A Hybrid Data Mining Method for Customer Churn Prediction

Elham Jamalian
Faculty of Technology and Information
Department of Information Technology
University of Qom
Qom, Iran

Rahim Foukerdi
Faculty of Management
Department of Industrial Management
University of Qom
Qom, Iran

Abstract—**The expenses for attracting new customers are much higher compared to the ones needed to maintain old customers due to the increasing competition and business saturation. So customer retention is one of the leading factors in companies' marketing. Customer retention requires a churn management, and an effective management requires an exact and effective model for churn prediction. A variety of techniques and methodologies have been used for churn prediction, such as logistic regression, neural networks, genetic algorithm, decision tree etc.. In this article, a hybrid method is presented that predicts customers churn more accurately, using data fusion and feature extraction techniques. After data preparation and feature selection, two algorithms, LOLIMOT and C5.0, were trained with different size of features and performed on test data. Then the outputs of the individual classifiers were combined with weighted voting. The results of applying this method on real data of a telecommunication company proved the effectiveness of the method.**

Keywords-customer churn; data mining; hybrid method; LOLIMOT; C5.0; weighted voting

I. INTRODUCTION

The way companies communicate with their customers has become a key point for competition in marketing. Concepts such as customer acquisition, maintenance and satisfaction are internalized in companies. According to studies, the expenses for attracting a new customer are 5-10 times more than maintaining an old one [1]. On the other hand, customer maintenance has its own expenses and it's not possible to have these expenses for all customers, because not all customers worth maintaining. Thus, churn management systems are looking for customers that want to leave the company. The key point which shows the importance of churn management systems is that, studies show that a 5% increase in customer retention will have a 25% to 95% benefit increase for companies [2]. A lot of studies have been carried out in churn management, and different methods for recognizing downfall reasons and churn prediction and prevention have been used. Among these, data analysis methods have been used widely.

Customer churn is a situation in which customers decide to leave the company. The meaning of churn is different in different areas [3-5]. Most definitions keep the related

behaviors with a product and a defined threshold with business rules in mind. When customer's transaction is less than the threshold, churn occurs [6]. Authors in [7], defined customer churn, in banking field, as those customers who close their accounts. Authors in [6] define churn as those who have less than 2500 Euro (save, and all other kind of properties) in banks. Technically according to (1) churn is the customer loss in a defined time period.

$$\text{Monthly churn} = (C0 + A1 - C1)/C0 \qquad (1)$$

Here C0 means the number of customers at the beginning of the period, C1 is the number of customers at the end of the period and A1 is the number of new customers in the period.

Generally, customer churn prediction is a binary classification that its outcome shows the probability of customer churn [10, 11]. However, the special nature of the churn prediction problem causes analysis algorithms to face some limitations. For example, in these problems, data are imbalanced. This means that lost customers are just a small part of the data. In addition, extended learning programs will face some kinds of noise. In any case, churn prediction needs to classify customers according to the probability of churn [12]. Different approaches work on this problem and they have some defects themselves. For example, although algorithms based on decision trees are used for classification, it is possible that some leaves have the same class probability. In addition, this method is noise sensitive. Neural networks are looking for suboptimal solutions and when the number of parameters of models increases an overfitting happen. Although genetic algorithms present accurate prediction models, they can't make the probability of occurrence clear. And finally methods like support vector machines usually don't lead to the best results. In Table I, a summary of the above mentioned methods and their characteristics is shown.

In this study, using data fusion and feature extraction techniques, a hybrid method is presented for a more accurate prediction of customer churn. After data preparation and feature selection, two algorithms, LOLIMOT and C5.0, are trained with different size of features and the outputs of individual classifiers are combined with weighted voting.

TABLE I. CHURN PREDICTION MODELS REVIEW

Reference	Goal	Applied Techniques	Data, Customer numbers, Features numbers	Metrics	Sampling	Feature Selection
[13]	Explanation and application of modeling systems for automatic prediction of mobile service providers	Neural network	Telecommunication, 500000 customers, 200 features, (1)	Lift	Under sampling	Forward feature selection with decision tree and genetic algorithm
[14]	Making churn prediction model by C4.5 algorithm with limited numbers of features	C4.5 decision tree	Telecommunication, 114000 customers, 12 features, (1)	Lift, false positive rate, decision error	Oversampling	Visible
[15]	Using a new data mining method for churn prediction	C4.5 decision tree, neural network, datamining by evolutionary learning	Telecommunication, 100000 customers, 251 features, (1)	Top 5 lift,	Under sampling	Visible
[16]	Churn prediction as a part of customer life value model (CLV)	Logistic regression, decision tree, neural network	Telecommunication, 16384 customers, 200 features, (2)	Error rate, lift	-	R^2 method
[17]	Studying variables and modeling methods for customer churn prediction	Logistic regression, linear regression, random forest, regression forests	Financial services, 100000 customers, 30 features, (1)	AUC	-	-
[18]	The comparative and functional study of customer churn prediction	Logistic regression, decision tree, applying neural network on clustered parts	Telecommunication, 160000 customers, 40 features, (1)	Hit rate, Top decline lift	Oversampling	Visible and EDA
[10]	Improving and developing churn prediction models and testing them on real customer retention campaigns	Logistic regression(with markof chain), random forests	Pay-TV, 143198 customers, 81 features, (1)	Accuracy, lift, AUC	-	-
[19]	Using of support vector machine for churn prediction for newspaper subscribers	Logistic regression, support vector machine (SVM), random forests	Newspaper subscribers, 90000 customers, 82 features, (1)	Accuracy, lift, AUC	Under sampling	-
[20]	Improving accuracy of churn prediction by combining the cost sensitive random weighted forests and oversampling of random forests	Improved balanced random forests (IBRF)	Chinese bank, 20000 customers, 27 features, (1)	Accuracy, Top Decline Lift	Increasing	The best features selection increasingly
[21]	Customers churn behavior analysis for predicting churn and developing retention strategies	Decision tree, neural network, C4.5, Logit, PLS, Nave bays	Telecommunication, 100000 customers, 171 features, (2)	Hit rate	-	Feature selection with PLS and by use of VIP scores
[22]	Presenting a framework for churn prediction in electronic commerce area	Neural network, decision tree, SVM, Extended SVM	Chinese E-Commerce website, 50000 customers, 27 features, (1)	Accuracy, hit rate, cover rate, lift	-	-
[23]	Direct usage of behavior data in prediction model without converting them to static data by using hierarchical model	Hierarchical multi kernel SVM (H-MKSVM)	Telecommunication and food shop database, 633- 8842 customers, 47-79 features, (2)			

(1) private data, (2) public data

II. PROPOSED MODEL

Due to the fact that combining individual algorithms usually leads to better and more accurate predictions, in this study a hybrid method for customer churn prediction is suggested. Generally three factors should be kept in mind when combining algorithms: (1) Training set creation, (2) selection of individual algorithms that must be combined and (3) method or rules that results are combined. The suggested model for churn prediction includes the following steps:

A. Data Preparation

Information needed for data mining models includes demographic data like living location, age, gender, number of children, salary, other financial data and bills and data related to customer usage [14]. There is no need for customers' personal information such as name, email address, mailbox, father's name, etc. and they can be deleted from the database in this stage. Also missing data can be omitted or replaced with special amounts like average or predicted amount resulted from predicting techniques etc.

B. Features Selection or Extraction and Ordering

This step is important because it uses important features and omits extra and noisy features and those which have little information helping to data cleansing and dimension reduction [8]. In this study principal component analysis (PCA) is used for this purpose. In this step, the goal is achieve to a set of features which are the most effective in prediction by the use of PCA. F_k is the kth ordered feature that:

$$F_1 \square \ F_2 \square \ ... \square \ F_n \tag{2}$$

C. Constructing Subsets of Data According to The Extracted Features and Training Model(s):

Regarding to the literature and strengthened C5.0 and LOLIMOT high performance [26], these models are used as individuals and by using weighted voting their accuracy is improved.

1) C5.0 Algorithm

C5.0 decision tree is a classification tree. C5.0 algorithm is an improved form of ID3 algorithm that constructs the decision tree based on information theory. Training data are classified sets of samples in the form of S= s1, s2,…. Each sample is a vector of features s1= X1, X2,… Training data include the vector that shows the class that each set belongs to, C= C1, C2,… In each tree node the feature which makes the best classification is selected in order to put the sample set in related classes. This classification is done by computing entropy (information gain). In this case the feature that has the highest information gain will be selected as decision maker.

2) Boosting Algorithm

Boosting algorithm is an ensemble method that trains a set of classification models [27]. It is the case when each train set is created based on the accuracy of previous classification models. In this algorithm, new classification methods are constructed to predict the samples, in which classification accuracy was weak, more accurately. This process is done by adaptive resampling. This means that samples that were classified incorrectly will have more chance to be selected in the next step. Each sample has a weight and at the end of the classification these weights will be up to date. Finally, the result of different methods will be ensembled with voting. Strengthening can increase the accuracy of C5.0 algorithm but it needs more time for training.

3) LOLIMOT Algorithm

Linear tree model or LOLIMOT is based on the divide and conquer strategy. In this method the complicated problem is solved by dividing to several smaller problems [9]. LOLIMOT algorithm for achieving a better outcome (outcome with smaller error) will divide the problem area to several local linear models (LLMs). And after finding the worst LLM, continues the algorithm by dividing it to two LLMs. This algorithm is used as a fast learning tree algorithm, in many pattern recognition and prediction problems and it has some remarkable results. LOLIMOT algorithm was first introduced in locally linear neuro fuzzy model [28]. Dividing the incoming area to small linear subsets by using fuzzy activation functions

is the basic strategy in these networks. In locally linear neuro fuzzy model, the neural network has one hidden layer. But computing operation in its neurons is much more complex than that of a regular neuron. In this network, there are (p+1)*M weights (M is the number of neurons and p is the number of inputs). In (3) and (4) the matrix of weights and inputs can be seen. Each neuron has a LLM and a validity function. Validity function will indicate validity area in the LLM. Validity function is also known as activation function, because it controls the actions of Locally Linear Models.

$$\underline{\omega} = [\omega_0, \omega_1, ..., \omega_p] \tag{3}$$

$$\underline{u} = [u_1, u_2, ..., u_p]^T \tag{4}$$

It is clear that the first output is achieved by (5). It is the output of LLM of the ith neuron.

$$\hat{y}_i = \omega_{i0} + \omega_{i1}u_{i1} + ... + \omega_{ip}u_{ip} \tag{5}$$

Activation function can be computed by using (6) and (7).

$$\phi_i(\underline{u}) = \frac{\mu_i(\underline{u})}{\sum_{i=1}^{M} \mu_i(\underline{u})} \tag{6}$$

$$\mu_i(\underline{u}) = \exp\left[\frac{(u_1 - c_{i1})^2}{-2\sigma_{i1}^2}\right] \times ... \times \exp\left[\frac{(u_p - c_{ip})^2}{-2\sigma_{ip}^2}\right] \tag{7}$$

where $\mu_i(\underline{u})$ is a Gaussian function which has two parameters, center C_{ij} and standard deviation σ_{ij}.

$$\hat{y} = \sum_{i=1}^{M} \hat{y}_i \varphi_i(\underline{u}) \tag{8}$$

$$\underline{X} = \left[\underline{X}_1^{(sub)}, \underline{X}_2^{(sub)}, ..., \underline{X}_M^{sub} \right] \tag{9}$$

$$\underline{X}_i^{(sub)} = \begin{bmatrix} 1 & u_1(1) & \cdots & u_p(1) \\ \cdots & \cdots & \cdots & \cdots \\ 1 & u_1(N) & \cdots & u_p(N) \end{bmatrix} \tag{10}$$

In LOLIMOT, two parameters, center and standard deviation, are considered to be fixed and the weights are calculated by the least squares function. With the assumption that having M neurons and N train vector, regression matrix and train vector output can be defined as (7), (8) and (9). The output vector for train data, is considered as:

$$\hat{y}_i = [\hat{y}_i(1), \hat{y}_i(2), ..., \hat{y}_i(N)] \tag{11}$$

And, finally, the network output is calculated by (12).

$$\hat{y} = \underline{X}\omega \qquad (12)$$

Now, by considering a target function (13) through changing the weights, we try to minimize the error.

$$I = \sum_{j=1}^{N} e^2(j) \rightarrow \min \qquad (13)$$

That, $e(j) = y(j) - \hat{y}(j)$

LOLIMOT algorithm starts the training process by starting from one neuron or a specified numbers of neurons and the following three stages are done repeatedly:

- The worst neuron with highest error is recognized by (14).

$$I_i = \sum_{i=1}^{N} e^2(j)\varphi_i(\underline{u}(j)) \qquad (14)$$

- All cases in which this neuron can be divided in one of its dimensions are considered and the case that has the least error is selected.

- The worst neuron is recomputed.

After selecting suitable data mining algorithms, subsets of training data are constructed based on feature sets and the proposed models train and test on each of these train sets. Models are evaluated and cleaned according to their accuracy.

4) Combining outputs of classifiers

In this phase, the outputs of the remained classifiers from the previous step will ensemble with one of the voting algorithms that will be explained below. The simplest method to combine classifiers is majority voting which gathers all outputs of individual classifiers C_1, C_2, …, C_n and then the output with the largest number of votes is selected as the final decision.

$$CT = \arg\max \text{Count}_i(V(C_i)) \qquad (15)$$

where V(Ci) is the vote of ith classifier.

The other voting method is the weighted voting, that is improved but more complicated version of majority voting. In this strategy, each classifier is assigned a weight according to its classification performance. In this case, the weights can be obtained by the following formula [29].

$$W_i = \frac{\alpha}{n} + (1-\alpha).(\frac{1}{2})^i \qquad (16)$$

where Wi is the weight of i_{th} ordered classifier in which the classifiers are arranged based on their accuracy performance, α is a regulatory parameter $0<\alpha<1$ and $\sum_1^N W_i \approx 1$.

Then, the voting output will become:

$$CT = \arg-\max \sum_{k \in V(C_i)} \qquad (17)$$

Based on this method different weights are given to outputs according to their accuracy. If the total weights of the methods that estimate customer is churner are more than the whole weights of other methods, the customer will be predicted as churner, if not the customer is loyal.

III. EVALUATION CRITERIA

The lift criterion is a performance measure which is the result of the ratio between the obtained outcomes with and without the use of the prediction model. The higher the lift means the model is more accurate, and instinctively, the more profitable a targeted proactive churn management program will be [5]. Calculating lift criteria for all of churn management data base is not logical and also not practical. It is also not economically possible for the organization to apply churn management for all customers, so it concentrates on those with more churn probability. For this, the two mentioned outcomes are usually kept apart, and are about 10-20% in marketing [26, 30].

As the top decile lift is the main criterion in the churn management area [5], this criterion has been used to evaluate the performance of applied algorithms. The top decile lift focuses on the customers predicted most likely to churn. The first 10% are the most critical customers (i.e. with high churn probability) and this percentage is an ideal portion for targeting the retention marketing campaign [9]. So for computing top decline lift, first customers are sorted from predicted most likely to predicted least likely to churn. Then the number of customers that were correctly recognized in the first 10% will be calculated. And finally this percent will be divided with the total churn. If all customers (in the first 10%) were recognized correctly and the ratio between churners and other customers is 50-50, the amount of top decline lift is equal to 100/50. In addition to lift, proposed algorithm was compared with some other algorithms that were mentioned in literature by the use of two other criterions, accuracy and area under ROC curve (AUC).

IV. DATASET

The dataset is provided by the Teradata Center at Duke University. The database contains of datasets of mature subscribers (i.e. customers who were with the company for at least six months) from a major U.S. telecommunication company. There are three different datasets in this database: calibration data, current score data and future score data. A total of 172 variables are included in the datasets, one for churn indication, and 171 variables for prediction. The prediction variables include three types of variables: behavioral data such as minutes of use, revenue, handset equipment, company interaction data such as customer calls into the customer service center, and customer household demographics. The churn response is coded as a dummy variable with churn=1 if the customer churns, and churn=0 otherwise. There are 100,000 records in the calibration dataset, 51,306 records in the current score and 100,462 records in the future score dataset. The actual average monthly churn rate is reported to be around

1.8%. Calibration sample is a balanced data set with 50-50 split between churners and non-churners. Also churn rate in the current score and future score datasets is about 1.8%. The calibration dataset (a balanced one) is used as the training set and future score dataset is used as the test set in our experiments. It is necessary to mention that the customers in each dataset are different. Figure 1 shows the training set in two dimensions. In this Figure, churner customers and non-churners have a noticeable overlapping which makes their recognition difficult. In Figure 1, red circles are churner customers and blue squares are non-churners.
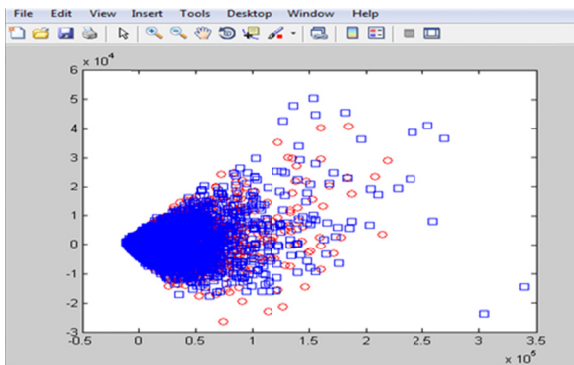

Fig. 1.   Training set in two dimensions

V.   METHODOLOGY

A. Data Preparation

At first all features with more than 30% missing value have been omitted. Then for non-numerical data a new value was given and for numerical data linear interpolation was used. Linear interpolation makes the replacement of missing value possible by using of linear relations between known and missing value. This procedure has been carried out in SPSS 18.0 platform.

B. Feature Selection

There are 171 features in the used dataset, and after omitting features with more than 30% missing value, 144 features remain of which 120 features were numerical and 24 were non-numerical. For doing other stages the non-numerical features were coded.

C. PCA Implementation

The database used in this study just like most other CRM databases is big, and with lots of dimensions that have high correlations as it can be seen in Figure 2. The correlation between features reduces the accuracy of models, while on the other hand the large amount of data makes the data analysis difficult. So, for reducing the dimensions and choosing the best and the most informative features, PCA was used. After PCA application, the number of principle components is defined. These components will be used as features at the next phase. For defining their number, variety ratio criterion was used, which is one of the most important ones in this field. This ratio is usually between 70-90% [31].
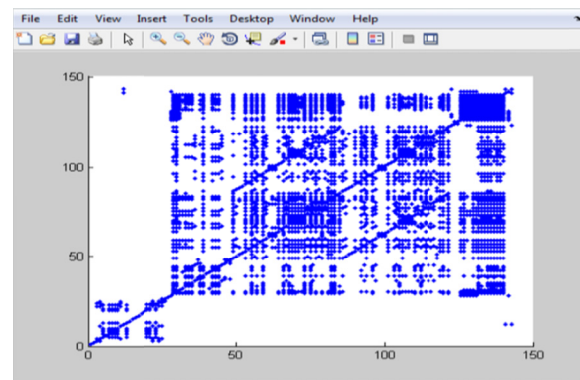

Fig. 2.   Correlation matrix with values over 0.5

In this phase, 55 features were selected that include 90% of variety. The result of this phase is 55 features ordered according with descending importance.

D. Feature Subsets Development

Feature subsets are constructed and then training subsets are developed based on them.

$$A_1 = \{F_1, F_2, F_3\} \; , \; A_2 = \{F_1, F_2, F_3, F_4\} \; ,$$

$$A_3 = \{F_1, F_2, F_3, F_4, F_5\} \; , \; A_{50} = \{F_1, F_2, F_3, F_4, F_5, ..., F_{55}\}$$

E. Model Training

LOLIMOT algorithm was implemented in MATLAB (R2011a) platform and C5.0 was performed in Clementine software. At first, LOLIMOT and C5.0-boosting were applied on the training subsets separately and then they were tested. After the testing the classifiers were ordered according to their accuracy and those with lower than 50% were omitted. Table II shows the classifiers with accuracy higher than 50%. Finally 18 classifiers remained.

TABLE II.   CLASSIFIERS WITH HIGHER THAN 50% ACCURACY

C5.0-boosting-10 features	LOLIMOT-15 features
C5.0-boosting-12 features	LOLIMOT-20 features
C5.0-boosting-13 features	LOLIMOT-25 features
C5.0-boosting-24 features	LOLIMOT-30 features
C5.0-boosting-37 features	LOLIMOT-32 features
C5.0-boosting-50 features	LOLIMOT-45 features
C5.0-boosting-55 features	LOLIMOT-48 features
LOLIMOT-5 features	LOLIMOT-50 features
LOLIMOT-10 features	LOLIMOT-55 features

F. Individual Classifiers Combination

For achieving the final result the outcomes of these classifiers must be combined with one of the voting methods. The results of weighted voting with different regulatory parameter ($0.1 < \alpha < 1$) were compared with majority voting (Figure 3). Also in Figure 4, the sensibility of weighted voting to the different values of α is shown. In Tables III-IV a comparison between the different methods mentioned in literature of churn prediction and the proposed model, according to top decile lift and accuracy criterion, is shown. Figure 5 represents the results of the methods mentioned in Table III. The proposed model is named as HWV in Figure 5.

In Table V different models are compared according to the AUC criterion. It can be that seen the suggested model has the highest AUC. Also, the suggested method has better top decline lift in comparison with other hybrid methods. Figure 6 shows the comparison between the different models according to AUC.
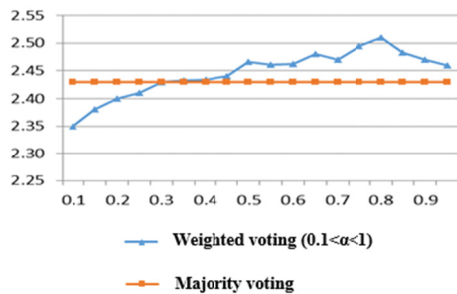

Fig. 3.   Comparison between majority and weighted voting
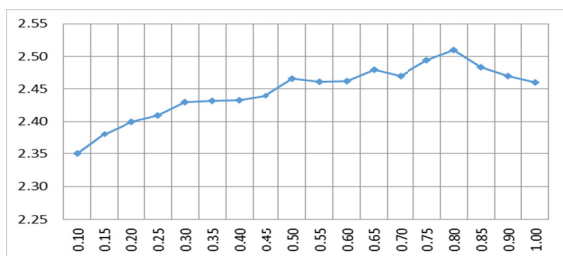

Fig. 4.   Sensibility of weighted voting to regulatory parameter (α)
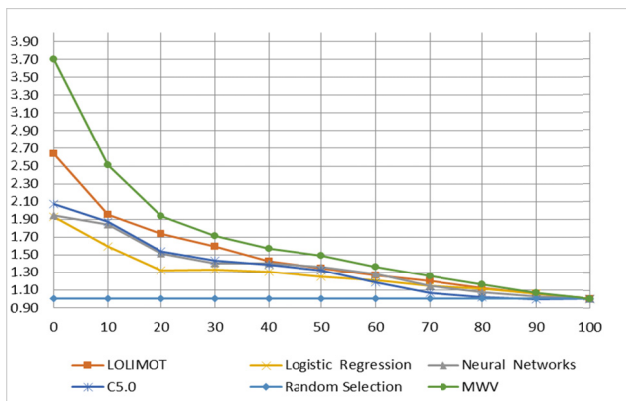

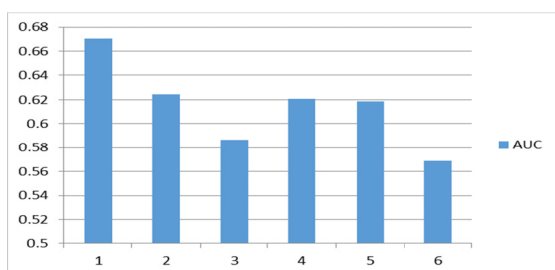Fig. 5.   Comparison between different models according to top decline lift

Fig. 6.   Comparison between different models according to AUC

TABLE III.   MODEL COMPARISON ACCORDING TO TOP DECLINE LIFT (FIGURE 5)

Algorithm	Top Decline Lift
Suggested model	2.51
LOLIMOT	1.94
C5.0	1.87
Neural network	1.84
Logistic regression	1.43
Beysian network	1.35

TABLE IV.   MODEL ACCURACY COMPARISON

Algorithm	Accuracy
Proposed model	92%
LOLIMOT	70%
Logistic regression	60%
C5.0	50%
Neural network	45%
Beysian network	44%

TABLE V.   MODEL COMPARISON ACCORDING TO AUC (FIGURE 6)

Algorithm	AUC
Proposed model (1)	0.67
LOLIMOT(2)	0.62
Logistic regression (3)	0.58
C5.0 (4)	0.62
Neural network (5)	0.62
Beysian network (6)	0.57

VI.   CONCLUSION

As churn data are usually noisy and imbalanced, non-hybrid models are often not that accurate. So, in this study a hybrid model was presented to increase the accuracy of classifiers. By using the presented hybrid method, after gathering suitable data and applying necessary preprocessing phases, the most important features were selected and were given to the predictors and the results were combined to achieve the final outcome. This method was evaluated by using data from a telecommunication company and its higher effectiveness, in comparison with available methods, was proved. Two algorithms, LOLIMOT and C5.0 were used in this method, but a larger number of algorithms can be combined, and other ensemble methods that lead to more accurate prediction can be examined (See fuzzy regression [32], ANFIS [33-34], GA-ANN [35] and ICA [36]) to produce a more efficient scheme. The number of features, feature subsets and training subsets can affect the prediction accuracy. Improving the methods for selecting the optimal number of these subsets can also improve the model.

REFERENCES

[1] E. Ko, S. H. Kim, M. Kim, J. Y. Woo, "Organizational characteristics and the CRM adoption process", Journal of Business Research, Vol. 61, No. 1, pp. 65–74, 2008

[2] S. Gupta, D. Hanssens, B. Hardie, W. Kahn, V. Kumar, N. Lin, N. Ravishanker, S. Sriram, "Modeling customer lifetime value", Journal of Service Research, Vol. 9, No. 2, pp. 139–155, 2006

[3] G. Nie, W. Rowe, L. Zhang, Y. Tian, Y. Shi, "Credit card churn forecasting by logistic regression and decision tree", Expert Systems with Applications, Vol. 38, No. 12, pp. 15273–15285, 2011

[4] J. Lu, "Predicting customer churn in the telecommunications industry—An application of survival analysis modeling using SAS", SAS User Group International (SUGI27) Online Proceedings, pp. 114–27, 2002

[5] S. A. Neslin, S. Gupta, W. Kamakura, J. Lu, C. Mason, Defection detection: Improving predictive accuracy of customer churn models, Tuck School of Business, Dartmouth College, 2004

[6]   N. Glady, B. Baesens, C. Croux, "Modeling churn using customer lifetime value", European Journal of Operational Research, Vol. 197, No. 1, pp. 402–411, 2009

[7]   D. Van den Poel, B. Larivière, "Customer attrition analysis for financial services using proportional hazard models", European Journal of Operational Research, Vol. 157, No. 1, pp. 196–217, 2004

[8]   J. Hadden, A. Tiwari, R. Roy, D. Ruta, "Computer assisted customer churn management: State-of-the-art and future trends", Computers & Operations Research, Vol. 34, No. 10, pp. 2902–2917, 2007

[9]   A. Ghorbani, F. Taghiyareh, C. Lucas, "The application of the locally linear model tree on customer churn prediction", International Conference of Soft Computing and Pattern Recognition, SOCPAR'09, Malacca, Malaysia, pp. 472–477, December 4-7, 2009

[10] J. Burez, D. Van den Poel, "CRM at a pay-TV company: Using analytical models to reduce customer attrition by targeted marketing for subscription services", Expert Systems with Applications, Vol. 32, No. 2, pp. 277–288, 2007

[11] J. Han, M. Kamber, J. Pei, Data Mining: Concepts and Techniques. Morgan Kaufmann, 2006

[12] P. C. Pendharkar, "Genetic algorithm based neural network approaches for predicting churn in cellular wireless network services", Expert Systems with Applications, Vol. 36, No. 3, pp. 6714–6720, 2009

[13] P. Datta, B. Masand, D. R. Mani, B. Li, "Automated cellular modeling and prediction on a large scale", Artificial Intelligence Review, Vol. 14, No. 6, pp. 485–502, 2000

[14] C.-P. Wei, I. Chiu, "Turning telecommunications call details to churn prediction: a data mining approach", Expert Systems with Applications, Vol. 23, No. 2, pp. 103–112, 2002

[15] W.-H. Au, K. C. Chan, X. Yao, "A novel evolutionary data mining algorithm with applications to churn prediction", IEEE Transactions on Evolutionary Computation, Vol. 7, No. 6, pp. 532–545, 2003

[16] H. Hwang, T. Jung, E. Suh, "An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry", Expert Systems with Applications, Vol. 26, No. 2, pp. 181–188, 2004

[17] B. Larivière, D. Van den Poel, "Predicting customer retention and profitability by using random forests and regression forests techniques", Expert Systems with Applications, Vol. 29, No. 2, pp. 472–484, 2005

[18] S.-Y. Hung, D. C. Yen, H.-Y. Wang, "Applying data mining to telecom churn management", Expert Systems with Applications, Vol. 31, Vo. 3, pp. 515–524, 2006

[19] D. Anil Kumar, V. Ravi, "Predicting credit card customer churn in banks using data mining", International Journal of Data Analysis Techniques and Strategies, Vol. 1, No. 1, pp. 4–28, 2008

[20] Y. Xie, X. Li, E. W. T. Ngai, W. Ying, "Customer churn prediction using improved balanced random forests", Expert Systems with Applications, Vol. 36, No. 3, pp. 5445–5449, 2009

[21] H. Cho, Y. Lee, H. Lee, "Toward Optimal Churn Management: A Partial Least Square (PLS) Model", Proceedings of the Sixteenth Americas Conference on Information Systems, Lima, Peru, AMCIS, August 12-15, 2010

[22] X. Yu, S. Guo, J. Guo, X. Huang, "An extended support vector machine forecasting framework for customer churn in e-commerce", Expert Systems with Applications, Vol. 38, No. 3, pp. 1425–1430, 2011

[23] Z.-Y. Chen, Z.-P. Fan, M. Sun, "A hierarchical multiple kernel support vector machine for customer churn prediction using longitudinal behavioral data", European Journal of Operational Research, Vol. 223, No. 2, pp. 461-472, 2012

[24] K. Coussement, K. W. De Bock, "Customer churn prediction in the online gambling industry: The beneficial effect of ensemble learning", Journal of Business Research, Vol. 66, No. 9, pp. 1629-1636, 2013

[25] K. W. De Bock, K. Coussement, and D. Van den Poel, "Ensemble classification based on generalized additive models", Computational Statistics & Data Analysis, Vol. 54, No. 6, pp. 1535–1546, 2010

[26] J. Basiri, F. Taghiyareh, B. Moshiri, "A Hybrid Approach to Predict Churn", 2010 IEEE Asia-Pacific Services Computing Conference (APSCC), pp. 485–491, Hangzhou, China, December 6-10, 2010

[27] Y. Freund, R. E. Schapire, "Experiments with a new boosting algorithm", Proceedings of the Thirteenth International Conference on International Conference on Machine Learning, Bari, Italy, July 3-6, 1996

[28] M. Mirmomeni, C. Lucas, B. N. Araabi, M. Shafiee, "Forecasting sunspot numbers with the aid of fuzzy descriptor models", Space Weather, Vol. 5, No. 8, 2007

[29] M. Siami, M. R. Gholamian, J. Basiri, "An application of locally linear model tree algorithm with combination of feature selection in credit scoring", International Journal of Systems Science, Vol. 45, No. 10, pp. 2213-2222, 2014

[30] A. Lemmens, C. Croux, Bagging and boosting classification trees to predict churn, Journal of Marketing Research, Vol. 43, No. 2, pp. 276–286, 2006

[31] I. Jolliffe, Principal Component Analysis, Wiley Online Library, 2005

[32] S. Nazari-Shirkouhi, A. Keramati, "Modeling customer satisfaction with new product design using a flexible fuzzy regression-data envelopment analysis algorithm", Applied Mathematical Modelling, Vol. 50, pp. 755-71, 2017

[33] S. Nazari-Shirkouhi, A. Keramati, K. Rezaie, "Improvement of customers' satisfaction with new product design using an adaptive neuro-fuzzy inference systems approach", Neural Computing and Applications, Vol. 23, No. 1, pp. 333-43, 2013

[34] V. Majazi Dalfard, M. Nazari Asli, S. Nazari-Shirkouhi, S. M. Sajadi, S. M. Asadzadeh, "Incorporating the effects of hike in energy prices intoenergy consumption forecasting: A fuzzy expert system", Neural Computing and Applications, Vol. 23, No. 1, pp. 153-69, 2013

[35] A. Azadeh, S. M. Asadzadeh, R. Jafari-Marandi, S. Nazari-Shirkouhi, G. Baharian Khoshkhou, S. Talebi, A. Naghavi, "Optimum estimation of missing values in randomized complete block design by genetic algorithm", Knowledge-Based Systems, Vol. 37, pp. 37-47, 2013

[36] S. Nazari-Shirkouhi, H. Eivazy, R. Ghodsi, K. Rezaie, E. Atashpaz-Gargari, "Solving the integrated product mix-outsourcing problem using the imperialist competitive algorithm", Expert Systems with Applications, Vol. 37, No. 12, pp. 7615-7626, 2010