Table of Contents

International Journal of Conceptual Structures and Smart Applications

Volume 5 • Issue 2 • July-December-2017 • ISSN: 2166-7292 • eISSN: 2166-7306

An official publication of the Information Resources Management Association

Research Articles

The *International Journal of Conceptual Structures and Smart Applications* is indexed or listed in the following: Cabell's Directories; DBLP; Google Scholar; INSPEC; MediaFinder; The Standard Periodical Directory; Ulrich's Periodicals Directory

International Journal of Conceptual Structures and Smart Applications
Volume 5 • Issue 2 • July-December 2017

Case Studies in Applying Data Mining for Churn Analysis

Susan Lomax, University of Salford, Salford, UK
Sunil Vadera, University of Salford, Salford, UK

ABSTRACT

The advent of price and product comparison sites now makes it even more important to retain customers and identify those that might be at risk of leaving. The use of data mining methods has been widely advocated for predicting customer churn. This paper presents two case studies that utilize decision tree learning methods to develop models for predicting churn for a software company. The first case study aims to predict churn for organizations which currently have an ongoing project, to determine if organizations are likely to continue with other projects. While the second case study presents a more traditional example, where the aim is to predict organizations likely to cease being a subscriber to a service. The case studies include presentation of the accuracy of the models using a standard methodology as well as comparing the results with what happened in practice. Both case studies show the significant savings that can be made, plus potential increase in revenue by using decision tree learning for churn analysis.

KEYWORDS

Business Intelligence, Churn Prediction, Data Mining, Decision Trees, Knowledge Discovery

DOI: 10.4018/IJCSSA.2017070102

INTRODUCTION

Organizations in many different domains such as wireless telecommunication and the telecommunication industry (Huang et al., 2012, Keramati, 2014, Mahajan et al., 2015), mobile phone (Kirui et al., 2013), internet service providers (Khan et al., 2010), energy providers and other industries such as insurance, retail banking (Mutanen et al., 2006), financial services and supermarkets are having increasing difficulty in attracting and retaining customers as reported by Shandiz (2015). This is in part owing to customers being able to access information regarding brands, products and price comparisons on many internet comparison websites (Mahajan et al., 2015). The cost of acquiring new customers is higher than retaining old ones (Kirui et al., 2013) and small changes in the retention rate have been shown to have significant impact on businesses (Van den Poel & Larivie`re 2004; Larivie`re & Van den Poel 2005). For example, in the banking industry, Reichheld & Sasser (1990) and Nie et al. (2011) conclude that a bank is able to increase profits by 85% as a result of a 5% improvement on its retention rate.

Thus, a key part of any business these days is to manage customer churn; that is, to avoid losing its customers and build its customer base (Mattison 2005; Tsai & Lu 2010; Nie et al., 2011; Kirui et al., 2013).

A number of authors have advocated the use of data mining techniques to develop models to predict possible churn (Kirui et al., 2013). Several methods have been proposed including decision trees, neural nets, K Nearest Neighbour, logistic regression, random forests, SVM, linear and quadratic discriminate analysis, GA, Markov model, cluster analysis and optimization (Hadden et al., 2005; Better et al., 2008; Nie et al., 2011; Shandiz 2015). Most of these studies provide useful results on benchmark data, showing the potential for applying these methods for prediction of churn in practice. The experiences and lessons learned from applying data mining for predicting customer churn are seldom reported. This is perhaps owing to commercial confidentiality. It can also be difficult to find public datasets which can be used for churn prediction owing to business privacy and confidentiality therefore case studies presented using a variety of features are particularly useful (Kirui et al., 2013).

Thus, this paper presents two case studies in applying data mining for predicting customer churn. Although limited to two case studies, the use of non-traditional data in one of the case studies attempts to address this issue.

BACKGROUND

Churn, also known as turnover, defection or attrition is the loss of clients or customers. Many domains such as banks, mobile phone companies, internet service providers and supermarkets use churn analysis and churn rates as a key business metric as it has been shown that the cost of retaining an existing customer is less than the cost of acquiring new customers (Wei & Chiu 2002; Hung et al., 2006; Huang et al., 2012). These existing customers tend to purchase more than new customers and it is more efficient to deal with existing customers than dealing with new customers (Fornell & Wernerfelt 1987; 1988; Reichheld & Sasser 1990; Bolton 1998).

Churn analysis identifies or attempts to identify those customers who are most likely to churn. In the case of customers who have used an organization for a longer period of time, their churning can lead to a greater revenue loss than customers who have not been with the organization as long (Bolton 1998). It therefore makes sense to try to identify customers and then to develop strategies in order to retain these customers.

There are two approaches to combat customer churn. Untargeted relies on a good product and uses mass advertising to increase brand awareness and loyalty, therefore retaining customers. The second is targeted, which relies on identifying customers who have a high likelihood of churning. These customers are then given incentives or a customized service plan in order to convince them to stay (Ngai et al., 2009; Khan et al., 2010).

Organizations are now focusing on Customer Relationship Management (CRM). Customer retention is the main concern of CRMs (Rygielski et al., 2002). Unknown behaviours of customers are important with customer satisfaction an essential condition for retaining customers. Using existing customer data is a good source for decision making (Tsai & Lu, 2010; Ngai et al., 2009; Nie et al., 2011).

Data mining techniques have been used extensively to develop models to predict possible churn with satisfactory performances (Kirui et al., 2013). Several methods have been used over time such as decision trees, neural nets, K Nearest Neighbour, logistic regression, random forests, SVM, linear and quadratic discriminate analysis, GA, Markov model, cluster analysis and optimization (Hadden et al., 2005; Better et al., 2008; Nie et al., 2011; Shandiz 2015).

Mutanen et al. (2006) has some success predicting those likely to churn in the retail banking domain using logistic regression. Kirui et al. (2013) improves accuracy and high true positive rates using call traffic figures and customer profiles for the mobile phone industry. This is achieved using Naïve Bayes, Bayesian Networks and decision trees. Nie et al. (2011) report accuracy rates of logistic regression models in the 80 to 89% range on credit card data.

There have been a number of notable studies examining the use of data mining methods to solve problems associated with churn prediction. However, this paper focuses on case studies to show how non-traditional data can be used to predict churn. The reader is referred to a comprehensive literature review and classification by Ngai et al. (2009) which classifies CRM tasks, including customer retention, in terms of the CRM task, data mining functions and techniques used. Additionally, Tsai and Lu (2010) present a useful survey on data mining techniques and reviews a number of studies compared in terms of domains, pre-processing techniques and prediction techniques used in churn analysis.

CASE STUDIES

This section presents the two case studies. The first is a digital agency and the second is a provider of intelligence, research and information products and services to the public sector. Both of them are based in the UK.

In each case, we present what was involved in preparing the data, the methodology adopted, the main findings along with application of the models and the validation of the results.

Case Study 1

This dataset comprises of orders which have been carried out over the last 15 years. This dataset has been extracted from a system designed to house the data. It has been rarely used for any purpose other than keeping track of orders through the financial system apart from extracting only recent month's orders to examine for accounting purposes.

Data Preparation

In order to determine whether using order data could predict churn in the same way as user activity, a dataset has been created from order data. The dataset has been created from only 2 sources. The main data has been extracted from a retired administration system. This system was designed to be an internal management system to capture core data in relation to clients' orders such as history of projects in progress and completed projects. The file extracted comprises of one row per order with information about the organization placing the order, the division undertaking the work, the type of work being undertaken and its value along with text descriptions of the work. In addition, a separate file with details of the organization requesting the work has been appended to this file. This relates to the type of organization it is and is entered by the users from a drop-down list. There are 32 different descriptions to choose from giving 175 combinations of descriptions. In order that these multiple entries could be included in the analysis, the multiple entries were aggregated to one of 11 groups using the most descriptive of the entered descriptions.

A class attribute has been added to the dataset to simulate subscription-like behaviour. A list of organizations which have a current project within the new administrative system has been obtained. When an organization is on this list the class attribute is assigned 1, otherwise it is assigned 0. This provides information with regard to 'membership'. The file is imbalanced with 87.8% of orders from organizations which do not have a current project and therefore are designated as having no relationship with the Company. The number of examples in the dataset is 11,288.

Methodology

A wide range of data mining methods, ranging from logistic regression, to neural networks could be used for predicting customer churn. Each has its merits, and the one selected was the use of decision tree learning, mainly because these are simple and transparent. In addition, many researchers cite the use of decision tree analysis as a good method for classification given that decision trees implicitly perform feature selection and they are easy to interpret to non-experts (Deshpande 2011). These are important issues here as it will be necessary to explain outcomes of the churn analysis.

The open source data mining software R using Rattle as an interface has been used as the trees produced using this software are less complicated and more compact than some other implementations (such as in WEKA). The dataset created was imbalanced and it was felt the imbalance would affect the performance of the algorithm. Using under-sampling, 10 datasets have been created. Each of the 10 datasets have all the examples belonging to class 1 with examples belonging to class 0 being drawn at random to balance the datasets. After some initial experimentation, 49 attributes were selected to be used to predict whether an organization has a current project or not. Each file was split by the software into a 70% training set from which the model was built and 30% testing set which was used to determine the accuracy of the model.

Table 1 presents the attributes which were chosen by the algorithm with the number of times chosen and the root attribute indicated. The accuracy is reported for each of the ten files and the results are averaged out over these 10 files. The overall accuracy averaged over the 10 files is 95.29%. The accuracy for class 0 is 93.17% and the accuracy for class 1 is 96.8%. More rules are required to classify class 0. The average number of attributes in the models is 4. Models 1 and 9 have the highest overall accuracy with model 3 having the most consistent accuracy for both classes. Of all the attributes in the datasets there are 10 of them which are in the models, the rest are ignored by the algorithm. Two attributes are in all the models: org2013 and oMeanVal2015. Table 2 summarizes the results from all models.

The decision trees produced are very similar with the larger decision trees containing some of the smaller ones as sub trees. The main patterns are that the most recent the last order year is, the more likely the organization is to have a current project. If the mean value of orders in 2015 is relatively high or if the average number of orders are below 12 they are more likely to have a current project. If organizations have regular numbers of orders each year with the total number of orders being 50 or over, they are more likely to have a current project. The main overall pattern is that if the organization has high number of orders which are of low value they are more likely not to have a current project and if the organization has a low number of orders but they are of a high value then they are more likely to have a current project.

Applying the Models

The file used to create the ten datasets has been used to find out which organizations are predicted to have a project or are predicted to have no project. The rules generated by each of the 10 decision tree models have been applied to this order file. The aim is to use the rules

Table 1. The 10 Attributes Chosen by Algorithm with Number of Times Appearing in Models

			Models									
Attribute	Description	#times used	1	2	3	4	5	6	7	8	9	10
lastYear	Year in which the last order for that organization was placed	3	Y				Y				Y	
ordersOrg	The number of orders over all years of data placed by that organization	4	Y		Y						Y	Y
org2008	The number of orders placed in year by that organization	4	Y		Y						Y	Y
org2011	The number of orders placed in year by that organization	1	Y									
org2013	The number of orders placed in year by that organization	10	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
avgOrdersOrg	The average number of orders by that organization	3	Y								Y	Y
VerticalGroup	The company description (vertical on system) assigned to this group	2					Y				Y	
oMeanVal2009	The mean value of orders placed in year by that organization	1									Y	
oMeanVal2010	The mean value of orders placed in year by that organization	3			Y	Y						Y
oMeanVal2015	The mean value of orders placed in year by that organization	10	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

Table 2. Results from All Models with Averages Over the 10 Models

								TN	FP	TP	FN	Class 0	Class 1	Overall
DataFile	N	Training size	Testing size	Class 0	Class 1	Correct	Incorrect	0 correct	0 pred as 1	1 correct	1 pred as 0	Accuracy	Accuracy	Accuracy
1	2782	1947	835	1405	1377	815	20	383	17	432	3	95.56%	99.31%	97.60%
2	2743	1920	823	1366	1377	771	52	384	37	387	15	90.36%	96.12%	93.68%
3	2789	1952	837	1412	1377	800	37	390	17	410	20	95.64%	95.12%	95.58%
4	2712	1898	814	1335	1377	770	44	365	27	405	17	92.60%	95.80%	94.59%
5	2733	1913	820	1356	1377	782	38	379	31	403	7	91.82%	98.26%	95.37%
6	2710	1896	814	1333	1377	765	49	364	31	401	18	91.48%	95.51%	93.98%
7	2793	1955	838	1416	1377	789	49	386	31	403	18	91.97%	95.53%	94.15%
8	2738	1916	822	1361	1377	781	41	373	26	408	15	93.03%	96.32%	95.01%
9	2776	1943	833	1399	1377	811	22	398	21	413	1	94.72%	99.76%	97.36%
10	2727	1908	819	1350	1377	783	36	381	21	402	15	94.49%	96.27%	95.60%
Average	**2750**	**1925**	**826**	**1373**	**1377**	**787**	**39**	**380**	**26**	**406**	**13**	**93.17%**	**96.80%**	**95.29%**

File 3 produces the tree where the accuracy for each class is balanced: File 2, 6, 7 and 8 produce the same tree
Class 0: does not have a current project; Class 1: has a current project

generated by each model to predict both classes. For each subset created when applying the rules, the predicted class was assigned to the organization. There is up to 10 predictions for each organization with both classes potentially being predicted. To determine the overall prediction for each organization composite voting took place using the following formula:

$$Prob_i = N_i / M \tag{1}$$

where Prob_i is the value used to determine class assigned to organization,
N_i is the number of predictions for class *i* and
M is the number of models the organization appeared in
If $\text{Prob}_i > \text{Prob}_j$ then class predicted is *i* otherwise class predicted is *j*

Next the predicted class is compared to the actual class; if they are the same attribute 'match' is assigned "yes" otherwise 'match' is assigned "no". The orders where the value for match is "no" are the ones which are of interest. The accuracy for predicting class 0 (no current project) is 89.74% and the accuracy for predicting class 1 (current project) is 89.65%.

Of the 896 organizations, 90 have predictions which do not match the actual class and it is these organizations which are of interest. 81 organizations do not have a current project but have been predicted to have one and 9 organizations have a current project but have been predicted as not having one. Those who have been predicted as having a current project when they have not means that given their characteristics they should have a current project. They could be encouraged to start a new project and restart the relationship. If each of the 81 organizations were to place an order, using the mean value of their past orders, this would obtain revenue of over £900,000. Those who have been predicted as not having a current project, when they in fact have a current project, could indicate that they wish to terminate the relationship. If this were to be prevented and each of these placed an order, using the mean value of their orders this would be a value of over £165,000, giving a total increase of revenue of just over £1million.

Validation of Results

In order to check the results of the churn analysis, the organization list has been compared to a list of organizations who have either had a project completed or have started a new project since July 2016. Even though there has been only a short time where the results can be verified there are some good positive results.

There were 9 organizations predicted not to have a current project when in fact they did. Of these 7 of them have completed projects and have no current project. They are not likely to stay in the relationship and have no project opportunities underway which shows an accuracy of 77%. Using the mean value of orders, the estimated loss of approximately £29,000. Of the other 2 organizations, they are likely to stay in the relationship with other project opportunities. These projects have an approximate value of £50,000.

There were 81 organizations predicted to have a current project when in fact they do not have a project. The list of organizations was given to the sales and marketing department in order that they might use it to target new business from existing customers. Of these, 7 now have a current project whose values are over £400,000, 18 organizations have project opportunities to the value of over £500,000 and 3 organizations who, as they are partners could potentially have orders to the value of approximately £25,000. This gives a total of 28 organizations which now have or will have a new project which is a successful prediction of 34.56%. This will bring in an approximate value of £1 million. This leaves 52 organizations of which 48 have been marked as possible opportunities to a value of over £500,000 (using mean values of previous orders). Of the 81 organizations, there are only 4 where there is nothing likely to happen with regard to new projects.

Case Study 2

The data used for case study 2 is the metadata and transactions from 2012 to 2015 from a document database focused on public and social policy and practice. Its purpose is to help subscribers make informed decisions to improve services and to understand policy environment. It provides knowledge services to local authorities, public agencies, research consultancies and commercial organizations across the UK. It is important therefore to be able to predict if any of these organizations are likely to churn and investigate what steps can be taken to prevent this. This will improve service to all existing members.

Data Preparation

The dataset has been created from a list of activity on the database. This shows user downloads and searches by type. It includes subscription data which gives a list of current subscribers and their subscription amounts. Transaction files comprise of one row per transaction with information such as the user name, organization to which the user belongs and the type of transaction such as download document, emailed document, basic search, advanced search and external link. To indicate whether a transaction belongs to a current member a class attribute was added using the subscription information. If an organization has a current subscription the class attribute value was assigned 1 otherwise it was assigned 0. Counts by transaction type per user and organization were calculated and added to the file, for example how many times had a user downloaded a document or how many times a user had searched the database over the period of the data.

Finally, download transactions from 2013 and 2014 were extracted out of the whole dataset and used to create the file for churn analysis as these are the transactions for complete years. The number of transactions by organization in 2013 and 2014 were counted for each organization and appended to the file. In order to prevent possible correlation, it was decided to have only those attributes which comprise of counts of different types of transactions by organizations and users. The file contains 35,214 examples with a class distribution of 83.3% for class 1 (current member) and 16.7% for class 0 (not a subscriber). The average number of transactions by organization is 827 in 2013 and 742 in 2014 with most organizations carrying out up to 500 transactions in both years. The maximum number of transactions by an organization was 2,133 in 2013 and 2,226 in 2014.

Methodology

The main task is to classify whether an organization is not a current subscriber to the database. The same data mining method decision tree learning has been used as before. Unlike the other dataset the imbalanced dataset did not cause any issues with the decision tree algorithm so it was not necessary to balance the data using sampling methods. It is important to note that although decision trees are very good at determining useful attributes, when there are too many in the dataset, it is useful to reduce the number of attributes in the dataset. Especially where there could be correlation between attributes. However, it is important not to miss potentially useful information. In order to eliminate these problems, different combinations of attributes were used to produce 9 different models. Each of the models can be used in a simple voting system when applying them. Table 3 lists all attributes used with which file they were used in, along with whether they were present in each tree.

Each file was split by the software into a 70% training set from which the model was built and a 30% testing set which was used to determine the accuracy of the model. The nine different attribute combinations used to induce the models are presented in Table 4.

All models apart from one have an overall accuracy of over 90%. The model which has a lower accuracy also predicts poorly the accuracy of the target class 0: not a current subscriber. The models which predict this class the best use mainly the transactions by the organizations in 2013 and 2014 or the overall transactions by type carried out by organizations. The most common root attribute is transaction type downloads with a threshold by organization in 2013 of 834 and 93 in 2014. There are 2 models which have a different root: searches by user in 2013 and transactions by organization in 2013. The highest number of times an attribute has been used in the tree construction is 5. There are 5 attributes which have appeared many times with 1 attribute appearing 4 times. There are some attributes which never appear in any of the trees.

Overall it is shown that there are many combinations of transactions which can indicate whether an organization is likely to churn. The most likely description of the pattern to explain churn would be extreme numbers of transactions; for example, low numbers of one transaction type with higher numbers of another transaction type indicate that an organization might churn.

Applying the Models

In order to apply the models, the whole dataset was used to extract out all transactions where the class attribute is 1 i.e. the transaction was completed by an organization who has a current subscription. The rules from each model, which resulted in an allocation of class 0, were implemented. The resulting subsets extracted for each model were combined

Table 3. Frequency of Attribute Usage

		How often	1		2		3		4		5		6		7		8		9	
Attribute	Description of attribute	chosen	In File	In Tree	In File	In Tree	In File	In Tree	In File	In Tree	In File	In Tree	In File	In Tree	In File	In Tree	In File	In Tree	In File	In Tree
pcactiveUsers	Percent of active users in an organization	1			1	1														
searchOrg	Number of searches by organization	3	1	1	1	1							1	1						
searchUser	Number of searches carried out by user	1	1		1								1	1						
dlOrg	Number of docs downloaded by org	3	1	1	1	1							1	1						
dlUser	Number of docs downloaded by user	0	1		1								1							
extLinkOrg	Number of external links followed by org	1	1	1	1								1							
extLinkUser	Number of external links followed by user	0	1		1								1							
advancedOrg	Number of advanced searches by org	2	1	1	1								1	1						
advancedUser	Number of advanced searches carried out by u	0	1		1								1							
basicOrg	Number of basic searches by org	1	1		1								1	1						
basicUser	Number of basic searches carried out by user	0	1		1								1							
sUser2013	Number of searches carried out by user 2013	1			1		1	1			1						1			
sUser2014	Number of searches carried out by user 2014	0			1		1				1						1			
sOrg2013	Searches by organization 2013	4			1				1	1	1	1			1	1			1	1
sOrg2014	Searches by organization 2014	5			1	1			1	1	1	1			1	1			1	1
dlOrg2013	Number of docs downloaded by org 2013	5			1	1			1	1	1	1			1	1			1	1
dlOrg2014	Number of docs downloaded by org 2014	5			1	1			1	1	1	1			1	1			1	1
dlUser2013	Number of docs downloaded by user 2013	1			1		1	1			1						1			
dlUser2014	Number of docs downloaded by user 2014	0			1		1				1						1			
exlOrg2014	Number of external links followed by org 2014	5			1	1			1	1	1	1			1	1			1	1
exlOrg2013	Number of external links followed by org 2013	2			1				1		1				1	1			1	1
exlUser2014	Number of external links followed by user 2014	1			1		1				1						1	1		
exlUser2013	Number of external links followed by user 2013	0			1		1				1						1			
trans2013	Total number of transactions by org in 2013	1	1		1				1		1						1	1		
trans2014	Total number of transactions by org in 2014	5	1	1	1	1			1	1	1	1					1	1		
hasSub	Whether a current subscriber or not		1		1		1		1		1		1		1		1		1	
	Total Attributes in file/tree		12	5	25	8	6	2	8	6	14	6	10	5	6	6	8	3	6	6

Table 4. Summary of Models Used in Analysis

ID	Description of Tree	Overall Accuracy	Accuracy for Class 0	Accuracy for Class 1
1	Counts from all years with total number of transactions by organizations in 2013 and 2014	98.6	94.5	99.4
2	All counts calculated plus percent of active users in organizations	98.2	91.7	99.5
3	Counts for users for 2013 and 2014	84.4	6.5	100.0
4	Counts for organizations for 2013 and 2014 with total number of transactions by organizations in 2013 and 2014	98.0	91.7	99.3
5	Counts for 2013 and 2014 both users and organizations with total number of transactions by organizations in 2013 and 2014	98.0	91.7	99.3
6	Counts of the transactions for all years	98.3	92.8	99.4
7	Counts for 2013 and 2014 both users and organizations	97.8	92.1	98.9
8	Counts for users for 2013 and 2014 with total number of transactions by organizations in 2013 and 2014	98.3	96.0	98.8
9	Counts for organizations for 2013 and 2014	97.8	92.1	98.9

resulting in one subset per model. Finally, a simple voting scheme was employed to determine whether the organization is predicted to churn. If an organization appeared in a subset, that organization receives 1 vote for that model. The likelihood of churning is greater the more votes an organization obtains.

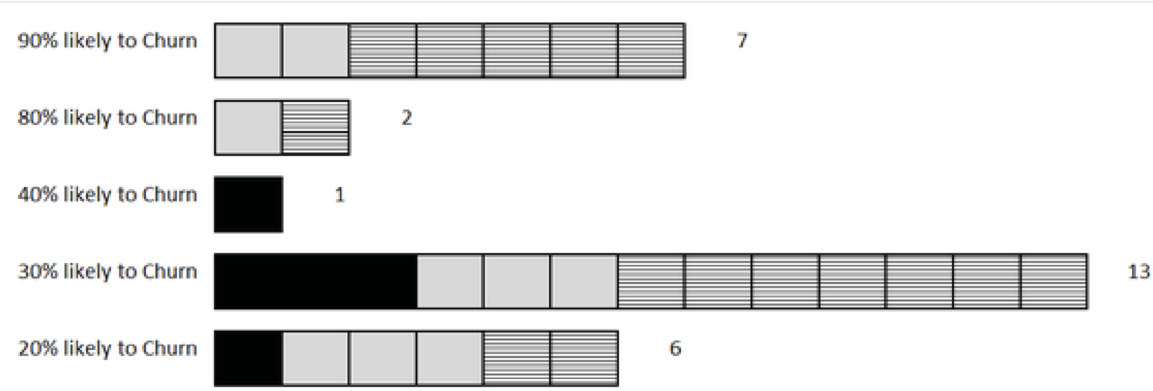
Of the 86 organizations in the dataset, 57 organizations appeared in at least one of the subsets of the decision tree models. However, 19 of these were in the model with low accuracy. This leaves 38 organizations which is 44% of the organizations. Figure 1 presents the number of organizations, colour-coded to show subscription levels and the likelihood to churn. This is calculated using the number of models that organization appears in.

If every organization whose probability to churn is 20% or above, churned, the value of the subscriptions lost would be over £80,000.

Validation of Results

In order to check the results of the churn analysis, the organization list has been compared to a list of organizations who have churned in the 18 months since the data was obtained. Of the 57 organizations predicted to churn, 28% (16) have churned or attempted to churn.

Figure 1. Number of organizations predicted to churn by level of subscription. Coded to include subscription level: dark > 5000; patterned > 1500; light <= 1500 90% likely: 8 trees; 80% likely: 7 trees; 40% likely: 4 trees; 30% likely: 3 trees; 20% likely: 2 trees.



The total lost in subscriptions from these 16 organizations is over £40,000, however 3 of the organizations were offered a new subscription deal so the actual loss is less than this.

There were 29 organizations who were not predicted to churn. Of these 14% (4) have actually churned. However, one of these organizations has 2 departments, one of which has not churned, one organization was offered a new subscription deal and another organization was closed by the government which would have been hard to predict. Again, although the loss of subscription is just over £23,000, with the accepted offer of a new subscription deal the actual loss is less than this. This shows that the predictive models are accurate and the Company can use them to successfully predict churn in the future.

CONCLUSION

This paper presents two case studies involving churn prediction.

The first case study used a non-traditional dataset comprising of orders of work of a Company. Datasets typically used for churn prediction tend to be customer service logs, complaint data, bill payment data and customer demographics. To see if churn prediction is possible on different data, 'membership' was imposed on this dataset in the form of determining if any of the organizations had a current project or not, simulating membership. Customer demographics are represented by having organization descriptions included in the dataset and work order details represent customer service logs or contractual data. On applying the model learned, an accuracy of over 89% was obtained. This identified 81 organizations that no longer had orders but had characteristics of organizations which had a current project, making them prime candidates to be targeted.

In the second case study, the dataset comprised of activity carried out by subscribers to a document database in the form of counts of the different activities on offer. Although the behaviour of the individual users did not predict churn with much success, when the activities were aggregated to organization level, they became very good at predicting churn. All models except one have an overall accuracy of over 90%. In order to validate the models, a list of organizations who had churned or attempted to churn during the course of data analysis was obtained. There were 20 organizations on the list. Of these 16 were predicted to churn, the other 4 were not predicted to churn. This shows that the predictive models are accurate and the Company can use them to successfully predict churn in the future.

In conclusion, these two case studies show that decision tree learning can result in good models for predicting customer churn and that using non-traditional datasets, such as company orders, can be successfully used for developing models that predict churn.

ACKNOWLEDGMENT

This work was supported by Innovate UK via a Knowledge Transfer Partnership [1008552]. The authors are grateful to Amanda Bolton for supporting this work and to the referees for their constructive comments, which have helped to improve the presentation of this paper.

REFERENCES

Better, M., Glover, F., Kochenberger, G., & Wang, H. (2008). Simulation optimization: Applications in risk management. *International Journal of Information Technology & Decision Making*, *7*(4), 571–587. doi:10.1142/S0219622008003137

Bolton, R. N. (1998). A dynamic model of the duration of the customers relationship with a continuous service provider: The role of satisfaction. *Marketing Science*, *17*(1), 45–65. doi:10.1287/mksc.17.1.45

Deshpande, B. (2011) Four Key Advantages of Using Decision Trees for Predictive Analytics. *Simafore.com*. Retrieved 13.7.2017 from http://www.simafore.com/blog/bid/62333/4-key-advantages-of-using-decision-trees-for-predictive-analytics

Fornell, C., & Wernerfelt, B. (1987). Defensive marketing strategy by customer complaint management: A theoretical analysis. *JMR, Journal of Marketing Research*, *24*(November), 337–346. doi:10.2307/3151381

Fornell, C., & Wernerfelt, B. (1988). Model for customer complaint management. *Marketing Science*, *7*(Summer), 271–286.

Hadden, J., Tiwari, A., Roy, R., & Ruta, D. (2005). Computer assisted customer churn management: State-of-the-art and future trends. *Computers & Operations Research*, *34*(10), 2902–2917. Retrieved from https://www.linkedin.com/pulse/application-machine-learning-data-mining-methods-rate-mohammad doi:10.1016/j.cor.2005.11.007

Huang, B., Kechadi, M. T., & Buckley, B. (2012). Customer Churn Prediction in Telecommunications. *Expert Systems with Applications*, *39*(1), 1414–1425. doi:10.1016/j.eswa.2011.08.024

Hung, S. Y., Yen, D. C., & Wang, H. Y. (2006). Applying Data Mining to Telecom Churn Management. *Expert Systems with Applications*, *31*(3), 515–524. doi:10.1016/j.eswa.2005.09.080

Keramati, A., Jafari-Marandi, R., Aliannejadi, M., Ahmadian, I., Mozaffari, M., & Abbasi, U. (2014). Improved churn prediction in telecommunication industry using data mining techniques. *Applied Soft Computing*, *24*, 994–1012. doi:10.1016/j.asoc.2014.08.041

Khan, A. A., Jamwal, S., & Sepehri, M. M. (2010). Applying data mining to customer churn prediction in an internet service provider. *International Journal of Computers and Applications*, *9*(7), 8–14. doi:10.5120/1400-1889

Kirui, C., Hong, L., Cheruiyot, W., & Kirui, H. (2013). Predicting Customer Churn in Mobile Telephony Industry using Probabilistic Classifiers in Data Mining. *IJCSI International Journal of Computer Science Issues*, *10*, 165–172.

Lariviere, B., & Van den Poel, D. (2005). Predicting Customer Retention and Profitability by using Random Forests and Regression Forests Techniques. *Expert Systems with Applications*, *29*(2), 472–484. doi:10.1016/j.eswa.2005.04.043

Mahajan, V., Misra, R., & Mahajan, R. (2015). Review of Data Mining Techniques for Churn Prediction in Telecom. *JIOS*, *37*(2), 183–197.

Mattison, R. (2005). *The Telco Churn Management Handbook*. Oakwood Hills, Illinois, USA: XiT Press.

Mutanen, T., Ahola, J., & Nousiainen, S. (2006) Customer Churn Prediction - A Case Study in Retail Banking. In *Proceedings of ECML/PKDD 2006 Workshop on Practical Data Mining: Applications, Experiences and Challenges*, Berlin, September 22 (pp. 13-19).

Ngai, E. W. T., Xiu, L., & Chau, D. C. K. (2009). Application of Data Mining Techniques in Customer Relationship Management: A Literature Review and Classification. *Expert Systems with Applications*, *36*(2), 2592–2602. doi:10.1016/j.eswa.2008.02.021

Nie, G., Rowe, W., Zhang, L., Tian, Y., & Shi, Y. (2011). Credit Card Churn Forecasting by Logistic Regression and Decision Tree. *Expert Systems with Applications*, *38*(12), 15273–15285. doi:10.1016/j.eswa.2011.06.028

Reichheld, F. F., & Sasser, W. E. (1990). Zero Defections: Quality Comes to Services. *Harvard Business Review*, *14*(March), 495–507. PMID:10107082

Rygielski, C., Wang, J.-C., & Yen, D. (2002). Data Mining Techniques for Customer Relationship Management. *Technology in Society*, *24*(4), 483–502. doi:10.1016/S0160-791X(02)00038-6

Shandiz M. A. (2015) Application of machine learning and data mining methods for churn rate prediction of customers.

Tsai, C.-F., & Lu, Y.-H. (2010). Data mining techniques in customer churn prediction. *Recent Patents on Computer Science*, *3*, 28–32.

Van den Poel, D., & Lariviere, B. (2004). Customer attrition analysis for financial services using proportional hazard models. *European Journal of Operational Research*, *157*(1), 196–217. doi:10.1016/S0377-2217(03)00069-9

Wei, C. P., & Chiu, I.-T. (2002). Turning telecommunication calls details to churn prediction: A data mining approach. *Expert Systems with Applications*, *31*(2), 103–112. doi:10.1016/S0957-4174(02)00030-1

Susan Lomax, PhD, MBCS has recently completed a project as a Knowledge Transfer Partnership Associate which aimed to demonstrate the value of data analysis using data mining methods. Her PhD focused on Cost-Sensitive Decision Tree Learning and was part of the Data Mining & Pattern Recognition Research Centre. She is a professional member of the BCS and a member of the IEEE Computer Society. She received an MSc by research from the University of Salford in 2006 and a First-Class Honours degree in Computing at the University of Bolton in 2004. She was also in receipt of the Athena Prize in 2004. She worked at the University of Manchester between 2007 and 2009 as a Computing Research Assistant at the Cathie Marsh Centre for Census and Survey Research.

Sunil Vadera is a Professor of Computer Science and the Dean of the School of Computing, Science and Engineering at the University of Salford. He is a Fellow of the British Computer Society, a Chartered Engineer (C. Eng) and Chartered IT Professional (CITP). He holds a PhD in Computer Science from the University of Manchester which was awarded in 1992.His research in machine learning and its applications which has been published in outlets such as ACM Transactions on Knowledge Discovery from Data, ACM Computing Surveys, Computer Journal, Expert Systems, Foundations of Science, and IEEE Transactions of Power Systems.