Predicting Customer Churn in Mobile Telephony Industry Using Probabilistic Classifiers in Data Mining

Clement Kirui[1], Li Hong[2], Wilson Cheruiyot[3] and Hillary Kirui[4]

[1] School of Information and Communication Engineering, Central South University
Changsha, 410083, China

[2] School of Information and Communication Engineering, Central South University
Changsha, 410083, China

[3] Institute of Computer Science and Information Technology, Jomo Kenyatta University of Agriculture and Technology
Juja, Nairobi, Kenya

[4] School of Human Resource and Development, Jomo Kenyatta University of Agriculture and Technology
Juja, Nairobi, Kenya

Abstract

Customer churn in the mobile telephony industry is a continuous problem owing to stiff competition, new technologies, low switching costs, deregulation by governments, among other factors. To address this issue, players in this industry must develop precise and reliable predictive models to identify the possible churners beforehand and then enlist them to intervention programs in a bid to retain as many customers as possible. This paper proposes a new set of features with the aim of improving the recognition rates of possible churners. The features are derived from call details and customer profiles and categorized as contract-related, call pattern description, and call pattern changes description features. The features are evaluated using two probabilistic data mining algorithms Naïve Bayes and Bayesian Network, and their results compared to those obtained from using C4.5 decision tree, a widely used algorithm in many classification and prediction tasks. Experimental results show improved prediction rates for all the models used.

Keywords: *Customer churn, data mining, classification / prediction, decision tree, Naïve Bayes and Bayesian Network.*

1. Introduction

One of the key aspirations of every business is to build and maintain a loyal customer base. However, the mobile telephony industry is characterized by stiff competition, increasing innovation as a result of new technologies, low switching costs, and deregulation by governments [1]. These, among other factors, have contributed to the risk of customers switching to competitors with ease. Customer churn can be defined as the ceasing of a customer from the subscription to a service [1] and can be broadly classified into two: voluntary and involuntary churn. Voluntary churn occurs when a customer initiates termination of a service contract while involuntary churn occurs when customers are disconnected by the company for fraud, non-payment, or under-utilization of subscribed services. In this work we focus on voluntary churn.

Customer churn is a costly risk that if not managed carefully, may bring a company to its knees. Costs associated with customer churn include loss of revenue, costs of customer retention and reacquisition, advertisement costs, organizational chaos, as well as planning and budgeting chaos [1]. In addition, previous studies have shown that the cost of acquiring new customers is much higher than the cost of retaining the existing ones [1], [2], [3], [4]. Therefore, it makes business sense for players in this industry to identify the subscribers who are likely to leave the company beforehand and develop intervention strategies in a bid to retain as many customers as possible.

One of the strategies that can be used to achieve this is by developing predictive models that can reliably identify possible churners in the near future. In the recent past, data mining techniques have been used extensively to develop these models with satisfactory performance [2], [4], [5], [6], [7]. Because of the importance of precision in churn prediction, much research efforts have been channeled towards improving churn recognition rates of prediction models. One way of achieving this is by using appropriate feature sets that have high predictive power.

Features that have been used previously in the wireless telephony industry include call details, contractual data, customer service logs, complaint data, bill payment data, and customer demographics [3], [6]. Due to business confidentiality and privacy, it is difficult to find public datasets on churn prediction and thus there is a challenge of standardizing the feature sets to use. Wei and Chiu (2002) in [3] used subscriber contractual information and call pattern changes extracted from call details to propose features that include minutes of use, frequency of use, sphere of influence, and changes in the three feature subsets. Huang et al (2009) in [5] used demographic profiles, grants, account information, call details, service orders, Henley segmentation, payment and bill information, and telephone line information in the landline telecommunication sector.

In this paper, we present a new subset of features in order to improve the accuracy of customer churn prediction in the wireless telephony industry. The new features are categorized as contract-related features, call patterns description features, and calls pattern changes description features. To evaluate the features, we performed experiments using two probabilistic data mining algorithms Naïve Bayes and Bayesian Network, and their results compared to those obtained from using C4.5 decision tree, a widely used algorithm in many classification and prediction tasks. Experimental results show improved prediction rates for all the models used.

The rest of the paper is organized as follows: the next section discusses the evaluation criteria while section 3 outlines the methodology adopted. Experimental setup is discussed in section 4 and the results analyzed in section 5. We make a conclusion and future research in section 6, and finally we give a list of references.

2. Evaluation Criteria

To test and evaluate the features, we used 10-fold cross validation. In this process the initial data are randomly partitioned into 10 mutually exclusive subsets or “folds,” each of approximately equal size. Training and testing is performed 10 times. In iteration i, partition D_i is reserved as the test set, and the remaining partitions are collectively used to train the model. For classification, the accuracy estimate is the overall number of correct classifications from the 10 iterations, divided by the total number of tuples in the initial data. This provides a good indication of how well the classifier will perform on unseen data.

A confusion matrix similar to the one shown in table 1 is generated [8].

Table 1 Confusion matrix

		Predicted Class	
		CHURN	ACTIVE
Actual Class	CHURN	TP	FN
	ACTIVE	FP	TN

Definitions:

- *True positive (TP):* Number of positive cases correctly predicted.
- *False negative (FN):* Number of positive cases wrongly predicted as negative.
- *False positive (FP):* Number of negative cases wrongly predicted as positive
- *True negative (TN):* Number of negative cases correctly predicted.

From the confusion matrix the following measures, among others, can be obtained.

Accuracy: The percentage of correctly classified instances over the total number of instances. $Accuracy = \frac{TP+TN}{TP+FP+TN+FN}$

True positive rate (TPR) or sensitivity: fraction of positive instances predicted correctly. $TPR = \frac{TP}{TP+FN}$

False positive rate (FPR): fraction of negative instances wrongly predicted as positive. $FPR = \frac{FP}{TN+FP}$

Precision: fraction of records that actually turn out to be positive in the group the classifier has declared as positive. The higher the precision is, the lower the number of false positive errors committed by the classifier. $Precision = \frac{TP}{TP+FP}$

Recall: Fraction of positive instances correctly predicted by the classifier. Its value is equivalent to true positive rate. The higher the value of recall the fewer the number of instances misclassified as negative. $Recall = \frac{TP}{TP+FN}$

Our main focus in this study is to improve the recognition of churners (Recall /true positive rate) and reduce the false positive rate.

3. Methodology

Our approach consists of data sampling, data preprocessing, model construction, and model evaluation phases. Data sampling randomly selects a set of customers with the required information, according the definition of churn. The data preprocessing phase includes data cleaning and feature extraction steps. Data cleaning removes the irrelevant information which includes wrong spelling words caused by human errors, special mathematical symbols, missing values, duplicated information, and so on. The feature extraction step extracts a set of features or variables to be evaluated during prediction. In the model construction phase, we build a classification/prediction model that predicts the potential behavior of customers in the near future. These steps are described in the following subsections.

3.1. Data sampling

The dataset that we used was obtained from a European telecommunications company and were collected in a period of three months from August to October 1997. Originally, the dataset consisted of 112 attributes and 106,405 instances of which 5.6% were churners and the rest were active subscribers. We then derived the additional features and added to the original dataset. The skew in the data presented a class imbalance problem [9] which makes it hard for minority class (churn) recognition by some data mining algorithms [8]. We handled this problem by applying stratified random sampling [10] to both the original and the modified datasets. In this technique, the data is grouped into two homogeneous strata, CHURN and ACTIVE. Random sampling is then used in each stratum independently to obtain data of the required sample size. In this study, we maintained the size of CHURN at 100% and varied the size of ACTIVE from 10% to 100%. We generated datasets of the same sample sizes for both the original and the modified datasets.

3.2. Data preprocessing

In this work, we focused on data cleaning and feature extraction in this phase.

3.2.1. Data cleaning

Noise is the irrelevant information which would cause problems for the subsequent processing steps. Therefore, noisy data should be removed. This irrelevant information includes wrong spelling words caused by human error, special symbols like mathematical symbols and punctuation marks, missing values, duplicated information (e.g. the same attributes with the same values are in different tables of a database). This noise can be removed by finding their locations and using the correct values to replace them, or some times by deleting them if the missing values are too many.

3.2.2. Feature extraction

Feature extraction plays an important role in determining the performance of predictive models in the terms of prediction rates (high TP and low FP) for churn recognition. If a robust set of features can be extracted in this phase, the prediction rates can be significantly improved. However, obtaining such a good seat of features is not an easy task. Most of the feature sets that have been proposed so far in mobile telephony industry [4], [5] still have a room for improvement. In order to improve the prediction rates for churn recognition, based on the dataset obtained for this study, we present a new set of features for customer churn prediction in mobile telephony industry in this section.

The dataset consisted of two subsets: call traffic figures and customer profiles. We merged the two data subsets to obtain a complete feature set for every customer basing on the customer id and then we came up with the following attributes:

Customer profiles: describe the demographic grouping of customers basing on their common characteristics. We chose value segment, rate plan, payment type, credit score, area, and penalties for non-payment.

Traffic details: describe both incoming and outgoing traffic basing on call types and customer activity.

- **Calls to/from competition:** include monthly number of calls made or received, number of SMS, and duration of incoming and outgoing calls to competitors.
- **Fixed line calls:** describe monthly incoming and outgoing calls to fixed line subscribers. They include duration of calls and the number of calls.
- **International calls:** describe monthly outgoing and incoming calls from international callers. They include duration of calls, number of calls, and the number of SMS sent or received.
- **On-net calls:** describe monthly outgoing and incoming calls within the network. They include duration of calls, number of calls, and the number of SMS sent or received.
- **VAS calls:** describe monthly outgoing and incoming calls from value added services. They include duration of calls, number of calls, and the number of SMS sent or received.
- **Activity:** describe monthly traffic activity of every subscriber. It includes number of days for outgoing or incoming traffic and the number of events on weekdays and on weekends.
- **Distinct callers:** describe monthly number of distinct callers and subscribers contacted.

From the features described above, we derived the following additional features.

Contract-related features

Tenure: describe the length of service or duration, measured in months, between the contract starting date

and termination date; otherwise it measures the number of months since the contract started. Usually, a customer who has stayed with a service provider for long is unlikely to churn in the near future. It is computed as:
$IF\ (STATUS_i = \text{"CHURN"})\ TENURE_i = DEACTIVATION\_MONTH_i - FIRST\_RENEWAL\_DATE_i\ else\ TENURE_i = REF\_DATE_i - FIRST\_RENEWAL\_DATE_i$

Calls pattern description features

- **Total call duration:** Total number of minutes of outgoing calls made by a subscriber in a month, given as: $VOICE\_OUT\_MINS_i = DUR\_CMP\_OUT_i + DUR\_FIXED\_OUT_i + DUR\_INTER\_OUT_i + DUR\_ONNET\_OUT_i + DUR\_VAS\_OUT_i$
- **Number of voice calls:** Total number of outgoing voice calls made by a subscriber in a month, given as: $VOICE\_OUT\_CALLS_i = NUM\_CALLS\_CMP\_GSM\_OUT_i + NUM\_CALLS\_FIXED\_OUT_i + NUM\_CALLS\_INTER\_OUT_i$
- **Total number of calls:** Total number of outgoing calls made by a subscriber in a month, given as: $TOTAL\_OUT\_CALLS_i = VOICE\_OUT\_CALLS_i + SMS\_OUT\_CALLS_i + NUM\_MMS\_OUT_i + NUM\_VAS\_OUT_i$

Calls pattern changes description features

- **Change in minutes of use:** Change in total number of minutes of outgoing calls between sub-period *i*-1 and *i* (for *i* =2, …, *n*), computed as: $CHG\_VOICE\_OUT\_MINS_i = \frac{TOTAL\_OUT\_MINS_i - TOTAL\_OUT\_MINS_{i-1} + \alpha}{TOTAL\_OUT\_MINS_{i-1} + \alpha}$

  Where α is a very small number used to avoid division by zero. We used $\alpha = 0.001$.

- **Change in frequency of use:** Change in total number of outgoing calls between sub-period *i*-1 and *i* (for *i* =2, …, *n*), computed as:

  $CHG\_TOTAL\_OUT\_CALLS_i = \frac{TOTAL\_OUT\_CALLS_i - TOTAL\_OUT\_CALLS_{i-1} + \alpha}{TOTAL\_OUT\_CALLS_{i-1} + \alpha}$

- **Change in sphere of influence:** Change in distinct number of receivers called by the subscriber between sub-period *i*-1 and *i* (for *i* =2, …, *n*), computed as: $CHG\_DISTINCT\_CALLERS\_OUT_i = \frac{DISTINCT\_CALLERS\_OUT_i - DISTINCT\_CALLERS\_OUT_{i-1} + \alpha}{DISTINCT\_CALLERS\_OUT_{i-1} + \alpha}$

- **Change in subscriber activity:** Change in number of days a subscriber is active between sub-period *i*-1 and *i* (for *i* =2, …, *n*), computed as: $CHG\_ACTIVITY_i = \frac{ACTIVITY_i - ACTIVITY_{i-1} + \alpha}{ACTIVITY_{i-1} + \alpha}$

- **Change in duration of calls to competitors:** Change in total number of minutes of outgoing calls to competitors between sub-period *i*-1 and *i* (for *i* =2, …, *n*), computed as: $CHG\_DUR\_CALLS\_CMP\_OUT_i = \frac{DUR\_CMP\_GSM\_OUT_i - DUR\_CMP\_GSM\_OUT_{i-1} + \alpha}{DUR\_CMP\_GSM\_OUT_{i-1} + \alpha}$

- **Percentage of calls to competitors:** Percentage of number of outgoing calls to competitors against total number of outgoing calls between sub-period *i*-1 and *i* (for *i* =2, …, *n*), computed as: $PCNT\_CALLS\_CMP\_OUT_i = \frac{NUM\_CALLS\_CMP\_GSM\_OUT_i}{TOTAL\_OUT\_CALLS_i} \times 100\%$

- **Percentage of outgoing voice calls:** Percentage of number outgoing voice calls against total number of outgoing calls between sub-period *i*-1 and *i* (for *i* =2, …, *n*), computed as: $PCNT\_VOICE\_OUT\_CALLS_i = \frac{VOICE\_OUT\_CALLS_i}{TOTAL\_OUT\_CALLS_i} \times 100\%$

3.3. Model construction

There are many techniques that have been proposed for customer churn prediction in the telecommunications industry. In this study, we used two probabilistic data mining algorithms Naïve Bayes and Bayesian Network to evaluate the feature subsets that were proposed. Their results were compared with those obtained from using C4.5 decision tree, a widely used algorithm in many classification and prediction tasks. The three techniques are briefly discussed as follows:

3.3.1. C4.5 Decision Tree

The C4.5 algorithm [11] uses the ‘divide and conquer’ method to construct a model based on a tree structure. Nodes in the tree represent features, with branches representing possible values connecting the features. A leaf representing the class terminates a series of nodes and branches. Initially, the method starts to search an attribute with best information gain at root node and divide the tree into sub-trees. Similarly, each sub-tree is further separated recursively following the same rule. The partitioning stops if the leaf node is reached or there is no information gain. Once the tree is created, rules can be obtained by traversing each branch of the tree.

3.3.2. Naive Bayes

Naive-Bayes algorithm [8], [11] is based on the Bayesian theorem. It analyses the relationship between each feature and the class for each instance to derive a conditional probability for the relationships between the feature values and the class. We assume that X is a vector of instances where each instances is described by attributes $\{X_1,...,X_k\}$ and a random variable C denoting the class of an instance. Let x be a particular instance and c be a particular class. During training, the probability of each class is computed by counting how many times it occurs in the training dataset. This is called the prior probability $P(C=c)$. In addition to the prior probability, the algorithm also computes the probability for the instance x given c. Under the assumption that the attributes are independent this probability becomes the product of the probabilities of each single attribute. Naive Bayes has achieved good results in many cases even when this assumption is violated.

The probability that an instance x belongs to a class c can be computed by combining the prior probability and the probability from each attribute's density function using the Bayes formula:

$$P(C=c|X=x) = \frac{P(C=c)\prod_i P(X_i|C=c)}{P(X=x)}$$

The denominator is invariant across classes and only necessary as a normalizing constant (scaling factor). It can be computed as the sum of all joint probabilities of the enumerator:

$$P(X=x) = \sum_j P(C_j)P(X=x|C_j)$$

3.3.3. Bayesian Networks

A Bayesian Network [8], [12] is a combination of a directed acyclic graph of nodes and links, and a set of conditional probability tables. Nodes can represent features or classes, while links between nodes represent the relationship between them. Conditional probability tables determine the strength of the links. There is one probability table for each node (feature) that defines the probability distribution for the node given its parent nodes. If a node has no parents the probability distribution is unconditional. If a node has one or more parents the probability distribution is a conditional distribution where the probability of each feature value depends on the values of the parents.

4. Experimental setup

Two sets of experiments were performed in order to evaluate the predictive importance of the new features proposed. To begin with, we performed data preparation as discussed in sections 3.1 and 3.2.

In the first set of experiments, we used the modified sampled dataset only because our interest was to compare the performance of the new feature subsets. We tested the features basing on their call types. The call types that were identified include calls to competition (CMP), fixed line calls (FIXED), international calls (INTER), on-net calls (ONNET), and value added service calls (VAS). We then tested the new proposed features altogether (NEW). We also tested features that describe user activity (ACTIVITY + NUM_EVENTS) and customer profiles (CP).

To further assess the impact of the new features, we performed a second set of experiments in which we used information gain attribute selection technique to select the first 60 attributes with the highest information gain for both the original and the modified datasets.

For the two sets of experiments, we used three modeling techniques: C 4.5 decision tree, Naïve Bayes, and Bayesian Network to test the features.

5. Results and analysis

From the experiments, we obtained the false churn rate (FP) and true churn rate (TP) of each of the feature subset for all modeling techniques. Based on these pairs of FP and TP, the ROC curves were plotted as shown in figures 1 to 3. The x-axis represents the false churn rates (FP) while the y-axis represents the true churn rates (TP). Each ROC curve consists of a sequence of points, each of which presents a pair of prediction rates (FP, TP) for a specified sampling rate. Generally, an ROC curve which is closer to the left-top corner presents the better prediction result.

From results of experiment 1 (figures 1 to 3), the predictive importance of each feature subset has been shown. For instance, it can be seen that the feature subset NEW (representing the proposed features) plays the most significant role in churn prediction for all the modeling techniques used.

While there is a clearer distinction in importance of each feature subset while using C4.5 and Bayesian Network, most of the feature subsets perform almost equally when using Naïve Bayes, save for "customer profiles" and NEW. We can also see that as the sampling rate $\mathrm{CHURN/ACTIVE}$ decreases, i.e. the number of ACTIVE increases, both the false churn rate and the true churn rate generally reduce. This can be attributed to the fact that as the sample size increases the ratio of CHURN to ACTIVE also increases. We can also observe that the two probabilistic classifiers achieve higher true positive rates and false positive rates than the decision tree.

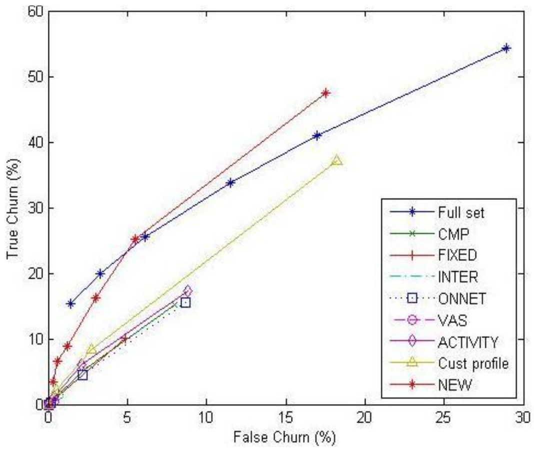

Fig. 1 C 4.5 Decision Tree results for each feature subset

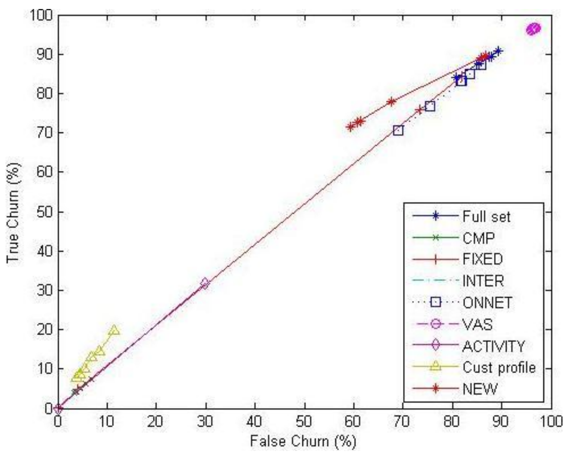

Fig. 2 Naive Bayes results for each feature subset.

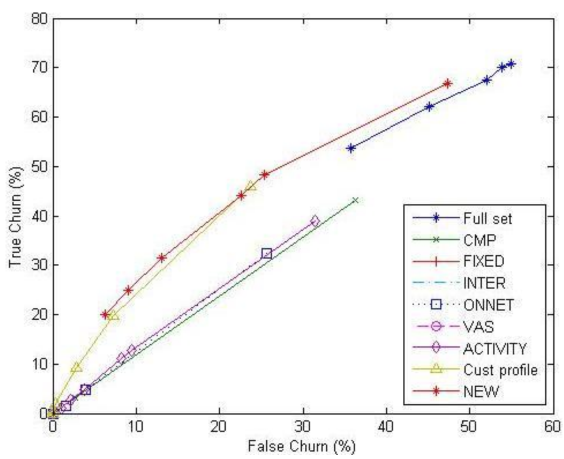

Fig. 3 Bayesian Network results for each feature subset.

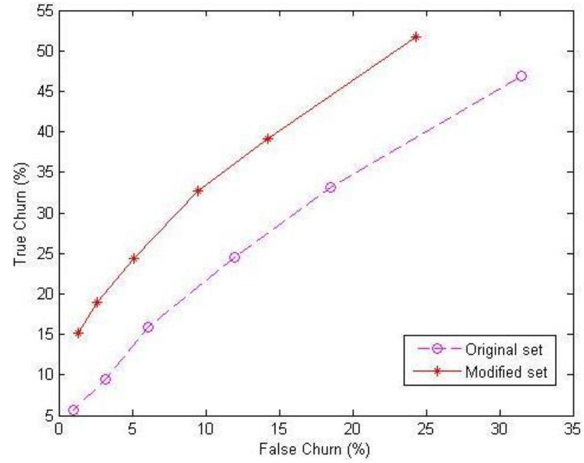

Fig. 4 C4.5 Decision Tree results for Information gain feature selection.

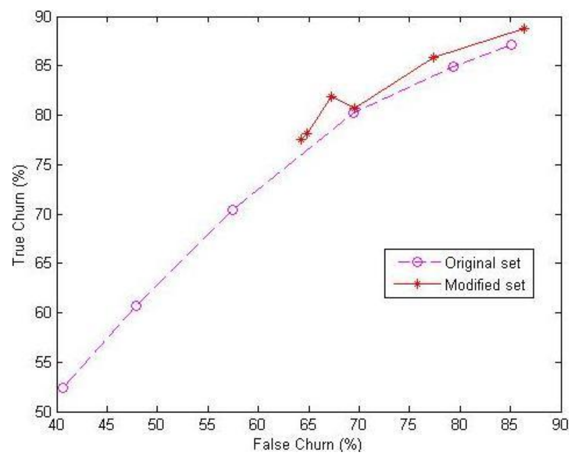

Fig. 5 Naive Bayes results for Information gain feature selection.

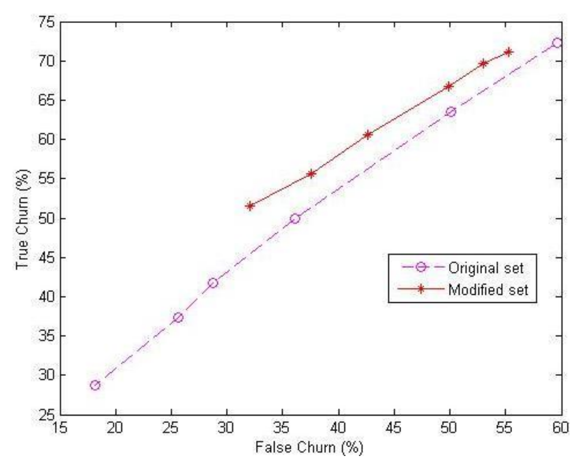

Fig. 6 Bayesian Network results for Information gain feature selection.

Information gain measure is based on information theory, which measures the information content of attributes [11]. Here, we used this attribute selection measure to select top 60 attributes with the highest information gain for both original and modified datasets. The number 60 was chosen merely because it represents almost half the number of attributes in both datasets. After the information gain of each attribute is calculated, the attributes are ranked in a descending order and the best ones selected as mentioned above. The selected attributes for various sample sizes were then tested using the three models and their true churn rate (TP) and false churn rate (FP) values recorded. Their ROC curves were then plotted and results shown in figures 4 to 6. From these results, it is clear that the attributes of the modified dataset have a higher significance in churn prediction than those of the original dataset. The improvement in prediction rates is attributed to the new features added to the feature set of the original dataset. This means that the new features are indeed of significant importance in churn prediction.

6. Conclusions and future work

Customer churn prediction is plays a central role in churn management in mobile telephony industry. In order to reduce the various costs associated with customer churn, it is imperative that mobile service providers deploy churn predictive models that can reliably identify customers who are about to leave. After the possible churners are identified, intervention strategies should be put in place with the aim of retaining as many customers as possible.

It is not uncommon for mobile service providers to have datasets with hundreds of features. However, not all these features are of significant importance in churn prediction. In order to improve the recognition of churners in prediction models, a careful selection of feature sets to be used should be done. In this study, we presented a set of new features categorized as contract-related, call pattern description, and call pattern changes description features that were derived from traffic figures and customer profile data. The features were evaluated using two probabilistic data mining algorithms Naïve Bayes and Bayesian Network, and their results compared to those obtained from using C4.5 decision tree, a widely used algorithm in many classification and prediction tasks. In order to further assess the impact of the proposed features, we ranked the feature sets of both the original and the modified datasets using information gain feature selection technique. Experimental results show improved prediction rates for all the models used. It was also observed that the two probabilistic classifiers achieve higher true positive rate than the decision tree. On the other hand, the decision tree performs better in overall accuracy. However, churn datasets are usually skewed and the minority class (CHURN) is usually the class of interest. Therefore, higher true positive rate and lower false positive rates are better measures than the overall accuracy.

Although churn prediction is important in churn management, there are many challenges that should be taken into consideration. For example, it is good to predict the possible churners beforehand, but it would be even better to know when these possible churners are likely to quit. This will help in determining appropriate intervention strategies to employ at any given time. Also, there is still room for improvement of prediction rates. In addition, churn datasets exhibit class imbalance problem whereby the class of interest is the minority. As stated previously, this makes it hard for some data mining techniques to recognize the minority class instances although they may achieve high overall accuracy. Therefore, there is a need to do further research on how address this issue. These are some of the issues we wish to address in our future research.

References

[1] R. Mattison, The Telco Churn Management Handbook, Oakwood Hills, Illinois: XiT Press, 2005.

[2] B. Huang, M. T. Kechadi and B. Buckley, "Customer churn prediction in telecommunications," *Expert Systems with Applications,* vol. 39, p. 1414–1425, 2012.

[3] C. P. Wei and I.-T. Chiu, "Turning telecommunication calls details to churn prediction: a data mining approach," *Expert Systems with Applications,* vol. 31, pp. 103-112, 2002.

[4] S. Y. Hung, D. C. Yen and H. Y. Wang, "Applying data mining to telecom churn management," *Expert Systems with Applications,* vol. 31, p. 515–524, 2006.

[5] B. Huang, T. Kechadi, B. Buckley, G. Kiernan, E. Keogh and T. Rashid, "A new feature set with new window techniques for customer churn prediction in land-line telecommunications," *Expert Systems with Applications,* vol. 37, p. 3657–3665, 2010.

[6] B. Huang, B. Buckley and T.-M. Kechadi, "Multi-objective feature selection by using NSGA-II for customer churn prediction in telecommunications," *Expert Systems with Applications,* vol. 37, p. 3638–3646, 2010.

[7] Y. Xie, E. N. Xiu Li and W. Ying, "Customer churn prediction using improved balanced random forests," *Expert Systems with Applications,* vol. 36, p. 5445–5449, 2009.

[8] P. N. Tan, M. Steinbach and V. Kumar, Introduction to

Data Mining, Pearson Education Asia Inc., 2006.

[9] J. Burez and D. V. d. Poel, "Handling class imbalance in customer churn prediction," *Expert Systems with Applications,* vol. 36, pp. 4626-4636, 2009.

[10] E. Babbie, The Practice of Social Research, Belmont, Califonia: Thomson Wadsworth, 2007.

[11] J. Han, M. Kamber and J. Pei, Data Mining Concepts and Techniques, Morgan Kaufmann Publishers, 2012.

[12] U. B. Kjærulff and A. L. Madsen, Bayesian Networks and Influence Diagrams: A Guide to Construction and Analysis, New York: Springer, 2008.

[13] D. T. Larose, Discovering Knowledge from Data: An Introduction to Data Mining, John Wiley & Sons, Inc., 2005.

[14] J. Haddena, A. Tiwaria, R. Roya and D. Rutab, "Computer assisted customer churn management: State-of-the-art and future trends," *Computers & Operations Research,* vol. 34, p. 2902 – 2917, 2005.

[15] L. Rokach and O. Maimon, Data Mining with Decision Trees: Theory and Applications, World Scientific Co. Ltd., 2008.

[16] I. H. Witten, E. Frank and M. A. Hall, Data Mining Practical Machine Learning Tools and Techniques, Morgan Kaufmann Publishers, 2011.

Clement Kirui received his Bachelor's degree (Information Technology) from Jomo Kenyatta University of Agriculture and Technology, Kenya, in 2008 and is currently pursuing his Master's degree (Computer Application Technology) in Central South University, China. He has previously worked with Acrobat Research Limited, Kenya (a market research firm) and currently works with the Municipal Council of Bomet, Kenya, at the IT Department. He is a Cisco Certified Network Associate (CCNA) and a Microsoft Certified Professional (MCP), with three Microsoft IT Professional certifications (MCITP Database Administrator 2008, MCITP Database Developer 2008, and MCITP Business Intelligence Developer 2008). He has co-authored one paper and five more are currently under review in different refereed journals. His research interests include Database Systems, Business Intelligence, Data Mining, Business Information Systems, and Internet Technologies.

Hong Li received his Bachelor's and Master's degrees from Peking University and PhD degree from Central South University in 2007. He is a professor and the director of the Department of Information and Communication Engineering, Central South University. His main research interests are data mining and DSP. He has published many papers in different refereed journals and conferences and has supervised students since 2001.

Wilson Cheruiyot received his Bachelor's degree (Mathematics and computer science) from Jomo Kenyatta University of Agriculture and Technology, Kenya, in 1994. He received his Master's degree and PhD (Computer Application Technology) both from Central South University, China, in 2002 and 2012 respectively. He is also a Microsoft Certified Professional (MCP) and a Microsoft Certified Database Administrator (MCDBA). He has previously worked with the Teachers Service Commission of Kenya and with the Kenya National Audit Office (KENAO). He is currently a professor of Computer Science at the Computing Department, Jomo Kenyatta University of Agriculture and Technology, Kenya. He has published over eleven papers in different refereed journals. His best paper award was a paper submitted to Springerlink journal of Multimedia systems, titled "Query quality refinement in singular value decomposition to improve genetic algorithms for multimedia data retrieval", whose impact factor was 1.176 and indexed by SCI and EI Compendex. Currently, he reviews articles for the Journal of Petroleum and Gas Engineering, www.academicjournals.org/JGE, and the Journal of Engineering and Technology Research, www.academicjournals.org/JETR/index.htm. His current research interests are: Multimedia Data Retrieval, Internet of Things, Evolutionary Computation for Optimization, Digital Image Processing, and ICT for Development.

Hillary Kirui received his Bachelor's degree (Electronics and Computer Engineering) from Jomo Kenyatta University of Agriculture and Technology, Kenya, in 2008 and is currently pursuing his Master's degree (ICT Policy & Regulation) at the same university. He has previously worked with Rapid Communications Limited, Kenya (a telecommunications firm) and currently works with the Kenya Airways Limited in the Flight Operations Department. He is a Cisco Certified Network Associate (CCNA) and is presently undertaking Cisco Certified Network Professional (CCNP) certification. His research interests include Data Communications, Network Security, ICT Policy Analysis, and ICT Regulatory Frameworks.