

Operational comparison of Logistic regression, Decision trees & Neural networks in modelling mobile service churn.

Arun Parekkat
SPSInfoquest U.K.
Group'e Business & Décision

Abstract

The objective of this study was to compare logistic regression (LR), decision trees (DT) and Neural networks (NN) techniques for prediction modeling of churn amongst mobile subscribers at Orange PCS U.K. The study also assessed these techniques with reference to their implementation on SAS Enterprise Miner 4.1 for Windows. On development of the models the accuracies were compared. While the sensitivity of the NN model was significantly better than the LR, there were no significant differences between the NN and LR on many of the other performance parameters. The results found NN models to be effective in predicting churn, while LR being a good starting point for model derivation and in understanding the model structure. With a rigorous methodology, SAS Enterprise Miner provides the facility to develop models faster, with a comprehensive set of techniques, and a considerable ease of usage with its close integration with the SAS environment.

Introduction

Churn prediction and management is of particular concern to the competitive and rapidly maturing telecommunications industry. Subscriber churn, with customers defecting from one mobile service provider to the other is the issue in focus, with the cost of replacement of a lost mobile service customer at £200 to £400 in terms of sales support, marketing, advertising and commissions (SAS, 2001). Other hidden benefits of existing customers, such as liquidity, better price sensitivity make losing existing service customers an increasingly costly proposition. Building appropriate churn prediction models is therefore crucial for the bottom line survival for Orange.

Three techniques – Logistic Regression (LR), Decision Tree (DT), Neural Networks (NN) were used to estimate the churn rate among contractual subscribers of Orange. The advantage of the LR and DT models is that these techniques provide insight into the relation between subscriber characteristics and the churn. The advantage of using neural nets is that these can detect complex non-linear relations between the independent variables and churn target and provide a more accurate estimate. Though interpretation of NN models is difficult.

Data analysis & sampling

Data is categorized into the two blocks namely: subscriber level demographic, contractual & marketing data and subscriber level call & service logs. The most important level is the subscriber level. A subscriber may have several accounts (e.g. business, personal etc.). The customer level is above the account level. This is used when the

customer signs up for say a new service and is not particularly useful in churn modelling.

Independent variables were individually tested for their R-square values (Coefficient of determination) with a cut-off criterion of 0.005. Coefficient of determination is the proportion of target variation explained by a variable i.e. here the effect on the target is calculated by measuring univariate R-square values. Variables were thus reduced from 91 variables to 35 variables.

A stratified random sample with 50-50 stratification of churners to non-churners was used. Predicted probabilities are adjusted to compensate for the over-sampling while generating results.

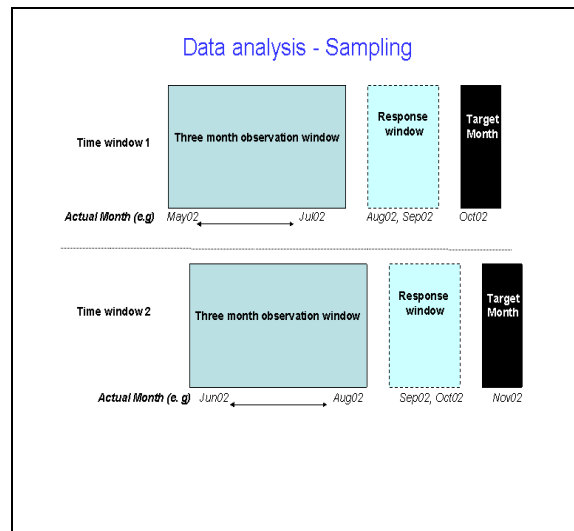


Figure1: Data set used for modelling

A retention period or response window of two months was built into the model between the target month of one month and the observation (past data) period of three months. A set of two time windows was used while designing the datasets.

Target & variables

The model target - churners are defined as subscriptions that have chosen to leave the contractual service on their own account.

Monthly churn rate = number of churners over month/ ((live base at the end of the month + live base size at the end of previous month)/2).

The target variable for customer churn, is Target = 1 for churners and Target = 0 for non-churners. Target is derived by considering only subscriptions for line 1 (several subscriptions could have more than one mobile line), selecting the personal account type, and voluntary churners. Another key variable that was developed was - tenure. For each customer in the sample, tenure indicates the period since start of customer's initiation of a service subscription.

Profile analysis of main variables is used to understand their value distributions, identify outliers and churn rate. Analyzing some of the profile charts, some variable levels were collapsed, formatted and new variables developed.

Modelling

This section looks at the modelling effort during the project. As a first step the most significant variables that strongly explain churn prediction is identified. This development was based on the training data set. Profile analysis, correlation analysis and univariate logistic regression was used to identify the subset of variables.

Criterion	Definition	Computation from 2 by 2 classification table
Sensitivity (Hit rate)	Probability that an actual churn case is predicted as churn.	$D/(C+D)$
Specificity	Probability that an actual non-churner is predicted as a non-churner.	$A/(A+B)$
False positive rate (False alarm rate)	Probability of an actual non-churner being predicted as churner.	$B/(B+D)$
False negative rate (Type I error) (Missed detection rate)	Probability of an actual churner being predicted as non-churner.	$C/(A+C)$
Misclassification rate	Summative measure of both the false positive and false negative rate.	$(C+B)/N$

Figure 2: Classification table & assessment

A 1:1 ratio between the minority/rare event - churners, and the majority event - non-churners is

used in modelling. As non-churners increase in proportion to the churners, misclassification rates (measure of false positive and false negative rates) increase and are higher than at lower target ratio.

A set of two-time window is used, as increasing the number of time windows, from 1 to 2 & 3 results in marginal improvement on prediction effectiveness. With more time windows, while there is an improvement in the capability to capture more subtle variations in subscriber trends. Increasing variability, in areas irrelevant to churn, offsets the increase. At any level of misclassification rate a shorter response window results in lower misclassification rates. This has to be balanced with the ability to get access to data, and the time required to action customer retention measures.

		Predict		Total (Actual)
		Non-Churner Target=0	Churner Target=1	
Actual	Non-Churner Target=0	A True -ve (Specificity)	B False +ve (False Alarm) Type II error	A+B
	Churner Target=1	C False -ve (Missed detection) Type I error	D True +ve (Hit rate/Sensitivity)	C+D
Total (Predicts)		A+C	B+D	N

Figure 3: Classification table measures

Finally, a selected subset of variables was used in building the final set of LR, DT and NN models. At the end of the process, the estimated probability were compared with the known outcomes.

Evaluation & analysis criterion

Some measures are statistically heavy; while others provide simpler measures of intelligence suited to customers of modelling assignments - the management team.

Sensitivity, specificity, ROC curve & AUC

Sensitivity - the measure of accuracy of predicting churn, is the ratio of true positives by total actual positives and specificity - accuracy of predicting non-events, is the ratio of true negatives by total actual negatives. The classification table is generated using a classification threshold of 0.95. Receiver-operating characteristic (ROC) curves for the models are plotted and area under the ROC curves (AUC) is calculated. ROC curve plots the false-positive rate (1-specificity) on the x-axis and the true-positive rate (sensitivity) on the y-axis. Area under the ROC curve is a measure of a model's discriminatory power. The closer an ROC

curve is to the upper left corner of the graph (as true-positive rate approaches 1 and false-positive rate approaches 0), the larger the area under the curve, and more accurate the prediction model. Each point on the curve represents a cut-off probability. A lower cutoff typically gives more false positive. A high cutoff gives more false negatives, a low sensitivity, and a high specificity (SAS EM reference, 2001).

The trapezoidal method was used to calculate AUC. AUC ranges from .5, for a worthless model, to 1 for a perfect classifier.

Classification chart measures

Missed detection rate - probability of an actual churning not predicted and detected as a churning. This is a high-risk business scenario, as the churning is lost.

Misclassification rate - for predictive models, two misclassifications errors are possible. Missed detection rate makes up the Type 1 error made when a high risk components is classified as low risk, while a Type 2 error is a low risk component which is classified as high risk. It is desirable to have both errors as low as possible (Lanubile, 1997).

Lift charts

In a lift chart (also known as a gains chart) for a binary target, the scored data set is sorted by the posterior probabilities of the event in descending order. Then the observations are grouped into deciles. The percent response vs percent captured response lift chart for binary targets, classifies responders and non-responders based on the event level of the target.

Model analysis

Modelling assessment

Analysis of the models have involved comparison of the three models using classification tables and related measures. This is the traditional method of model assessment (Fedenczuk, 2002).

On comparing the three techniques, we find that differences of the model sensitivities/hit rate are not very different. The largest difference is seen with regards to the decision tree model, which turns out to be disappointing in its performance. The sensitivity rate is 41.4% with neural nets model, decision trees at 40.17% and LR at 36.73%. The high risk, Type I error rate of LR, DT and NN models is 1.02%, .99% and .95% respectively. Thus, NN model is selected as the more promising. Further, from the ROC curves of the three models, we calculate the AUC measure. We find that NN show a significantly higher AUC than the other two models.

Finally, when the accuracy (sum of true positives and true negatives divided by total number of observations) of the models show that LR yields accuracy of prediction of 92%, neural nets - 91.89% and decision trees yield 90% accuracy.

To sum, classification tables and attendant statistics propose the NN model for its significant sensitivity with average misclassification and accuracy rates, while LR model is proposed for its high overall accuracy levels.

Advanced assessment with existing model

Businesses require easier, graphical assessment measures that enable quicker decisions. Lift charts, percentage cumulative churn response are just such comparison measures. Beyond that, in this sub-section, we also compare the model performances with the performance of the existing regression model (Ex. Reg) at Orange PCS.

Lift measurements is the much preferred comparison tool by business customers. At each decile it demonstrates the model's power to beat the random approach or average performance. The cumulative captured response chart identifies customers who churned in each of the deciles. We see that in the first top decile of the NN model captures churners 48% of all the churners, while the top 5 deciles capture around 93% of all churners.

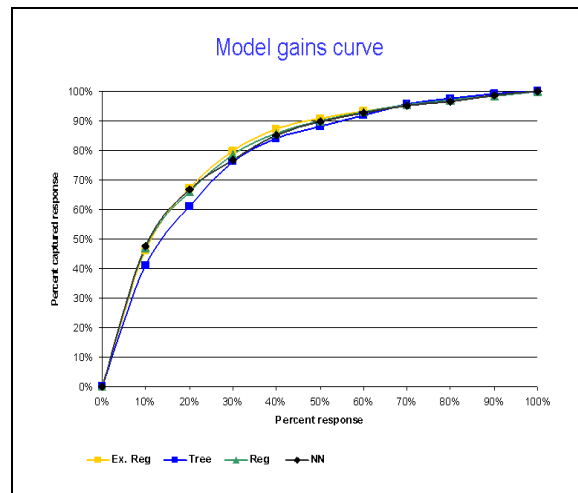


Figure 4: Percentage captured response lift/gains curve

The cumulative lift curve; illustrated next, on the other hand provides a comparison of how much better does each model perform with regards to average captured response. We see that in the top decile the NN model is a 4.76 times better predictor of the churn event.

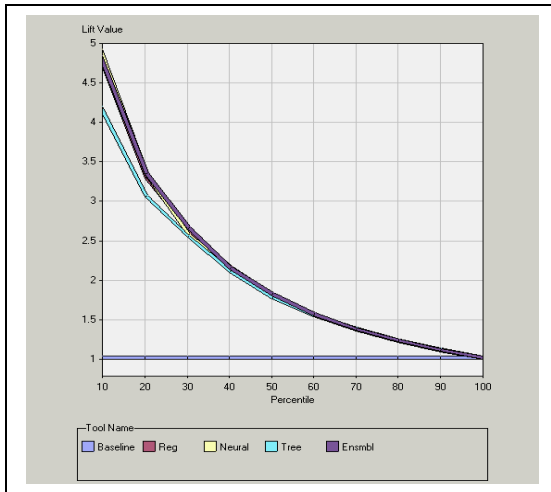


Figure 5: Model lift chart

The Model

Target & model explanatory variables

The outcome variable, or the dependent variable describes the subscriber's status of churn.

Studies show that churn is closely related to subscriber's tenure with the service. During the months leading to end of the first contractual period (contracts are for twelve months), there are more churners than the other months. The other key variable is the service plan. For example, some service plans exhibit high churn, with large subscriber volumes. Older people and young people exceed the average churn rate with regards to age. To sum the following variables explain subscriber churn – Tenure, service plan summary, age, and delivery channel. Additionally, some of the significant variables such as number of total incoming calls, duration of outgoing WAP calls, geo-demographic measures such as Mosaic also effect subscriber churn.

SAS Enterprise Miner & modelling

Process modelling and data mining can be accomplished with comparative ease using the SAS EM and provide a quicker tool infrastructure for the adept data-miner. It can be said that SAS Enterprise Miner provides a rigorous tool set, with close integration to Base SAS and SAS STAT. SAS EM relies on defaults that are well-researched parametrical combinations, which provide a good starting point in building models. These options with the flexibility provide consultants a good environment to build models at a faster pace with ease. On the other hand SAS EM diagnostics is not flexible as some of the traditional displays using Base SAS or some other competing packages.

Conclusions & lessons

LR modelling and its analysis for final model derivation is recommended. Several other works such as Nugyen et.al.(2002), provide support to such a recommendation, albeit in the context of other data sets. LR use linear combinations of variables, though it is not very adept at modelling nonlinear interactions. Given its linearity, LR is simpler to understand and interpret.

Predictions of very high quality are obtained using NN. There being many possible NN architectures, combined with the large choice of input variables makes structuring an NN a complex task. Data search engines allow testing of number of possible combinations to generate best-fit models. This work proposes the use of NN as the outcome prediction tool in conjunction with LR as an initial model derivation tool.

SAS EM's selection of techniques and methods provide an ease-of-use, shortening the development time significantly. An average of a week was the time spent in building a model from data that had undergone the exploratory data analysis step.

Data-mining infrastructure to improve customer retention, besides meaning effective churn models, must also include process changes to undertake more effective retention efforts, intelligence transfer in building these retention efforts and an ability to initiate changes in implementation through a feedback process.

References

- Fedenczuk, L. L. (2001)**, "To Neural or Not to Neural? – This is the question", Gambit Consulting Ltd., Proceedings of the SUGI27 2001.
- Lanubile, F. et. al. (1997)**, "Evaluating predictive quality models derived from software measures: lessons learned", J.Systems Software, Elsevier Science Inc.
- Nguyen, T. (2001)**, "Comparison of prediction models for adverse outcome in paediatric meningococcal disease using artificial neural network and logistic regression analyses", Journal of clinical epidemiology.
- SAS Institute (2001)**, "Enterprise Miner Reference", SAS.

Contact Information

Your comments and questions are valued and encouraged.

Contact the author at:

Arun Parekkat,
SPSInfoquest
4-6 Spicer street,
St.Albans,
Hertfordshire – AL3 4PQ
arun.parekkat@spsinfoquest.com