What is churn? an overview

Churn is the phenomenon where a customer switches from one service to a competitor's service (Tsai & Chen, 2009: 2). There are two main types of churn, namely voluntary churn and involuntary churn. Voluntary churn is when the customer initiated the service termination. Involuntary churn means the company suspended the customer's service and this is usually because of non-payment or service abuse.

Companies, in various industries, have recently started to realise that their client set is their most valuable asset. Retaining the existing clients is the best marketing strategy. Numerous studies have confirmed this by showing that it is more profitable to keep your existing clients satisfied than to constantly attract new clients (Van Den Poel & Larivière, 2004: 197; Coussement & Van Den Poel, 2008: 313).

According to Van Den Poel and Larivière (2004: 197) successful customer retention has more than just financial benefits:

1. Successful customer retention programs free the organisation to focus on existing customers' needs and the building of relationships.
2. It lowers the need to find new customers with uncertain levels of risk.
3. Long term customers tend to buy more and provide positive advertising through word-of-mouth.
4. The company has better knowledge of long term customers and they are less expensive with lower uncertainty and risk.
5. Customers with longer tenures are less likely to be influenced by competitive marketing strategies.

6. Sales may decrease if customers churn, due to lost opportunities. These customers also need to be replaced, which can cost five to six times more than simply retaining the customer.

1. 1. Growth in Fixed-line Markets

According to Agrawal (2009) the high growth phase in the telecommunications market is over. In the future, wealth in the industry will be split between the companies. Revenues (of telecommunication companies) are declining around the world. Figure 2 shows Telkom's fixed-line customer base and customer growth rate for the previous seven years. The number of lines is used as an estimate for the number of fixed-line customers.

Figure 2-Telkom's fixed-line annual customer base (Idea adopted from Ahn, Han & Lee (2006: 554))

With the lower customer growth worldwide, it is becoming vital to prevent customers from churning.

1. 2. Preventing Customer Churn

The two basic approaches to churn management are divided into untargeted and targeted approaches. Untargeted approaches rely on superior products and mass advertising to decrease churn (Neslin, Gupta, Kamakura, Lu & Mason, 2004: 3).

Targeted approaches rely on identifying customers who are likely to churn and then customising a service plan or incentive to prevent it from happening. Targeted approaches can be further divided into proactive and reactive approaches.

With a proactive approach the company identifies customers who are likely to churn at a future date. These customers are then targeted with incentives or special programs to attempt to retain them.

In a reactive targeted approach the company waits until the customer cancels the account and then offers the customer an incentive (Neslin et al., 2004: 4).

A proactive targeted approach has the advantage of lower incentive costs (because the customer is not " bribed" at the last minute to stay with the company). It also prevents a culture where customers threaten to churn in order to negotiate a better deal with the company (Neslin et al., 2004: 4).

The proactive, targeted approach is dependent on a predictive statistical technique to predict churners with a high accuracy. Otherwise the company's funds may be wasted on unnecessary programs that incorrectly identified customers.

1. 3. Main Churn Predictors

According to Chu, Tsai and Ho (2007: 704) the main contributors to churn in the telecommunications industry are; price, coverage, quality and customer service. Their contributions to churn can be seen from Figure 3.

Figure 3 indicates that the primary reason for churn is price related (47% of the sample). The customer churns because a cheaper service or product is available, through no fault of the company. This means that a perfect retention strategy, based on customer satisfaction, can only prevent 53% of the churners (Chu et al., 2007: 704).

1. 4. Churn Management Framework

Datta, Masand, Mani and Li (2001: 486) proposed a five stage framework for customer churn management (Figure 4).

The first stage is to identify suitable data for the modelling process. The quality of this data is extremely important. Poor data quality can cause large losses in money, time and opportunities (Olson, 2003: 1). It is also important to determine if all the available historical data, or only the most recent data, is going to be used.

The second stage consists of the data semantics problem. It has a direct link with the first stage. In order to complete the first stage successfully, a complete understanding of the data and the variables' information are required. Data quality issues are linked to data semantics because it often influences data interpretation directly. It frequently leads to data misinterpretation (Dasu & Johnson, 2003: 100).

Stage three handles feature selection. Cios, Pedrycz, Swiniarski and Kurgan (2007: 207) define feature selection as " a process of finding a subset of features, from the original set of features forming patterns in a given data set…". It is important to select a sufficient number of diverse features for the modelling phase. Section 5. 5. 3 discusses some of the most important features found in the literature.

Stage four is the predictive model development stage. There are many alternative methods available. Figure 5 shows the number of times a statistical technique was mentioned in the papers the author read. These methods are discussed in detail in Section 6.

The final stage is the model validation process. The goal of this stage is to ensure that the model delivers accurate predictions.

5. 5. 1Stage one – Identify data

Usually a churn indicator flag must be derived in order to define churners. Currently, there exists no standard accepted definition for churn (Attaa, 2009). One of the popular definitions state that a customer is considered churned if the customer had no active products for three consecutive months (Attaa, 2009; Virgin Media, 2009; Orascom Telecom, 2008). Once a target variable is derived, the set of best features (variables) can be determined.

5. 5. 2Stage two – Data semantics

Data semantics is the process of understanding the context of the data. Certain variables are difficult to interpret and must be carefully studied. It is also important to use consistent data definitions in the database. Datta, et al. (2001) claims that this phase is extremely important.

5. 5. 3Stage three – Feature selection

Feature selection is another important stage. The variables selected here are used in the modelling stage. It consists of two phases. Firstly, an initial feature subset is determined. Secondly, the subset is evaluated based on a certain criterion.

Ahn et al. (2006: 554) describe four main types of determinants in churn. These determinants should be included in the initial feature subset.

Customer dissatisfaction is the first determinant of churn mentioned. It is driven by network and call quality. Service failures have also been identified as " triggers" that accelerate churn. Customers who are unhappy can have

an extended negative influence on a company. They can spread negative word-of-month and also appeal to third-party consumer affair bodies (Ahn et al., 2006: 555).

Cost of switching is the second main determinant. Customers maintain their relationships with a company based on one of two reasons: they " have to" stay (constraint) or they " want to" stay (loyalty). Companies can use loyalty programs or membership cards to encourage their customers to " want to" stay (Ahn et al., 2006: 556).

Service usage is the third main determinant. A customer's service usage can broadly be described with minutes of use, frequency of use and total number of distinct numbers used. Service usage is one of the most popular predictors in churn models. It is still unclear if the correlation between churn and service usage is positive or negative (Ahn et al., 2006: 556).

The final main determinant is customer status. According to Ahn et al. (2006: 556), customers seldom churn suddenly from a service provider. Customers are usually suspended for a while due to payment issues, or they decide not to use the service for a while, before they churn.

Wei and Chiu (2002: 105) use length of service and payment method as further possible predictors of churn. Customers with a longer service history are less likely to churn. Customers who authorise direct payment from their bank accounts are also expected to be less likely to churn.

Qi, Zhang, Shu, Li and Ge (2004?: 2) derived different growth rates and number of abnormal fluctuation variables to model churn. Customers with

growing usage are less likely to churn and customers with a high abnormal fluctuation are more likely to churn.

5. 5. 4Stage four – Model development

It is clear from Figure 5 that decision tree models are the most frequently used models. The second most popular technique is logistic regression, followed closely by neural networks and survival analysis. The technique that featured in the least number of papers is discriminant analysis.

Discriminant analysis is a multivariate technique that classifies observations into existing categories. A mathematical function is derived from a set of continuous variables that best discriminates among the set of categories (Meilgaard, Civille & Carr, 1999: 323).

According to Cohen and Cohen (2002: 485) discriminant analysis makes stronger modelling assumptions than logistic regression. These include that the predictor variables must be multivariate normally distributed and the within-group covariance matrix must be homogeneous. These assumptions are rarely met in practice.

According to Harrell (2001: 217) even if these assumptions are met, the results obtained from logistic regression are still as accurate as those obtained from discrimination analysis. Discriminant analysis will, therefore, not be considered.

A neural network is a parallel data processing structure that possesses the ability to learn. The concept is roughly based on the human brain (Hadden, Tiwari, Roy & Ruta, 2006: 2). Most neural networks are based on the

perceptron architecture where a weighted linear combination of inputs is sent through a nonlinear function.

According to de Waal and du Toit (2006: 1) neural networks have been known to offer accurate predictions with difficult interpretations. Understanding the drivers of churn is one of the main goals of churn modelling and, unfortunately, traditional neural networks provide limited understanding of the model.

Yang and Chiu (2007: 319) confirm this by stating that neural networks use an internal weight scheme that doesn't provide any insight into why the solution is valid. It is often called a black-box methodology and neural networks are, therefore, also not considered in this study.

The statistical methodologies used in this study are decision trees, logistic regression and survival analysis. Decision tree modelling is discussed in Section 6. 1, logistic regression in Sections 6. 2 and 6. 3 and survival analysis is discussed in Section 6. 4.

5. 5. 5Stage five – Validation of results
Each modelling technique has its own, specific validation method. To compare the models, accuracy will be used. However, a high accuracy on the training and validation data sets does not automatically result in accurate predictions on the population dataset. It is important to take the impact of oversampling into account. Section 5. 6 discusses oversampling and the adjustments that need to be made.

1. 5. Adjustments for Target Level Imbalances

From Telkom's data it is clear that churn is a rare event of great interest and great value (Gupta, Hanssens, Hardie, Kahn, Kumar, Lin & Sriram, 2006: 152).

If the event is rare, using a sample with the same proportion of events and non-events as the population is not ideal. Assume a decision tree is developed from such a sample and the event rate (x%) is very low. A prediction model could obtain a high accuracy (1-x%) by simply assigning all the cases to the majority level (e. g. predict all customers are non-churners) (Wei & Chiu, 2002: 106). A sample with more balanced levels of the target is required.

Basic sampling methods to decrease the level of class imbalances include under-sampling and over-sampling. Under-sampling eliminates some of the majority-class cases by randomly selecting a lower percentage of them for the sample. Over-sampling duplicates minority-class cases by including a randomly selected case more than once (Burez & Van Den Poel, 2009: 4630).

Under-sampling has the drawback that potentially useful information is unused. Over-sampling has the drawback that it might lead to over-fitting because cases are duplicated. Studies have shown that over-sampling is ineffective at improving the recognition of the minority class (Drummond & Holte, 2003: 8). According to Chen, Liaw & Breiman, (2004: 2) under-sampling has an edge over over-sampling.

However, if the probability of an event (target variable equals one) in the population differs from the probability of an event in the sample, it is

necessary to make adjustments for the prior probabilities. Otherwise the probability of the event will be overestimated. This will lead to score graphs and statistics that are inaccurate or misleading (Georges, 2007: 456).

Therefore, decision-based statistics based on accuracy (or misclassification) misrepresent the model performance on the population. A model developed on this sample will identify more churners than there actually are (high false alarm rate). Without an adjustment for prior probabilities, the estimates for the event will be overestimated.

According to Potts (2001: 72) the accuracy can be adjusted with equation 1. It takes prior probabilities into account.

With:

: the population proportion of non-churners

: the population proportion of churners

: the sample proportion of non-churners

: the sample proportion of churners

: the number of true negatives (number of correctly predicted non-

churners)

: the number of true positives (number of correctly predicted churners)

: the number of instances in the sample

However, accuracy as a model efficiency measure trained on an under-sampled dataset is dependent on the threshold. This threshold is influenced by the class imbalance between the sample and the population (Burez & Van Den Poel, 2009: 4626).