WHEN CHURN HAPPENS IN A MIDDLE EAST TELECOM OPERATOR

João Oliveira and Orlando Belo
Algoritmi R&D Centre, University of Minho
4710-057 Braga, PORTUGAL
oliveira.gmr@gmail.com, obelo@di.uminho.pt

ABSTRACT

The purpose of this article is to study in detail the relationship between customer and supplier, using a data mining system to detect and define the concept of the churn customer in the telecommunications sector. Research was carried out using a real database of a telecommunications company in the Middle East, after thorough cleaning and auditing with respect to the robustness of the data and its validity within its market. The outcome of this work is the acquisition of acknowledgment on the building, understanding and definition of the churn customer, starting by defining customers with a high probability of leaving the company, and eventually identifying the loyal customers who represent higher profits.

KEYWORDS

Telecommunications, Data Mining, Classification systems, Churning Identification, Middle East.

1 INTRODUCTION

At the end of the 20th century and in the early of 21st century telecommunications became one of the most dynamic emerging sectors of the world economy. In less than 100 years, 60% of the world's population acquired access to some form of telecommunication. Growing at a surprising rate, wireless telecommunications stopped being the hallmark of secret institutions and a niche of millionaires, becoming a common good in all corners of the world. Telecommunications services are now the status symbol of youth, sophistication and modernity in nowadays' society. The world climate of globalization and opening-up of markets and the stimulation of competition, as the favoured means of consumer defence by providing better services at more competitive prices, are the main factors contributing to the rapid growth, driven by a permanent technological improvement resulting in a portfolio of new and more complex services. This growth impose to service providers the creation of a network to support the distribution and dissemination of products and services with a geographic coverage in line with the expansion strategy, whether directed at consumer niches or product popularization. Whatever the defined strategy, quality of the offer is essential to ensure customer loyalty and competitiveness in attracting and retaining higher value customers. The factors that led to the meteoric growth of telecommunications are the same as those that contribute to the emergence of the churn – customer dropout phenomenon.

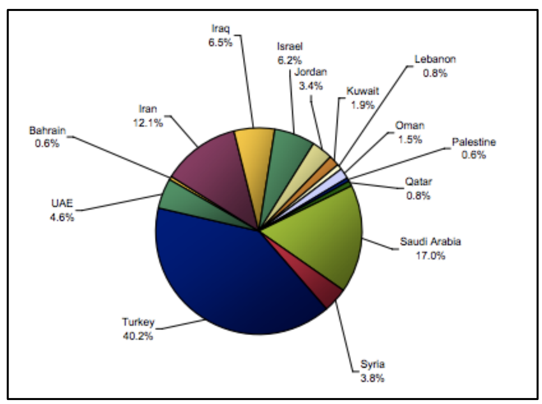

Figure 1. Mobile market in the Middle East - 2nd Quarter 2009 by country [4]

The unrelenting technological advances that enabled the generalization of mobile services are also responsible for prompting in customers the will to change. The incessant appearance of new, more sophisticated and affordable alternatives on the markets, in a search for continuous improvement and diversification of supply, generates the favourable conditions, which impel customers to consider change. The technology associated with telecommunications in fact begets its own obsolescence in an incredibly short time.

One of the regions which has seen the most growth in the telecommunications sector is the Middle East, where market liberalization along with the extension of services by multinational conglomerates and strong competition, has contributed to a real revolution in communications. Saudi Arabia, in particular, has enjoyed growth of about 100% in *personal communication systems* (PCSs) in the last 5 years, being expected that this growth will be even more marked in the next few years (Fig. 1). In emerging countries with a rapid economic growth, where *average revenue per user* (ARPU) is normally higher, PCSs have had an even more significant impact on the economy, and can be twice that of more developed countries [4].

2 CHURN – CASE STUDIES

In most markets, telecom operators have invested significant resources to understand the churn behaviour and its impact during the last decades. In this paper two previous studies in this area are presented due to its strong relationship with the models developed, which served to compare both options and results. One case study was a work presented in a Brazilian telecommunications company [11]. This study involved the analysis of a real-world telecommunications database that featured a segment of high monthly traffic customers, with a tendency to defect. An analysis of the behaviour of the variables led the author to infer that between the eighth and ninth months there were variables which could by themselves explain and classify the customer with respect to churn. This finding enabled the initial choice of 12 months of data to be cut to a smaller sample. The three analytical methods used, i.e. distribution of the standard mean over time, distribution of the determinant of the covariance matrix over time and eigenvalue analysis, confirmed the validity of reducing the sample, facilitating the processing cost in the analysis process. To classify the variables the author used predictive analysis through three synthesized variables: level, tendency and volatility. Each of the characteristics used to synthesize the new variables showing extreme efficiency in representing and splitting the groups. Examination of the unsupervised models used revealed a movement of the centroids. This means that, in the absence of controlled modelling (churn and non-churn classes). The fact that it was possible to eliminate the last month from the sample demonstrates the robustness and efficiency of the model, which was particularly important to analysing and consequently taking action to retain customers in good time. The third case study [15] was undertaken in a large telecommunications operation in Brazil, using data on the broadband service between January and December 2006. The aim was to find the relationship between quality of service and the customer churn rate. The study focused on voluntary reasons for defection related to the quality of the service provided by the operator to derive an equation that represented the churn rate as a function of quality of service, using multiple regressions. This regression method was chosen to provide a means of assessing the degree and nature of the relationship between the dependent and independent variables, through the statistical variable, based on the independent variable. The tool used was SPSS because it is easy to use, is well documented and has a good track record in multiple regression analysis. Analysis of the correlation matrix showed that the values of the independent variables did not exceed 0.5 and it was concluded that there was no multicolinearity - independent variables that have exact or nearly exact linear relationships. Values for the variance inflation factor are quite close to one, so it was concluded that multicolinearity did not exist. After adjustment and making use of normal distribution, a distribution consistent with the residuals in relation to the theoretical distribution was found, that is, they are close to the normal curve. Analysis of the data obtained showed the existence of correlation between the quality indicators determined and the churn rate. The four variables most relevant to the churn rate were identified from a set of 12: mean time to repair; mean time to install, mean time for prevention, and repair rate. This makes it possible to construct a set of recommendations and remedial actions that contribute directly to a decline in churn. It is extremely hard to forecast churn customers in telecommunications companies, both because of the large amount of data and because segmentation of the data marts is normal and frequent, which complicates the consolidation of the information needed to classify churn customers. Taking the case studies as examples, the huge importance of the use of data, the choice of variables and the choice of the most appropriate techniques to derive the most efficient models that present the best results is abundantly clear.

3 THE WORLD OF CHURN

Many authors have been studying the various forms of churn and its business perspectives. According to Strousse, 'Churn is a disloyal act by a customer, that is, the loss of a customer to a rival company' [5], thereby ending the link with the former company, and can be defined by the 'annual rate of turnover of the customer base' [5]. Strouse also says that if there is a 5% annual rate of churn, this means that the company has lost 5% of its customers during the year because they are in principle dissatisfied with something about its services. In a market as competitive as telecommunications companies must do their utmost to try and retain loyalty in an effort to prevent defection and the consequent high cost of regaining these customers. Advance warning systems are one potential solution that could anticipate and prevent customer defection by implementing techniques and measures to retain them. Companies could thus both retain their most lucrative customers and enhance their loyalty level. Another equally important objective is to learn from situations that involve customers who leave the company, so that they can forecast more effectively and prevent future defection. Lejeune [6] believes that another factor to consider is that of restricting the customer base in opposition to the overall analysis of customers, for reasons of efficiency of earnings. Those customers who represent significant value for the company should be considered. In these circumstances efforts should be made to appraise what kind of customer is under analysis, since it could be one of high value to the company, whose defection would have a considerable impact – financial, for instance. Analysis of churn customers in telecommunications companies makes it possible not only to identify those with high defection rates, but also to single out the most valuable ones. So we have managed to proactively apply a number of actions, chosen with a view to ensuring both customer retention and customer loyalty. It is important to remember that operator-customer behaviours change over time, and these changes stem from causal, social, cultural, personal and psychological factors [7].

3.1 Voluntary churn

When we try to understand the exact reasons why customers choose to leave an operator, we realise that little effort has been made on interpreting their profile, as a customer, and on their initial motivation to sign up for the service, or why they want to switch operator. While the price factor is the main reason a customer leaves, the decision that leads to switching operator is more complex since there are many other factors involved (technological, economic, quality, and many others).

Voluntary churn occurs when a customer initiates the process to terminate the contract [1]. It can be subdivided into 2 major groups: accidental, this kind of churn does not happen because the customer has planned it, but because some event in their life has forced them to terminate the contract. Termination of the contract is seen as a collateral effect of this event [1]; deliberate, while in accidental churn there is no objective motivation of the customer that might prompt preventive action from the operator – a situation that does not arise very often – in deliberate churn the customer makes a deliberate decision.

3.2 Involuntary churn

This is the commonest kind of churn for most telecommunications operators, and it has the biggest financial impact. In this kind of churn customers have no intention of switching operator but a series of circumstances force them to cancel the subscription [1]. The reasons or circumstances for this can be placed into 3 sub-categories: **f**raud, when illegal activities are perpetrated; churn due to non-payment, customers that have financial problems; churn due to under-usage, customers do not use the service (e.g. pre-paying customers who do not make a monthly payment and do not use the service).

Churn can occur in the telecom sector in several ways, with each operator applying the definition most appropriate to its situation, based on the standard understanding of churn customers. This process takes into account a number of demographic, geographic, economic and behavioural variables. Since the market of the telecommunications company under consideration has certain characteristics, this study focused on questions such as:

- What is the typical behaviour of churn customers?
- What are the thought processes and motivations of voluntary churn customers?
- Which are the non-payment involuntary churn customers?

3.3 Internal churn

Different telecom operators use the reference of 'internal churn' differently. Usually, it's associated with customers who want to end a particular service to sign up to another product or method of payment. One example of this is the migration of a customer from a post-paid tariff plan to a prepaid one. This is one of the commonest circumstances in telecommunications, arising from consumers' financial constraints. This is the least important of the three kinds of churn described because the customers do not in fact leave the company. This type of churn is not examined in this study.

3.4 Mining Systems

Having described the various kinds of churn we shall now analyse and discuss the ways and means at our disposal today to detect them. There are all kinds of approaches to this phenomenon. But unfortunately they are not all sufficiently effective or useful. This work does not address all the strategies and techniques for identifying and tackling cases of churn. This study therefore emphasizes techniques related to classification systems, e.g. decision trees, neural networks and logistic regression [9], which are the most popular techniques in the studies analysed (section 2) for predicting churn.

Decision Trees - Many decision tree applications were used in the past to supplement statistical techniques such as logistic regression to refine and process data. But today this technique is increasingly used in predictive models. The idea that decision trees could only be used to explore data is clearly undergoing a change. A decision tree algorithm 'looks' for all the possible questions that can split the available group of data into homogeneous segments with respect to the different classes envisaged. Some decision tree algorithms can also use heuristics to choose the questions.

Neural Networks – This is one of the most popular techniques for building predictive models. They are systems for processing distributed parallel information and consist of processing units, usually called nodes, neurons or cells, interconnected by unidirectional arcs, also known as links or connections. The nodes have a local memory and can carry out localized information processing operations. Each cell has a single exit, which can branch into multiple collateral connections (each branch has the same output signal as the neuron). All the processing in each unit must be completely local, that is, it must only depend on the current values of the input signals coming from the neurons via the connections. These values act on the values stored in the cell's local memory [16]. The advantages of neural networks is that the user does not need to know much about how they work, or about how the predictive model is built, or even about the database. Most neural networks can be used without reorganization or modification of the data, yet the opposite is very often true since important decisions need to be taken in designing the model to effectively use a neural network, in particular:

- How the network's nodes should be connected?

- How many neurons (processing units) should be used?
- When should training be stopped to prevent over fitting?

As in all predictive models, neural networks require special attention to prevent over fitting. From this standpoint neural networks are relatively good with training data but less so with new data. This is, indeed, the main problem with neural networks, since it is very hard to see how the model works. Whereas decision trees and nearest-neighbour methods can quickly attain 100% prediction efficacy in the battery of training data, neural networks can be trained forever without reaching 100%.

Linear Regression – this technique considers powerful methods and often used in statistics and data mining, normally in prediction procedures. There are several kinds of regression in statistics, but the main objective they share is the estimation of a conditional expected value. The simplest form of regression is linear regression, a statistical method to estimate the conditional (expected value) of one variable Y, given the values of other variables X [10]. In general, regression deals with the estimation of a conditional expected value. Linear regression is so called because it is considered that the relation of the response to the variables is a linear function of some parameters composed of a predictor and a prediction. In predictive models the trick is to find the model that best minimizes the error. The most common way to calculate the error is the square of the difference between the predicted value and the actual value.

4 A DATA MINING APPROACH

The various predictive models that are described in this study are designed to extract patterns and discover knowledge about information on the history of customers of a telecommunications company, so as to classify and predict churn customers who are very likely to leave the company for one of its rivals. The basic structure of the models is usually based on the 'KDD [method], the process of identifying valid, new, potentially useful and comprehensive patterns in data' [2]. However, because of its flexibility and iteration over the course of several steps, it was decided to use CRISP-DM [3] since this approach supports the development of data mining projects.

4.1 Understanding the Business Rules

In the previous year the telecommunications company that is the subject of this work had an annual churn rate of 4.8%. The results of the first quarter of this year indicate that the churn rate was 0.9%. Bearing these figures in mind, the retention department decided to implement a project that would make it possible to identify the valuable customers most likely to churn and so take action to retain them that encourage their loyalty and make them profitable. Implementation of this project required a study of the business that would identify the problem in detail so that it could subsequently be turned into a typical data mining problem [8].

4.2 Characterizing the Churn

Churn is the act of discontinuing the main service at the request of a customer or for non-performance by the latter of part of her/his contractual commitment. The main services considered for the termination of a contract are: the mobile service (prepaid and post-paid) and the fixed service (ADSL and dial-up). The prepaid mobile service relates to voice calls, SMS and MMS services associated with a mandatory periodic top-up payment, and post paid relates to a subscription voice, SMS and MMS service, but with a fixed-base monthly tariff.

With respect to the fixed service, this can relate to ADSL and Dial-up Internet services, where the first is a rapid data transmission technology via cable and the second is data transmission via a telephone connection. After determining how to identify churn customers in operational systems, we get to the final step of the analysis, with the definition of the concept, or target. The target is the outcome of an action triggered by the customer. The action can thus be characterized by the act of switching, prompted by the customer, where the target is a binary variable that represents the positive or negative action taken by the customer: 1, the customer is going to leave the company (churn) – negative actions; 0, the customer is going to stay with the company – positive action.

4.3 Defining the Mining Objectives

Prediction must be undertaken three months in advance for post-paid mobile service customers who are contactable, of at least 6 months standing and belong to the 5-star segment (highest value segment). The following are used to achieve the final objective: network traffic history, billing, tariffs, products, and demographic information.

4.4 Understanding the data

One of the most important steps in identifying and characterizing churn must involve the analysis and effective understanding of the data that will support this process. This stage starts by selecting a significant sample of the available data, identifying problems related to the quality and understanding of the data and detecting subsets that are useful to formulating hypotheses relating to any items of information that may be hidden.

The attributes and volume of the data comprising each table were identified in the description of the data. In the exploring the data phase, analysis of the quality of information to understand which attributes to aggregates, which fields may be removed because they do not add value, and the level of attribute redundancy and duplication were carried out. Once the data exploration completed, they had to be validated to ensure that all the data had been correctly chosen from the different sources. Validation of their relevance to the case study and the detection of any errors and the frequency with which they occurred is crucial to any data mining process.

In this case study it was decided to eliminate occurrences for which the variables had no value since there is no standard value that could be correctly applied to such situations.

4.4 Preparing the data

The main purpose of data cleaning is to ensure the minimum quality necessary to develop the models. It became necessary to take action to clean the data involved in the process so as to obtain the expected results. One of the commonest problems in this phase is the presence of unknown values in the attributes, such as entries with incomplete data due to errors in the selection or validation process. Several techniques have been used to handle these unknown values, including replacing them with the average or attribute mode, or replacement with predicted values, using the learning algorithm. In the event of prediction, the algorithms provided by the tools do not offer effective results on data with unknown values, even utilizing simple techniques, such as the replacement of all the unknown values with a global variable, can entail unwanted results, thereby warping the prediction process. In this case the unknown value was interpreted as a positive factor that cannot be justified [13]. Once the data have been cleaned the attributes can be constructed. The data used in the mining process in this study employed user induction to construct the attributes in the definition of the final sample. One generic example is the ACCOUNT_ ACTIVATION_DATE attribute that contains information on the service activation date. Based on induction technique the SNAP_ACTIVATION_DAYS attribute was construct, which represents the number of days elapsed from the activation date up to the date of churn measurement. Data integrity involves the analysis of the relationships allowed between the attributes.

Having examined the business requirement and created the data extraction process, the integration of the information was assured and can be said that the source table for the models considered contained one entry per customer. The objective of transformation phase is the representation of the data so as to overcome the limitations of the algorithms used in the mining process. Deciding on the transformations to adopt depends on the algorithms used in the mining process, and mathematical functions usually evolve when obtaining the data, to avoid the limitations of the algorithms. The commonest transformation techniques are: normalization by standard deviation, normalization by variance, transformation of quantitative attributes into qualitative ones and vice-versa, transformation of complex data and data reduction. Regarding normalization, this is a process that transforms the intervals of the values of the attributes into a specific interval (-1,1). This kind of transformation is important for methods that calculate distances between neighbours [14], though it is not very useful for most methods that induce representations (decision trees), since normalization reduces understanding of the model generated. When applied to neural networks it helps achieve better training processes.

With respect to the transformation of the quantitative attributes into qualitative ones and vice-versa, some algorithms are limited when it comes to working with attributes of this kind. The values must be converted in accordance with the limitations of the algorithms. Several approaches can be used. For the transformation of qualitative variables into quantitative ones, replacement by numerical values is often implemented. It is the other way round for transforming quantitative into qualitative variables [14]. Although none of the algorithms used requires only nominal attributes in the training sample, it was decided to carry out a transformation of all the qualitative attributes into quantitative ones. In relation to the transformation of complex data, the great majority of the algorithms used in data mining are unable to work with date and time data. In general the transformation method used for data of this kind is conversion into an integer attribute. A practical example of this kind of transformation is the *SUBSCRIPTION_DATE* attribute, which represents the date on which the customer subscribed to the service with the format "20080201". The initial transformation converts the value of the date format into integer format, proceeding to the construction of a new variable (*SNAP_ACTIVATION_DAYS*). This will contain the value of the difference of dates between the value of the *SUBSCRIPTION_DATE* attribute and the current day.

Regarding to data reduction, there are clear advantages in using large amounts of information. But in practice the set can be too big and exceed the capacity of the prediction tool [13]. The data reduction technique is of the utmost importance in this study, since the large amount of entries with high processing cost would jeopardize the building of the models and analysis of the results. The data reduction model proposed by Weiss [13], considers a set (S1) of initial data, these are reduced in terms of the amount of values (or attributes), and a new set is produced (S2). When the size of the initial set lie within acceptable limits, no data reduction techniques are necessary. Once obtained, the reduced set of data can be divided into training and testing sets. Weiss describes a series of techniques for data reduction in data mining processes. Cluster sampling was the technique used in this study, in which the elements of a set of data form a cluster (customers belonging to the 5-star segment). In the case study, the amount of the sample is 5,000 entries, 4,701 (94.20%) of which belong to the non-churn segment and 299 (5.98%) belong to the churn segment.

4.4 Modelling

Several algorithms are used for prediction in data mining. The first method studied was the neural network. Several studies were carried out using different methods, with normalized and non-normalized samples. The pruning method exhibited the best behaviour. It main feature is that, starting with a large network, it prunes within the hidden and weakest input units, as the training progresses. Though it may be one of the slowest network algorithms, it usually gives the best results. The configuration of neural networks parameters influences the performance of the algorithms.

Using the SPSS Clementine, several tests were performed to find the best configuration parameters of the neural networks, and those giving the best results are: use partitioned data, if there is a partitioned field this option ensures that only the training data will be used to build the model; stop on, use of default criterion, training

ISBN: 978-0-9853483-3-5 ©2013 SDIWC

Table 1. Results obtained from the predictive algorithms

Methods	Training				Test				Validation			
	Cor.	Inc.	% Cor.	% Inc.	Cor.	Inc.	% Cor.	% Inc.	Cor.	Inc.	% Cor.	% Inc.
Neural Network - Prune	1994	6	99.7	0.3	1495	5	99.67	0.33	1404	96	93.6	6.4
Decision Tree - CHAID	1979	21	98.85	1.05	1477	23	98.47	1.53	1465	35	97.67	2.33
Logistic Regression	1943	57	97.15	2.85	1445	55	96.33	3.67	1422	78	94.8	5.2

only stops when the network attains the optimal state; optimize, the use of the Memory option to force the algorithm to use disk spilling in the processing phase. The decision tree was the second method studied, and the *CHAID* algorithm performed best. The best parameters chosen to obtain the best model were: use partitioned data, if there is a partitioned field this option ensures that only the training data will be used to build the model; mode, generate model was chosen to create the model; levels below root: if the algorithm needs to increase the number of levels to construct the tree, the limit is 15 levels. Finally, logistic regression was studied, using the stepwise, backward, and forward algorithms [10].

The stepwise approach exhibited the best behaviour since it starts by choosing the best discriminating variable. The initial variable is paired with each of the other independent variables, one at a time, and the variable that is best able to improve the discriminating power of the function in combination with the first variable is chosen. The third variable (and others) are chosen by the same process.

5 RESULTS EVALUATION

The preceding sections have set out the data mining processes undertaken to identify churn in a telecommunication company. The way they were carried out, led to the finding and achievement of some quite interesting results. In fact the use of more than one data mining algorithm, the working option adopted from the outset, helped these results and situations to emerge more naturally. It is true that the data chosen also played a part in the results obtained (Table 1). Comparing the various methods with the validation sample, we see that the different algorithms used perform well and very similarly: 93.60% in the neural networks with the pruning algorithm; 97.67% in the decision tree with the CHAID algorithm; and 94.80% in logistic regression with the stepwise algorithm. Based on these indicators it was found that the results of the study, though consistent, are too high in comparison with the churn prediction models in the case studies presented in section 2, whose results range between 50% and 70%.

The results of the prediction algorithms presented in table 1 are too high, which is a consequence of the churn being measured almost as it happened, when most of the customers had already left the company. Due to non-payment, most of them did so involuntarily. Bearing in mind the sample used in the validation given in Table 1, it can be seen that the decision tree was the method that best classified the non-churn class (99.83%), while the neural network provided the best results in the churn classification (82.22%). However, considering the results obtained for the churn and non-churn classes with the neural networks and decision tree methods, and bearing in mind the specific features of each of them, it is concluded that the latter method best meets the customer's needs.

In fact there is no perfect model for the various situations, but certain models do perform better, depending on the different types of data presented and the customer's requirements. In the case of neural networks, and based on the algorithm that provided the best results (pruning algorithm), examination of the information in table 2 shows that there is a marked decline in the successful classification of churn when we compare the testing (95.56%) and validation samples (82.22%). This does not happen when the training (96.67%) and testing (95.56%) samples are compared, which indicates that overfitting is involved. When using this method it was found that the variables that best represent the classification of churn customers are: eve_last_out_ call_date (0.3954), eve_last _call_charged_date (0.3946) and eve_first _out_call_date (0.21).

Table 2. Hit percentage for each class

Methods	Classe	Training	Test	Validation
Neural Network - Prune	Non-Churn	99.89	99.93	94.33
	Churn	96.67	95.55	82.22
Decision Tree - CHAID	Non-Churn	100	100	99.93
	Churn	82.5	74.44	62.22
Logistic Regression - Stepwise	Non-Churn	99.36	99.76	98.58
	Churn	62.5	42.22	46.67

The coefficients presented can be interpreted as follows: the value of the coefficient for the variable 'date of last call made' eve_last_out_call_date is 0.3954, which means that, with the other variables being separated out, for a 10% normalized increase in the eve_last_out_call_date variable there is a normalized increase of 4% in the rate of churn, which, in non-normalized figures, relates to an average increase of 60 customers.

The other indicators can be interpreted in the same way. In fact, when analysis of the meaning of each variable shows that customers whose last call date is quite old, that the date of the last call paid for is also old and that the date of the first call made was quite some time ago, it may be concluded that these customers are extremely likely to leave the company. But this information is not enough on its own to confirm whether the cause of the churn is related to a voluntary or involuntary circumstance since the model did not choose any traffic variable, which would be essential to determine the cause of churn.

Churn = 2.727 * SNAP_ACTIVATION_DAYS +
21.08 * SNAP_GRACE_PERIOD +
24.2 * INV_COUNT_VOICE_OUT +
13.75 * INV_OUT_USAGE_VOICE +
50.52 * EVE_FIRST_OUT_CALL_DATE +
-56.32 * EVE_LAST_OUT_CALL_DATE +
11.62 * INV_AVG_AIRTIME_CALL_OUT_INT +
25.56 * INV_AVG_USAGE_CALL_OUT_ROAM +
+ 0.3971

Figure 2. Logistic Regression – Stepwise

If we look at the decision tree method, in particular the CHAID algorithm, we find that the most important variables were: eve_first_ out_call_date (0.722), inv_ count_ voice_int (0.221), eve_last_ call_date (0.03) and eve_last_3m_usage _avg (0.028). The coefficients presented should be interpreted in the same way as described for the neural networks model. Analysis of the meaning of each of the above variables showed that customers who made the first call a long time ago, who had not made calls in the recent past, and whose quarterly average of calls made was very low, are clearly involuntary churn customers due to non-payment. The operator explains this, barring calls made after the customer has failed to pay a bill. It may be concluded that the variables related to customer traffic are hugely important in this case since they enable us to explain the reason for the churn. Regarding the logistic regression approaches, the stepwise algorithm provides the best results. In the equation yielded by the relevant algorithm the most important variables and which comprise the mathematical equation are:

The coefficients of the stepwise model can be interpreted as follows: the value of the coefficient for the 'days of service activation' variable up to the time churn was measured snap_activation _days is 2.727, which means that, with the other variables of the equation being separated out, for a 10% normalized increase in that variable there is normalized increase of 3% in the rate of churn, which, in non-normalized figures, related to an average increase of 45 customers in the churn rate. To sum up, it was found that the most critical situations of churn were detected by means of the logistic regression model, and this finding was borne out by the results achieved, by the data mining models and strategies followed, by the analysis of case studies and by the corporate management policies followed by the target company. Table 2 provides a summary of these cases. Even though the neural networks show a higher performance percentage than the other models, a number of factors support the choice of the decision tree model as the one offering the best performance, given the business requirements. These factors are: the question of over fitting inherent to neural networks, the good results of the decision trees and the fact they are easy to interpret, and the values obtained in the second case study, which bear out the validity of this study. The results achieved, though credible, do not represent reality because they have a high percentage in the classification of churn customers, which is due to the measurement being taken at almost the same time as the event, as noted earlier. This is why the construction of two churn classification models should be considered at the analysis stage: one model to classify the involuntary churn customers through failure to pay or fraud, and the other to classify the voluntary churn customers, since from the business point of view it is important and necessary to study and classify the two types. If this approach were considered at the analysis stage, not only would the algorithms yield different results, but these results would surely be more realistic and true to life.

6 CONCLUSIONS AND FUTURE WORK

In fact the impact of churn on telecommunications companies is huge because of the intense competition in current markets. This happens due to the large variety of offers that appear in telecommunications market, most of which are quite attractive and whose quality-price relation prompts customers to constantly switch telecoms operators. Therefore, it is important to study the concept of churn, since the cost of gaining new customers, or regaining old ones, is higher than the cost of retaining them. In this study it was decided to subdivide churn into 3 distinct groups: voluntary, involuntary and internal. However, the main purpose of this study was to identify the two first kinds of churn since until they have been properly studied internal churn is of no relevance. In addition, we wanted to identify the customers that would be most likely to switch companies, in this particular case telecommunications companies (churn customer). The efficacy of the processes developed and their appropriate monitoring were ensured by means of the CRISP-DM approach, which is, according to most of its followers, flexible and iterative in the development of several steps of a data mining process. Three predictive models were developed based on neural network, decision tree and logistic regression methods. These methods were chosen because they are the ones traditionally used to predict churn customers [12]. The decision tree was the method selected, for various reasons: it harmonized with the business requirements, it performed well with the three samples tested (most efficient method) and it offered added value in the analysis and support of decision making. This model correctly classified all the non-churn class cases in the training and testing samples and in the validation sample it correctly classified 99.93% of the cases.

Relatively to the churn class, there was a relatively constant decrease in the correct classification of cases in all three samples, as shown in table 2. The explanation for the high results stems from the fact that the churn customers were measured at almost the same time as the event occurred, and most of the churn customers are involuntary, through non-payment. Regarding the churn classification variables, as alluded to in section 5, those that best represent this model permit the inference that one of the main factors for classifying churn customers is the fact that such customers have made their first call a long time ago, have not made calls in the recent past and have a very low quarterly average of calls made. In this case we have an involuntary churn customer whose service has been barred for non-payment of a bill. Finally, and bearing in mind that the construction of two separate models (to handle voluntary and involuntary for non-payment churn) was not considered in the analysis phase, it is concluded that the results, although restrictive, do not impede the use of the decision tree in other classification and prediction problems that involve issues of customers defecting and switching to rival companies.

Finally, and as suggestions for possible future work, we need to try and improve the processes implemented and the results achieved, with special focus on: the differentiation between voluntary and involuntary churn customers in the training samples; the improvement of samples by including more derived variables and continuous variables related to information on incoming calls – information from the CDR bypass; and the application of mining systems to all segments so as to obtain information on the whole customer base.

7 REFERENCES

1. Mattison, Rob, The Telco Churn Management Handbook, USA-2005.
2. Fayyad, Usama; Piatetsky-SHAPIR, Gregor; SMYTH, Padhraic, From Data Mining to Knowledge Discovery in Databases, January 1997.
3. SHEARER C,"The CRISP-DM Model: The New Blueprint for DataMining". JOURNAL of Data Warehousing, v5, n4, p. 13-22, 2000.
4. Hottelecom, Middle East – Mobile subscribers Statistics and Forecast by country, December 2007.
5. Strousse, K. G., Marketing Telecommunications Services: New Approaches for a Changing Environment. Norwood: Artech House, 1999, p.271.
6. Lejeune, M. A. P. M., Measuring the Impact of Data Mining on Churn. Internet Research, v11 n5, 2001, p375-87.
7. Kotler P.; Armstrong G., Princípios do Marketing (9. Edição). Prentice Hall, 1999, P.77.
8. Swift R. S., Accelerating Customer Relationships: Using Crm and Relationship Technologies. Prentice Hall, 2001, P97-110.
9. Han, J.; Kamber, M., Data Mining: Concepts and Techniques, 2a ed. Morgan Kaufmann Publishers, March 2006.
10. Montgomery, D. C.; Peck E. A.; Vining, G. G., Introduction to Linear Regression Analysis (3rd Edition). Wiley Series In Probability and Statistics, 2006.
11. Andrade, D. F., Uma Analise De Cancelamento Em Telefonia Utilizando Mineração de Dados. Rio De Janeiro, 2007.
12. Hung, S.; Yen, D. C.; Wang, H., Applying Data Mining To Telecom Churn Management. Taiwan: Department of Information Management, National Chung Cheng University, 2006.
13. Weiss, S. M.; Indurkhya, N., Predictive Data Mining – A Practical Guide. San Francisco: Morgan Kaufmann Publishers Inc., 1998, P230.
14. Batista, G. E., Pré-Processamento de Dados em Aprendizagem de Máquina Supervisionada. São Carlos: Instituto de Ciências Matemáticas e de Computação, 2003.
15. Neto, A. P, F.; Mattozo, T. C.; Costa, J. A. F., Uma Aplicação de Mineração de Dados no Gerenciamento do Churn em Serviços de Banda Larga de Telecomunicações. Niteroi, Brasil, 2008.
16. Hecht-Nielsen, R.Applications of Counterpropagation Networks. Neural. Networks Journal, 1988. p. 131-140.