customers only contact a call centre if they have a specific need or complaint. Predictive Call Centre has been designed to turn these inbound calls into sales opportunities by predicting customers needs and preferences in real time (SPSS, 2004i).

The five mentioned products are just a sample of the tools that are available from SPSS for customer relationship building. Mentioning all available products and performing a deep analysis of each would be unpractical for the purpose of this report. Further reading can be done by visiting http://www.spss.com/products/.

Stage 4 – Former Customers

As can be seen from the products already mentioned, SPSS software incorporates many predictive features. Several of the products can be used for churn analysis, such as the Predictive Call Centre application. SPSS suggest that the call centre could be the businesses last chance to retain a customer, having this information available as the call arrives could be valuable in preventing that customer from churning (SPSS, 2004i).

Churn Management Tool

There is a product available through SPSS called LexiQuest. LexiQuest is a solution that SPSS says is based on 25 years of research. It is a solution similar to text mining, only it analyses text not as a collection of words or letters, but as a collection phrases and paragraphs. LexiQuest is meant for a variety of purposes, one being churn management. SPSS say that LexiQuest can be used to strengthen customer satisfaction, and minimise churn (SPSS, 2004j).

Churn Prediction

SPSS offer various data mining solutions, however LexiQuest, as mentioned in 3.8.2 appears to be a good example of a churn prediction solution. More information on LexiQuest can be found at http://www.spss.com/lexiquest/.

Technologies Used

SPSS have made use of a wide variety of technologies, the most well documented ones are the predictive models that SPSS have encorporated into many of their products. These predictive technologies will be summarised in the next section (see 3.8.5, Forecasting).

Forecasting

SPSS has taken advantage of many of the available statistical forecasting methods available. An example of some of the forecasting that can be done using SPSS is as follows (SPSS, 2004a): -

> **Estimate Nonlinear Equations** – Aimed towards companies that has models with nonlinear relationships. Purpose is for such tasks as predicting coupon redemption as a function of time etc.
>
> **Fit Latent Variable Models** – Used to test and confirm hidden variables in the data.

SPSS recommend other techniques for numerical predictions, which have also been used in various SPSS models, these techniques are as follows: -

> **Linear Regression –** To explore the relationships between predictors and what is required to predict.
>
> **Weighted Least Squares Regression –** This is appropriate when the variance in the dependant variable isn't constant.
>
> **Two Stage Least Squares –** Used for when the predictor and the outcome have reciprocal effects.
>
> **Survival Analysis Procedures –** Examine the time to an event when a second event hasn't been recorded. The example given in the paper is customers who are still loyal.

Customisation

No real information has been found regarding customisation. It is unknown if user interfaces and services can be customised to match the business requirements at this stage.

After Sales Service and Support

SPSS has a full customer support web site which offers customisation to the customer (login required). The website has a search bar which can be used to search the SPSS technical support database, containing over 7000 resolutions based on customer questions and common technical problems. It also has links for downloading patches and utilities and enhancements. No mention has been found from this website regarding other technical support services (SPSS, 2004h).

Training

SPSS offer training courses worldwide, on a wide variety of services. The general training categories are listed as follows. Each of the training categories has a number of different courses available under them: -

Training Packages
- Three courses to choose from

Getting Started With SPSS
- Four courses to choose from

Analysis and Reporting Courses
- Twelve courses to choose from

Survey Courses
- Four courses to choose from

Application Development Courses
- Four courses to choose from

Customer Analytics Courses
- Three courses to choose from

Business Intelligence Courses
- Six courses to choose from

Web Analytics Courses
- Two Coursers to choose from

It should be noted that these are the courses that are available in the UK. Courses offered by SPSS in other countries may differ (SPSS, 2004f).

CustomerSat

CustomerSat is another CRM vendor that offers hosted online solutions. The service has been developed in an attempt to enable a company to act quickly to recognise and save at-risk customers, better prioritise corporate investments and optimise overall business performance (CustomerSAT, 2003).

Customer Life-Cycle

Stage 1 – Prospects

One of the ways that CustomerSAT has achieved support for this initial stage of the customer life-cycle through an application that they call WEBconnect. WEBconnect allows a company to capture feedback from new customers seeking information about a particular service or product. WEBconnect delivers a focused popup in an attempt to capture information about the user's attitudes and contact information. An example of the type of popup the users may receive can be seen in fig 3 (CustomerSAT, 2004c).

fig 3 CustomerSat WEBconnect

Stage 2 – Responders

In response to this stage of the customer life-cycle CustomerSAT has developed SalesConnect. SalesConnect has been developed to identify the prospects in the system, accelerate the sales cycle and deliver messages that resonate with prospects. It also offers the power to qualify prospects and obtain useful information from the wins and losses, providing a company with insight about why a campaign was successful or unsuccessful. Internet based processes are used to qualify all leads in the suspect database. The key features of this software are as follows (CustomerSAT, 2004f): -

Automatically move leads from suspect to prospect, hot prospect or removal
Instantly alert sales directors of hot prospects for immediate action
Discovery of new sales potentials
Sharpen sales and marketing messages

Stage 3 – Established Customers

For this stage of the customer life-cycle CustomerSAT offer eFinanceConnect. This service has been developed in an attempt to offer the following benefits (CustomerSAT, 2004d): -

Measure and build customer satisfaction with all online transactions
Increase customer usage of online services
Improve customer retention
Improve up-sell opportunities
Accelerate online services ROI

Stage 4 – Former Customers

CustomerSAT offer a certain level of support for the final stage of the customer life-cycle in the form of an analytical tool named 'Apostle' Modelling. This is segmentation modelling that segments customers into groups depending on their level of satisfaction and purchase intentions (CustomerSAT, 2004a).

Churn Management Tool

Churn management is achieved using the 'Apostle' modelling tool. As mentioned, it segments customers into various groups depending on levels of loyalty. They have taken an interesting approach to segmentation in this area. The example that they give of groups are as follows (CustomerSAT, 2004a): -

Loyal Customers
Defectors
Hostages
Etc

Churn Prediction

As can be seen from section 3.9.2 Churn Management Tool, segmenting customer into these groups is a means of predicting at risk customers. There are obviously many benefits into segmenting customer according to their level of loyalty (CustomerSAT, 2004a).

Technologies Used

CustomerSAT uses web based technologies in order to deliver their services via the internet. There does not appear to be very much documentation regarding technologies but an assumption could be made that CustomerSAT uses Data Mining techniques and statistics etc (CustomerSAT, 2004b).

Forecasting

CustomerSAT offer a variety of different forecasting services. The list is as follows (CustomerSAT, 2004a): -

Customer Value Analysis and Management – An analytical service that provides information on market perceived quality and service relative to competitors. This service results in analysis and recommendations.

Balanced Scorecard – defines strategy focused metrics including measures of financial performance including customer loyalty and workforce commitment

Segmentation Modelling – Offers the ability to segment customers by such categories as profitability, loyalty, demographics, technographics etc.

'Apostle' Modelling- As already mentioned, this service offers the ability to segment customers depending on their levels of customer loyalty.

Customisation

CustomerSAT provide a service which offers in-depth expertise in customisation programming and other areas. CustomerSAT has designed a six-stage methodology and approach for designing and implementing enterprise-wide feedback solutions (CustomerSAT, 2004e).

After Sales Service and Support

CustomerSAT's goal for technical support is to strengthen their clients overall satisfaction, and reinforce relationships. The key components of CustomerSAT's customer support are as follows: -

Fast acknowledgement, relief and resolution of technical issues
Effective management of service requests.
Fast response to client questions
Pro-active notification of scheduled maintenance
Documentation of and training for new features
Fast response to training requests

Services are provided via phone, e-mail and postings on client portals. Should a project manager be unable to answer a specific question, it is escalated to the appropriate specialist. There are two forms of technical support on offer, standard support and extended support, the key benefits for these are detailed as follows (CustomerSAT, 2004g): -

Standard Support
- Provided during normal business hours
- Available to all CustomerSAT customers at no extra cost

Extended Support
- Includes all benefits of standard support
- Extended hours – 24/7 365 days a year
- Access via pager for critical issues when out of hours
- Optional and ordered separately

Training

CustomerSAT offer a variety of training courses aimed at maximising the value that users gain from CustomerSAT products and to provide the skills needed for effective for effective leadership and management of enterprise-wide customer and workforce feedback programs. Training courses are offered online, at the company's site, at regional events and at CustomerSAT headquarters. The following courses are offered (CustomerSAT, 2004h): -

Enterprise-wide feedback system design and management
User of CustomerSAT ECEM systems
Action Management
Advanced online analysis and reporting
Introduction to online analysis and reporting
Measuring the ROI of real-time feedback systems

SAS

SAS understand that business is harsh. It is this fact that has lead to the growth of CRM vendors. SAS (2004) says "organisations turn towards business intelligence

(BI) applications in the hopes of extracting greater insights from all the data generated by operational and transactional systems". The same paper then goes on to add "even after acquiring conventional BI, true competitive differentiation often remains elusive. The problem is what many software vendors call 'business intelligence' is simply query and reporting software with a thin veneer of so-called 'analytics'". Bearing this statement from SAS in mind the following analysis of their product should prove very interesting.

Customer Life Cycle

Stage 1 – Prospects

SAS provide a CRM solution for cross-selling and up-selling which has been targeted, not only at existing customers but also acquiring new ones by focusing on improving campaign result rates.

In an attempt to identify the most valuable potential customers, SAS also offer credit risking facilities. The credit risking software has been designed to quickly evaluate both existing and potential customers to ensure that firstly, the company isn't being over precautious by declining products and services to customers who have ever intention of paying, and secondly, to ensure the company isn't to lenient by providing products and services to customers that are genuinely high risk. It is important that a business gets the right balance with credit management and SAS feel that they have achieved this equilibrium.

Stage 2 - Responders

Once responders have been identified, a business would be able to target those with the potential of having the most valuable prospects, using the credit management software and target these with a proactive marketing campaign to maximise the businesses chances of converting these to established customers.

Marketing is an important step for a business, and SAS offer two marketing tools to help an organisation gain maximum selling potential. The two products on offer as follows: -

> **Marketing Automation –** offering segmentation and communication services aimed at allowing a business ease with managing sophisticated and personalised customer communication.
>
> **Marketing Optimisation** – This part of the SAS marketing software has been developed to offer extensive reporting capabilities, combining technology, methodology and the industry expertise needed to optimise customer communication based on and individual company's constraints.

Stage 3 - Established Customers

Established customers are catered for in most aspects of SAS CRM, from the cross-Sell up-Sell software to credit management to marketing. In addition to these services, SAS also offer score carding capabilities for strategic business analysis. An example of a use for this service would be to analyse business services for identification of the most and least popular sellers.

SAS also offer web analytics in an attempt to increase customer retention by providing a better customer experience and enhancing service quality. The mentioned are just a few of the services offered by SAS for improving customer loyalty, it would be inappropriate for the purpose of this paper to detail every service offered but more information can be found at www.sas.com.

Stage 4 – Former Customers

SAS offer customer retention strategies aimed at giving a company the information needed to identify at risk customers and those customers worth keeping. Using historic data from scenarios of customer churn specific to a business, SAS provides information on the variables that influence customer churn as to identify which customers are about to leave, but more importantly, why?

Churn Management Tool

As mentioned in stage 4 of the customer life-cycle, SAS offer a customer retention tool which identifies those customers who are at risk of moving to a competitor. This customer retention tool offers the following solutions: -

Reporting on which customers are thinking about churning
Information on the foremost factors influencing the decision to leave
Information regarding which customers are actually leaving
Proactive rule-based analysis of account behaviour
An early warning alert system

Churn Prediction

The customer retention tool previously mentioned (3.10.2) offers churn prediction by use of data mining and analytics, consulting services and industry specific data architectures. SAS understand the importance of customer retention and churn prediction. In their churn analysis paper (2001) SAS say "in the telecommunications industry, annual churn rates of 25 – 30 percent are the norm, and carriers at the upper end of this spectrum will get no return on investment on new subscribers". The paper the goes on to state the reason for this as being "It typically takes three years to pay back the cost (approximately $400 in the US and $700 in Europe) of replacing each lost customer with a new one (customer acquisition).

Technologies used

The technologies that SAP has taken advantage of include data mining techniques and analytics. The SAP website contains a huge amount of information about their services and products including a large amount of white papers. Due to time constraints a further search on SAP technologies is not possible.

Forecasting

Forecasting is offered for customer retention purposes as previously covered and also marketing, and selling. A marketing analysis can be performed in order to identify which campaigns work the best and which products sell the best, also forecasting can be performed using the cross-sell up-sell feature of SAS in an attempt to discover which customers would be most likely to purchase which products, in order to gain best results from customers.

Software Customisation

The author has been unable to locate a great deal of information on customisation, however it has been identified that SAS can tailor a CRM solution to suit a specific organisations needs.

After Sales Service and Support

SAS offer customer support at no additional cost. Specifically for the UK, the SAS customer support head-quarters is located in Marlow and consists of nineteen dedicated support consultants, having access to a vast amount of technical consultants and resources. Consultants have been split into two teams: -

The Intelligence Architecture – Specialises in support for the back end functions, installation and the setup of software and configuring and using SAS on a variety of operating systems.

The Business Intelligence – Specialising in the support of front end systems including the analytical areas of SAS and development.

Training

SAS offer a wide variety of training courses which they have designed to allow an organisation to gain the knowledge .leverage and expertise necessary to get the best use out their service. Some of the training courses available cover the following areas: -

Business intelligence
Customer intelligence
Data management and data quality
Data warehousing
Applications development
Data mining and predictive modelling
SAS 9

Table of Analysis

	Customer Life-Cycle				Churn Management Tool	Churn Prediction	Technologies Used	Forecasting
	Stage 1	**Stage 2**	**Stage 3**	**Stage 4**				
Siebel Systems	Lead Management Capabilities	Siebel's Sales Components	Offers Customer Service and Analytical Tools	No substantial Documentatio n Found	No Substantial Documentation Found	No Substantial Documentation Found	Statistics, Web Based Tools	View of Current Market, Pipeline Analysis, Historic Perspective
Amdocs Ltd	Sales and Ordering Software	Sales and Ordering Software	Opportunities Advisor, Predictive Analytics	Predictive Analytics	No Substantial Documentation Found	Predictive Analytics	Data Warehousing, XML, PMML, OLAP	Customer Retent Sales Analysis, C Risk Analysis, Camping Respon Analysis
PeopleSoft	Sales Software	Sales Software	e-Ordering	Customer Behaviour Modelling	No Substantial Documentation Found	Customer Behaviour Modelling	Web Based and Stand Alone Based	Marketing, Custo Segmentation, Customer Churn
SAP	MySap Sales	MySap Sales	Customer Service Tools, Planning and Forecasting Tools, Etc	Analytical Tools	No Substantial Documentation Found	Analytical Tools	.NET, Java, Web Based,	Various Forecast Possibilities
Oracle	Marketing Tools	Sales Tools	Accurate Delivery Dates, Response to Changing Needs, On time Shipment	Separate Product	No Substantial Documentation Found	Separate Product	Java	Separate Produc
Microsoft	Lead and Opportunities Manager	Sales Process Management	Customer Service Tools	No Substantial Documentatio n Found	No Substantial Documentation Found	No Substantial Documentation Found	.Net Technology	Reporting Tools a Microsoft Excel
Zero Attrition	No Substantial Documentation Found	cRisk	cValue	cRecovery	cRecovery	cRecovery	ASP Delivery	Churn Prediction What-If Analysis
SPSS	No Substantial Documentation Found	No Substantial Documentatio n Found	Many Products Available	Predictive Analytics	Predictive Analytics	Predictive Analytics	Predictive Techniques	Many Statistical Forecasts Are Available
CustomerS AT	Web Connect	Sales Connect	eFinanceConnect	Apostle Modeling	Apostle Modelling	Apostle Modelling	Web Based Technologies	Cust Value Analy and Managemen Balanced Score Card, Segmentat Modelling
SAS	Cross-Sell Up-Sell, Credit Man.	Cross-Sell Up-Sell, Credit Man.	Both previous mentioned and web analytics	Customer retention strategies	Customer retention strategies	Customer retention strategies	Data Mining and analytics	Churn prediction Marketing Predic

Key Observations

The main observation that can be made from this report is that most of the CRM vendors do not take churn management and churn prediction seriously. Most have focused primarily on customer satisfaction. Customer satisfaction is obviously an important aspect of customer management, but as the saying goes, you can not please all of the people all of the time. The thought is, no matter what a company does to try and satisfy customers, there will always be those that feel they have a reason to move to a competitor. It is important that these customers are identified before they actually churn, to at least give the company a chance to fight for them.

Some of the vendors do take churn into consideration, however even most of these do not offer a specific churn management tool. It is perceived by the author that a churn management tool would be very valuable for businesses. A company's most profitable customers are its loyal customers. Failing to recognise when a loyal customers is considering churning could be very costly to a business.

Another observation that has been made in this research is how CRM businesses have opted for developing web based services. Web Based services maybe easy to setup and configure, with low maintenance costs, however larger organisations already have a large percentage of the equipment required to run CRM software and engineers qualified to maintain it. It is believed by the author that the real reason behind hosting web based applications is that it is a permanent source of income, through subscription charges. This is most likely a service that is more beneficial to some organisations than others.

Limitations, Future Research and Conclusions

Limitations

This research has several limitations. The first being the fact that there are so many CRM vendors in the field at the moment that it would be virtually impossible to write an analysis that contains information about all of them. That is why, for the purpose of this paper, nine of the most popular vendors were identified through initial research and informal talks with industry professionals, and an analysis offered on only these companies.

The second limitation was the overall lack of any real valuable or technical information that could be used for a thorough analysis of the products. It appears as though companies are unwilling to share certain types of information such as technical information about how they have developed their analytical tools, and information about the technologies that they have used to create their services. Much of the information available on the vendor's web sites, as could probably be predicted, is mainly sales information hyping up the service offered. Some companies provide huge amounts of data making it difficult to identify anything of real value, while other companies provide very limited data.

Future Research

The author believes that future research in this area should be around identifying the methods used by vendors for their analytical tools and identifying the strengths and weaknesses of these techniques. In terms of churn analysis, it would also be interesting to discover which customer variables are used by the companies that do provide customer churn analysis, the estimated accuracy rates for these predictions and if improvements on the accuracy could be made by using alternative technologies and added variables. It is anticipated that the information needed for this proposal of future research would be very difficult to acquire and would most certainly require conversations with developers from the vendors in question.

Conclusion

Customer relationship management and churn management are important issues that should be taken seriously by any organisation if they wish to succeed in business in today's competitive world. The authors recommendation for an organisation planning to buy a solution, would be to ensure that the product they are purchasing in terms of CRM, covers every stage of the customer life-cycle, and that the product also includes a substantial amount of support for customer churn management. Also it is important to test the trial offers to ensure that the tools they offer firstly conform to the documentation provided by that company, and secondly meet all of the organisations requirements.

As there are many vendors offering CRM software, the author recommends that each software should be analysed against the set of criteria before the final decision is made.

Reference List

Amdocs (2003), *Share Prospect Information Anytime, Anywhere,* available at: http://www.amdocs.com/productpage.asp?SID=362&ID=348 (accessed 2004).

Amdocs (2004a), *ClarifyCRM Course Catalog,* available at: http://www.amdocs.com/documents/AMDOCS_CUSTOMER_TRAINING_SERVICES_CATALOG_ClarifyCRM.pdf (accessed 2004a).

Amdocs (2004b), *ClarifyCRM Solutions,* available at: http://www.amdocs.com/documents/ClarifyCRM4Enterprise.pdf (accessed 2004b).

Amdocs (2004c), *Opportunity Advisor,* available at: http://www.amdocs.com/documents/Opportunity_Advisor.pdf (accessed 2004c).

Amdocs (2004d), *Predictive Analytics,* available at: http://www.amdocs.com/documents/predictive_analytics.pdf (accessed 2004d).

Amdocs (2004e), *Predictive Analytics Webpage,* available at: http://www.amdocs.com/productpage.asp?SID=359&ID=353 (accessed 2004e).

CustomerSAT (2003), *Survey Action Management: Saving At Risk Customers and Protecting Recurring Revenue,* available at: http://www.customersat.com/Resources/index2.asp (accessed 2004).

CustomerSAT (2004a), *Analytic Modelling Services,* (accessed 2004a).

CustomerSAT (2004b), *Company Overview,* available at: http://www.customersat.com/About/index.asp (accessed 2004b).

CustomerSAT (2004c), *EFinanceConnect,* available at: http://www.customersat.com/Resources/Data_Sheets/efinanceconnect.asp (accessed 2004c).

CustomerSAT (2004d), *EFinanceConnect,* available at: http://www.customersat.com/Resources/Data_Sheets/efinanceconnect.asp (accessed 2004d).

CustomerSAT (2004e), *Professional Services,* available at: http://www.customersat.com/Services/professional.asp#2 (accessed 2004e).

CustomerSAT (2004f), *SalesConnect,* available at: http://www.customersat.com/Resources/Data_Sheets/salesconnect.asp (accessed 2004f).

CustomerSAT (2004g), *Technical Support,* available at: http://www.customersat.com/Services/technical.asp (accessed 200g).

CustomerSAT (2004h), *Training* , available at: http://www.customersat.com/Services/training.asp (accessed 2004h).

Microsoft (2002a), *Customer Satisfaction,* available at: http://www.mscrm-cabc.co.uk/ (accessed 2004a).

Microsoft (2002b), *Sales,* available at: http://www.mscrm-cabc.co.uk/ (accessed 2004b).

Microsoft (2003), *Microsoft CRM: Brochure,* available at: http://www.mscrm-cabc.co.uk/ (accessed 2004).

Microsoft (2004a), *Microsoft's Business Solutions Training,* available at:

http://www.microsoft.com/BusinessSolutions/training_options.aspx (accessed 2004a).

Microsoft (2004b), *Service Plans for Microsoft Business Solutions CRM,* available at: http://www.microsoft.com/BusinessSolutions/crm_serviceplan.aspx#CRM_Premier (accessed 2004b).

Oracle (2003), *Enhancing the Customer Experience,* available at: http://www.oracle.com/applications/index.html (accessed 2004).

Oracle (2004a), *Advanced Analytics,* available at: http://www.oracle.com/solutions/business_intelligence/sol_analytics.html (accessed 2003a).

Oracle (2004b), *Category Listing,* available at: http://education.oracle.com/web_prod-plq-dad/plsql/show_desc.redirect?redir_type=14 (accessed 215b).

Oracle (2004c), *Marketing,* available at: http://www.oracle.com/applications/marketing/index.html (accessed 2003c).

Oracle (2004d), *Oracle Order Management,* available at: http://www.oracle.com/applications/order_mgmt/intro.html (accessed 2004d).

Oracle (2004e), *Oracle Technologies,* available at: http://www.oracle.com/technologies/index.html (accessed 2004e).

Oracle (2004f), *Sales,* available at: http://www.oracle.com/applications/sales/index.html (accessed 2004f).

Oracle (2004g), *Support,* available at: http://www.oracle.com/support/index.html (accessed 2004g).

PeopleSoft (2003a), *2003 Annual Report,* available at: www.peoplesoft.com (accessed 2004a).

PeopleSoft (2003b), *Giving Sales the Right the Right Information at the Right Time* , available at: http://www.peoplesoft.com/media/en/pdf/white_paper/giving_sales_right_info_wp_0104.pdf (accessed 2004b).

PeopleSoft (2003c), *Improving Sales Operations While Reducing Operational Expenses* , available at: http://www.peoplesoft.com/media/en/pdf/white_paper/improving_sales_ops_wp_1103.pdf (accessed 2004c).

PeopleSoft (2004a), *CRM Technical Information,* available at: http://www.peoplesoft.com/corp/en/products/ent_one/tools_tech/features/index.jsp (accessed 2004a).

PeopleSoft (2004b), *Customer Behaviour Modelling,* available at: http://www.peoplesoft.com/corp/en/products/epm/module/cust_bhvr_mdl.jsp (accessed 2004b).

PeopleSoft (2004c), *Customer Self Service,* available at: http://www.peoplesoft.com/media/en/pdf/datasheet/ent1_crm_self_serve_ds_1203.pdf (accessed 2004c).

PeopleSoft (2004d), *Education Overview,* available at: http://www.peoplesoft.com/corp/en/training/index.jsp?site=corp (accessed 2004d).

PeopleSoft (2004e), *PeopleTools 8.45,* available at: http://www.peoplesoft.com/media/en/pdf/datasheet/ent_tools_tech_ptools_ds_0604.pdf (accessed 2004e).

PeopleSoft (2004f), *Predictive Analytics* , available at:

http://www.peoplesoft.com/corp/en/products/ent/crm/module/predictive_analy.jsp (accessed 2004f).

PeopleSoft (2004g), *Sales Brochure,* available at: http://www.peoplesoft.com/media/en/pdf/datasheet/ent1_crm_sales_ds_0104.pdf (accessed 2004g).

PeopleSoft (2004h), *Support Services,* available at: http://www.peoplesoft.com/media/en/pdf/brochure/support_services/overview.pdf (accessed 2004h).

Rosset S and Neumann E (2003), 'Integrating Customer Value Considerations into Predictive Modeling', *IEEE International Conference on Data Mining,*

SAP (2003a), *Customer Service Network,* available at: http://www.sap.com/services/maintenance/ITservice.asp (accessed 2004a).

SAP (2003b), *Knowledge Is Power, Knowledge Is Productivity,* available at: http://www.sap.com/education/ (accessed 2004b).

SAP (2003c), *MySap: CRM,* available at: http://www.sap.com/solutions/crm/brochures/ (accessed 2004c).

SAP (2003d), *MySap CRM: Analytics,* available at: http://www.sap.com/solutions/crm/brochures/ (accessed 2004d).

SAP (2003e), *MySap CRM: Sales,* available at: http://www.sap.com/solutions/crm/brochures/ (accessed 2004e).

SAP (2003f), *MySap: CRM Service,* available at: http://www.sap.com/solutions/crm/brochures/ (accessed 2004f).

SAP (2003g), *SAP NetWeaver,* available at: http://www.sap.com/solutions/netweaver/ (accessed 2004g).

Siebel (2004a), *Product Description,* available at: http://www.siebel.com/crm/analytics/customer_analytics/product_module_descr.shtm (accessed 2004a).

Siebel (2004b), *Siebel CRM OnDemand Best Practice: Triangulated Forecasting,* available at: http://www.siebel.com/common/includes/pdf_frame.shtm?pdfUrl=http://www.siebel.com/downloads/products/crm_ondemand/crmondemand_forecasting_bp.pdf (accessed 2004b).
Notes: Registration Required

Siebel (2004c), *Siebel CRM OnDemand: Brochure,* available at: http://www.siebel.com/common/includes/pdf_frame.shtm?pdfUrl=http://www.siebel.com/downloads/products/crm_ondemand/crmondemand_br.pdf (accessed 2004c).
Notes: Login Required

Siebel (2004d), *Siebel CRM OnDemand: Consulting Services,* available at: http://www.siebel.com/common/includes/pdf_frame.shtm?pdfUrl=http://www.siebel.com/downloads/products/crm_ondemand/crmondemand_consulting_ds.pdf (accessed 2004d).
Notes: Registration Required

Siebel (2004e), *Siebel CRM OnDemand: Customer Care,* available at: http://www.siebel.com/common/includes/pdf_frame.shtm?pdfUrl=http://www.siebel.com/downloads/products/crm_ondemand/OnDemand_CustomerCare_DS.pdf (accessed 2004e).
Notes: Registartion Required

Siebel (2004f), *Siebel CRM OnDemand: Marketing,* available at: http://www.siebel.com/common/includes/pdf_frame.shtm?pdfUrl=http://www.siebel.com/downloads/products/crm_ondemand/crmondemand_mktg_ds.pdf (accessed 2004f).

Siebel (2004g), *Siebel CRM OnDemand: Sales,* available at: http://www.siebel.com/common/includes/pdf_frame.shtm?pdfUrl=http://www.siebel.com/downloads/products/crm_ondemand/crmondemand_sales_ds.pdf (accessed 2004g).
Notes: Login Required

SPSS (2004a), *End-To-End Software For Statistical Analysis,* available at: http://www.spss.com/products/index.cfm?source=homepage&hpzone=nav_bar (accessed 2004a).

SPSS (2004b), *From Web Data to Business Understanding,* available at: http://www.spss.com/pwa/ (accessed 2004b).

SPSS (2004c), *Improve Customer Relationships With Text Mining,* available at: http://www.spss.com/predictive_text_analytics/ (accessed 2004c).

SPSS (2004d), *Improve Marketing Productivity,* available at: http://www.spss.com/predictivemarketing/marketing_productivity.htm (accessed 2004d).

SPSS (2004e), *Increase Marketing Campaign Revenues,* available at: http://www.spss.com/datadistilleries/ (accessed 2004e).

SPSS (2004f), *SPSS Training,* available at: https://secure.spsstraining.com/index1.html (accessed 2004f).

SPSS (2004g), *Target The Right People More Effectively With AnswerTree,* available at: http://www.spss.com/answertree/ (accessed 2004g).

SPSS (2004h), *Technical Support,* available at: http://support.spss.com/tech/default.asp (accessed 2004h).

SPSS (2004i), *Turn Inbound Calls Into Sales Opportunities,* available at: http://www.spss.com/predictivecallcenter/ (accessed 2004i).

SPSS (2004j), *Unlock The Value in Textual Data,* available at: http://www.spss.com/lexiquest/ (accessed 2004j).

Two Crows Corporation (2000), *Building Profitable Customer Relationships With Data Mining,* available at: http://www.spss.com/new_searchaction.cfm?searchstring=executive%20brief&resultset=1 (accessed 2004).

Zero Attrition (2002a), *A Customer Managemetn Philosophy for the 21st Century,* available at: http://www.zeroattrition.com/ecvm.html (accessed 2004a).

Zero Attrition (2002b), *Profit From Effective Customer Management,* available at: http://www.zeroattrition.com/products.html (accessed 2004b).

Zero Attrition (2002c), *Should We Fight to Keep This Customer?,* available at: http://www.zeroattrition.com/crecovery.html (accessed 2004c).

Zero Attrition (2002d), *What Is ASP Delivery,* available at: http://www.zeroattrition.com/delivery.html (accessed 2004d).

Zero Attrition (2002e), *What Will a Customer Be Worth,* available at: http://www.zeroattrition.com/crisk.html (accessed 2004e).

Zero Attrition (2002f), *Where Does Our Portfolio Stand Today?,* available at: http://www.zeroattrition.com/cvalue.html (accessed 2004f).