International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2015): 78.96 | Impact Factor (2015): 6.391

Customers Churn Prediction and Attribute Selection in Telecom Industry Using Kernelized Extreme Learning Machine and Bat Algorithms

S. Induja[1], Dr. V. P. Eswaramurthy[2]

[1]Research Scholar, Periyar University, Salem – 636011, Tamilnadu, India
Assistant Professor, Department of Computer Science, Government Arts and Science College, Komarapalayam, Tamilnadu, India

[2]Assistant Professor, Department of Computer Science, Government Arts and Science College, Komarapalayam, Tamilnadu, India

Abstract: ***With the fast development of digital systems and concomitant information technologies, there is certainly an incipient spirit in the extensive overall economy to put together digital Customer Relationship Management (CRM) systems. This slanting is further more palpable in the telecommunications industry, in which businesses turn out to be increasingly digitalized. Customer churn prediction is a foremost aspect of a contemporary telecom CRM system. Churn prediction model leads the customer relationship management to retain the customers who will be possible to give up. Currently scenario, a lot of outfit and monitored classifiers and data mining techniques are employed to model the churn prediction in telecom. Within this paper, Kernelized Extreme Learning Machine (KELM) algorithm is proposed to categorize customer churn patterns in telecom industry. The primary strategy of proposed work is organized the data from telecommunication mobile customer's dataset. The data preparation is conducted by using pre-processing with Expectation Maximization (EM) clustering algorithm. After that, customer churn behavior is examined by using Naive Bayes Classifier (NBC) in accordance with the four conditions like customer dissatisfaction (H_1), switching costs (H_2), service usage (H_3) and customer status (H_4). The attributes originate from call details and customer profiles which is enhanced the precision of customer churn prediction in the telecom industry. The attributes are measured using BAT algorithm and KELM algorithm used for churn prediction. The experimental results prove that proposed model is better than AdaBoost and Hybrid Support Vector Machine (HSVM) models in terms of the performance of ROC, sensitivity, specificity, accuracy and processing time.***

Keywords: Churn prediction, Expectation Maximization, Kernelized Extreme Learning Machine, data preparation, pre-processing, attribute selection, BAT algorithm, Naive Bayes Classifier

1. Introduction

Customer churn is the most important measure of lost customers. Telecommunication companies often lose potential customers and, thus, revenues to the competition. The telecommunication industry has gone through drastic changes for the last few decades such as addition of new services, technical improvements and increased competition due to deregulation [1]. Customer churn prediction in telecommunication has, thus, become important to industry workers in order to protect their loyal customer base, organization growth, and improve its Customer Relationship Management (CRM) [2, 3]. Retaining customers with high churn risk is one of the toughest challenges in telecommunication industry today [4]. Due to greater number of service providers as well as more intense competition, customers today have a variety of options to churn. Thus, the telecommunication industry workers are waking up to the importance of retaining existing customers as opposed to acquiring new ones [2].

Telecommunication companies are routinely deals with large quantity of superior data, basically rich customer base and growing rapidly with highly competitive environment [5]. In this fiercely competitive market, the subject of customer retention, loyalty, and churn becomes the notable attention in many industries. The most important and difficult problem meet out by Mobile industry is customer churn. Customer churn refers to the periodic loss of potential customers in an organization. Customers can choose various multiple service providers and actively exercise their rights of switching from one service provider to another. When a customer quits, we lose not only the future revenue from this customer but also the resources we spent to acquire the potential customer like new recruits of manpower, cost of publicity and offer discounts. Attracting thousands of new subscribers is worthless if an equal number are leaving. Attracting new customers costs almost five to six times more than retaining the old customers. [6] The company likely know exactly why a customer chooses to aim for other company is important in the product-range or even services by detecting problems. Appealing new customers charges roughly five to six times greater than maintaining the old customers. The prerequisite of retaining customers desires for correct customer churn prediction models. A model needs to be designed to recognize the reasons, that why you should churn along with the enhancements necessary to retain customers.

With this paper, Kernelized Extreme Learning Machine (KELM) algorithm is proposed to categorize customer churn habits in telecom industry. Preliminary phase of data preparation, will be the collection of data, integrated and it should be cleaned. The data preprocessing is performed in this particular phase using EM algorithm. And then the customer behaviour is analyzed and classified as four categories using naive bayes classifier. And the attributes are selected using BAT algorithm to |pursue relevant attributes, eradicate irrelevant or redundant ones. Later on, the reduced data together with only relevant attributes are passed on into classifiers trained using KELM algorithm.

The remaining part of this paper is designed as follows: section 2 will focus on related work of churn prediction. Section 3 uses the proposed methodology of customer churn prediction. Section 4 will depict the experimental results and discussions. Section 5 gives the conclusions and directions for further research.

2. Related Work

Oseman et al. [7] presented how to put into application grouping decision tree methods for churn examination in telecommunication industry. An illustration set is used to carry out a test of customer churn issue using the ID3 decision tree. In their outcomes they establish that the area of subscriber was main classification characteristics that contributed to client churn, other than two minor reasons for customer to churn.

In Taiwan, Wei & Chiu [8] put into use C4.5 based procedures on one of the largest local mobile telecommunication companies & it recognized 28.32% of the subscribers that restricted some of true churners with the lift factor of 2.30 & the preservation time of 14 days. This can be associated to research by Jahormi et al. [9] that the aimed at evolving a predictive model for client churn in pre-paid mobile telephony establishments. They applied decision trees methods like C5.0 with the neural network & it was exposed that based on improvement measure decision trees executed better than neural networks. An associated study was approved out by Yeshwanth [10] wherein he combined J48 decision tree along Genetic algorithm & designed a hybrid evolutionary methodology for churn prediction in mobile networks. Author attained 72% exact outcomes for largest telecom company in growing nations.

Kaur et al. [11] useful Naive Bayes, J48 & support vector machineries classifiers to process data so as to classify the important characteristics of the customers that help in forecasting churn of the bank clients. In their findings, they decided that achievement forecast of the loyal class is less than prediction success rate % of the churn class. Additionally they also originate that the J48 decision tree had enhanced performance related to other methods. Soein & Rodpysh [12] preformed some tests in Iran involving relating numerous well-known data mining methods: C5.0, QUEST, and CART, CHAID, Bayesian networks & Neural networks to find out the best technique of customers‘ the churn prediction in the Iranian Insurance Company. The outcomes presented that CART decision tree had improved performance than other methods.

Hadden et al. [13] had the purpose of specifying the most appropriate model for churn prediction analysis. They showed an estimate on the different algorithms such as CART trees, neural networks& regression &confirmed their correctness in predicting customer churn. They originate that decision trees outperform rest of other methods with an overall correctness percentage of 82%. Au et al. [14] believe that the largest limitation of neural networks is that they hardly uncover patterns in an easily understandable manner. Their study also had shown that neural networks outdo decision trees for prediction of churn through identification of more churners compared to C4.5 decision trees.

Qureshi et al. [15] in their examination imagine active churners in the Telecom industry by applying numerous methods of data mining such as, K-Means Clustering, Logistic Regression, Neural Network, Linear &,Exhaustive CHAID, CART, QUEST, &CHAID. They found that Exhaustive CHAID performed well related to all other methods. 60% was the percentage of correctly recognized churners which was highest % among altogether other methods. However, other decision trees variants did not demonstrate as high presentation in addition to Exhaustive CHAID. Jahromi et al. [16] showed research with aim of emerging predictive model for the customer churn in pre-paid companies of mobile telephony. They accepted tests on performance of numerous model-building algorithms such as Neural Networks, C5.0, CART, & CHAID.

Lazarov & Capota [17] discussed commonly used data mining algorithm in customer churn analysis and prediction. Regression tree techniques were discussed along with other popular data mining methods like Decision Trees, Rule based learning and Neural Networks. The conclusion was that good prediction models have to be constantly developed and a combination of the proposed techniques has to be used.

Zhang et al. [18] who presented in their research a hybrid method of the k-nearest neighbor algorithm as well as the logistic regression method for building a binary classifier named KNN-LR. They performed a comparison between KNN-LR with logistic regression, C4.5 and radial basis function (RBF) network. The outcome was that KNN-LR outperformed RBF on almost all the four benchmark datasets. In addition, this also outperformed logistic regression on these types of benchmark data sets, just that they have identical performance on the Wisconsin breast cancer data set. The result also indicated its superiority over RBF and C4.5 but C4.5 just exceeded KNN-LR on telecom dataset.

3. Proposed Methodology

In this section, the data preparation is discussed and the customer behaviour also analysed. Then the attribute selection methods and churn prediction are explained. The overall performance of proposed system is explained in system overview.

3.1 System Overview

The functionality the suggested system is illustrated in fig1. It implies that the preliminary stage of data preparation, it will be collected the data, integrated and thoroughly cleaned. The data preprocessing is performed in this stage by applying EM algorithm. And then the customer behaviour examined and categorized as four different types using naive bayes classifier. Subsequently, the customer churn behaviour is assessed in keeping with the (H_1,H_2,H_3,H_4) factors. These conditions of customer behaviors are associated with the probabilities of customer churn. The probability estimated with the guidance of the naive bayes classifier. Along with the attributes are chosen by means of BAT algorithm to check for} related attributes, reduce unnecessary or redundant ones. Later on, the reduced data having only related attributes will be passed on into classifiers skilled using KELM algorithm.

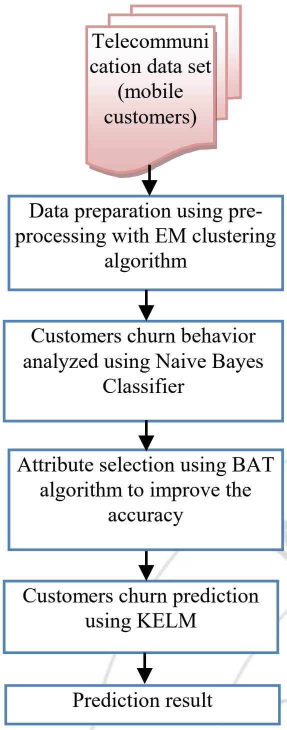

Figure 1: Overall Architecture Proposed System

3.2 Data preparation using EM clustering algorithm

On this data preparation work, information is obtained, incorporated and also cleaned. Integration of data really need extraction of data from various resources. Now that the data have been organized in tabular method, it will need to be completely recognized.

Data has to be cleaned by solving any specific ambiguities, flaws. The data cleaning is performed by preprocessing stage. In preprocessing phase, data cleaning is performed and also eliminates the inappropriate information that also includes incorrect spelling speech because of human errors, unique mathematical figures, lost values, repeated information, and so on. Additionally redundant and troublesome data items should be taken down at this stage. Not every areas of the database will always be ideal for modeling objectives. Areas with distinctive values, such as contact information or even individual unlock codes are does not need to be used. Most of these should not have predictive value since they distinctively recognize each row. Moreover areas having only a single value are left out, because these characterize a marginal section of the data. Then finally, areas with a lot of “null” values are too neglected. The data arranged, will make use of EM clustering algorithm. It will be utilized in optimum likelihood approximation in which the issue includes two data sets of arbitrary variables which often one, X, is noticeable ,as well as other, Z, is hidden. In simple phrase algorithm performs in following two steps:

E-step: Estimates the expectation of the missing value i.e. unlabeled class information. This step corresponds to performing classification of each unlabeled document. Probability distribution is calculated using current parameter.

M-step: Maximizes the likelihood of the model parameter using the previously computed expectation of the missing values as if were the true ones.

Algorithm 1: Pre-processing of telecommunication mobile customer data
Inputs: Collections of mobile customer telecommunication data set, no of attributes (words)
Output: predict the cleaned data
maximum probability calculation method:
Initial naive Bayes classifier, $\hat{\Theta}$, from the input datasets, use maximum a posteriori parameter estimation to find $\hat{\Theta}$ argmax $P(D/\Theta)P(\Theta)$

- Loop while classifier parameters improve, as measured by the change in $L_c(\Theta/D, z)\, c$
- **E-step)** Use the current classifier, $\hat{\Theta}$, to estimate component membership of each unlabeled dataset, *i.e.*, the probability that each mixture component (and class) generated each dataset, $P(c_j / d_i ; \hat{\Theta})$.
- **(M-step)** Re-estimate the classifier, $\hat{\Theta}$, given the estimated component membership of each dataset. Use maximum a posteriori parameter estimation to find $\hat{\Theta}=\text{argmax}_{\Theta} P(D/\Theta)P(\Theta)$.

Preprocessing steps:
1. Remove periods, commas, punctuation, stop words. Collect data's that have occurrence frequency more than once in the data set.
2. View the frequent words as data sets by matching the words which are in the vocabulary as well as training datasets.
3. Search for matching word set(s) or its subset(containing items more than one) in the list of data sets collected from training data with that of subset(s) (containing items more than one) of frequent data set of new dataset.
4. Collect the corresponding probability values of matched data set(s) for each target class.
5. Calculate the probability
6. Apply z score algorithm to calculate the range in which the attributes must be lying.
7. Calculate the probability class by applying expectation maximization algorithm.
8. Categorize the dataset in the class having maximum probability as cleaned dataset.

4. Customer Behaviour Analysis Using NBC

The variable specification of customer behavior is specified in table 1. Right here, assumed to be a collection of customers, each one of whom is associated with usual communications with a market depending on customer's position from effective use with the services on a normal basis) to non-use (determining never to make use of temporarily without requiring churned until recently) or even floating (getting closed by the service provider) since influenced defection and also from effective use to churn.

The customer churn determinants is analyzed in accord with the four circumstances such as customer dissatisfaction (H_1), switching costs (H_2), service usage (H_3) and customer status (H_4). The customer churn behavior is analyzed in accord with the customer churn determinants.

With the use of customer transaction and billing data, this type analyzes determinants of customer churn in the telecommunication service area. The determinants of customer churn benefits {point highlight that call quality-associated conditions persuade customer churn; conversely, customers entering into membership card programs are too, greater possible to churn.

A customer position totally transform clarifies the association among churn determinants along with probability of churn. The probability of churn is recognized through Naive Bayes classifier (NBC).

Table 1: Varible Description

Variable name	Description
	Customer dissatisfaction (H_1)
Call drop	Rate proportion of call drops a customer experiences out of the overall number of call trials
Call failure	Rate proportion of call breakdowns a customer experiences out of the overall number of call trials
Number of complaints	Number of times a customer makes complaints to customer service center about the difficulties with billing, call coverage, membership cards, etc
	Switching costs (H_2)
Loyalty points	Quantity of credits customers earned, which are valid for a wide variety of goods and services, for instance, retail gifts and coupons
	Service usage (H_3)
Billed amounts	Total monthly charge
Unpaid balances	Overall unpaid balances
Number of unpaid monthly bills	Number of times in which a customer did not pay his/her monthly bills in time
Customer status (H_4)	Customers with a non-use or suspended status are considered more likely to churn than customers with an active use status

Naive Bayes learning produces a probabilistic model of the customer behaviour data. Assumed the training set of instances, each is characterised as a vector of features [H_1,H_2,H_3,H_4], the assignment is gaining knowledge from data to be prepared to forecast the majority of feasible customer behaviour class, of the latest instance whose category is unidentified. Naive Bayes applies the Bayes's theorem to think the probability of the classes. The probabilities are relies on four circumstances or popular features of customer behaviour.

$$P(y_j|H_1,H_2,H_3,H_4) = \frac{P(y_j)P(H_1,H_2,H_3,H_4|y_j)}{P(H_1,H_2,H_3,H_4)} \quad (1)$$

where (y_j) is the prior probability of class y_j which is projected as its existence frequency in the training data. $P(y_j|H_1,H_2,H_3,H_4)$ is the subsequent probability of class y_j after observing the data. $P(H_1,H_2,H_3,H_4|y_j)$ denotes the conditional probability of observing an occurrence with the feature vector $[H_1,H_2,H_3,H_4]$ among those having classyj. And (H_1,H_2,H_3,H_4) is probability of detecting an instance with feature vector $[H_1,H_2,H_3,H_4]$ regardless of the class. Ever since the amount of the succeeding probabilities, completely classes is one $\Sigma_{y_j \in \mathbb{C}} P(y_j|H_1,H_2,H_3,H_4) = 1$, denominator on eq. (1)'s right hand side is normalizing factor & can be omitted.

$$P(y_j|H_1,H_2,H_3,H_4) = P(Y_j)P(H_1,H_2,H_3 \quad (2)$$

An instance will be labeled as the particular class which has the highest posterior probability y_{MAP}.

$$y_{MAP} = argmax_{y_j \in \mathbb{C}} P(y_j)P(H_1,H_2,H_3,H_4|yj) \quad (3)$$

In order to estimate the term $P(H_1,H_2,H_3,H_4|y_j)$ by counting frequencies, one needs to have a huge training set where every possible combinations $[H_1,H_2,H_3,H_4]$ appear many times to obtain reliable estimates. Naive Bayes solves this problem by its Naïve assumption that features that define instances are conditionally independent given the class. Therefore the probability of observing the combination $[H_1,H_2,H_3,H_4]$ is merely the product of the probabilities of viewing every individual characteristic value $P(H_1,H_2,H_3,H_4|y_j) = \prod_{i=1}^{d=4} P(H_i|y_j)$. Substituting this approximation into equation (3) to derive the Naive Bayes classification

$$y_{MAP} = \underset{y_j \in \mathbb{C}}{argmax} P(y_j) \prod_{i=1}^{d=4} P(H_i|y_j) \quad (4)$$

The customers churn probability in accord with the H_1,H_2,H_3,H_4 classifiers of customer. The over-all reports achievable probabilities are estimated with the help of NBC. Subsequently, the probability churn relationship is compared against the whole churn determinants conditions. Finally, the customer churn behaviour is examined.

5. Attribute selection using BAT algorithm

For prediction of particular results attribute selection provides an important role to enhances the prediction and reduces the curse dimensionality issue. Attribute selection is targets to locate the most essential information from a given list of attributes. Since this task can be viewed like an optimization issue, the combinatorial development of the possible remedies might be in-viable for a thorough search. In this particular research, 20 kinds of attributes are viewed just as, Gender, Senior Citizen, Partner, Dependent, tenure, Phone Service, Multiple Lines, Internet Service, Online Security, Online Backup, Device Protection, Tech Support, Streaming TV, Streaming Movies, Contract, Paperless Billing, Payment Method, Monthly Charges, Total Charges, Churn. These types of attributes are included information about Customers who left during the last month – the column is called Churn. Services that every customer has signed up for – phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies. Customer account information – how long they've been a customer, contract, payment method, paperless billing, monthly charges, and total charges. The demographic details of customers such as– gender, age range, and also verifies if they possess partners and dependents. These kinds of attributes are classified as numerical and categorical. The categorical attributes are contained 'yes' or 'no' and '0' or '1' etc., extremely important attributes or variables of the customer are chosen depending on the bat algorithm.

To start with, the initial attribute x_i, velocity v_i and frequency f_i are initialized for each attribute b_i. For each time step t, being T the maximum number of iterations, the movement of the virtual attributes is given by updating their velocity and position using Equations 5, 6 and 7, as follows:

$$f_i - f_{min} + (f_{min} - f_{max})\beta \quad (5)$$

$$v_i^j(t) = v_i^j(t-1) + [\hat{x}^j - x_i^j(t-1)]f_i \quad (6)$$

$$x_i^j(t) = x_i^j(t-1) + v_i^j(t) \quad (7)$$

where β denotes a randomly generated number within the interval [0, 1]. Recall that $x_{ji}(t)$ denotes the value of decision variable j for attribute i at time step t. The result of f_i (Equation 1) is used to control the pace and range of the movement of the customers. The variable $\hat{x}_j$ represents the current global best location (solution) for decision variable j, which is achieved comparing all the solutions provided by the m attributes.

In order to improvise the variability of the possible solutions, primarily, one solution is selected among the current best solutions, and then the random walk is applied in order to generate a new solution for each bat that accepts the condition in Line 5 of Algorithm 2:

$$x_{new} = x_{old} + \in \bar{A}(t) \quad (8)$$

In which $\bar{A}(t)$ stands for the average noisiness of all the attributes at time t, and $\epsilon \in [-1, 1]$ attempts to the direction and strength of the random walk. For each iteration of the algorithm, the noisiness A_i and the emission pulse rate r_i are updated, as follows:

$$A_i(t+1) = \alpha A_i(t) \quad (9)$$

And

$$r_i(t+1) = r_i(0)[1 - exp(-\gamma t) \quad (10)$$

Where α and γ are ad-hoc constants. At the first step of the algorithm, the emission rate $r(0)$ and the loudness $A_i(0)$ are often randomly chosen.

Algorithm 2: Attribute selection using BAT algorithm
Input: set of attributes contains customer's information
Output: important attributes are extracted
Objective function $f(x), x = (x^1, \ldots, x^n)$.
Initialize the customers information x_i and $v_i, i = 1, 2, \ldots, m$
Define pulse frequency f_i at $x_i, \forall_i = 1, 2, \ldots, m$.
Initialize pulse rates r_i and the noisiness $A_i, i = 1, 2, \ldots, m$.
1. While $t < T$
2. For each attributes bi, do
3. Generate new solutions through Equations (5), (6) and (7).
5. If $rand > r_i$, then
6. Select a solution among the best solutions.
7. Generate a local solution around the best solution.
9. If $rand < A_i$ and $f(x_i) < f(\hat{x})$, then
10. Accept the new solutions.
11. Increase r_i and reduce A_i.
12. Rank the bats and find the current best $\hat{x}$.

With the above expression, the most important information's are examined from a given set of attributes. And due to this result increases the prediction and reduces the notable dimensionality problem.

6. Customers Churn Prediction Using KELM

The customer churn separation or prediction is reviewed using kernelized extreme learning machines. The KELM is Single Hidden Layer Feed Forward Networks (SLFNs) with different type of hidden attributes as well as kernels and shows that the simple unified algorithm of ELM can be received for customer churn prediction result.

The initial parameters of hidden layer need not be tuned and almost all nonlinear piecewise continuous functions can be used as the hidden attributes. Therefore, for N arbitrary distinct samples$\{(x_i, t_i)\}_{i=1}^N$ where x_i is the input feature vector and $t_i \in R^m$ is the corresponding target vector, the output of a SLFN network with L hidden attributes can be expressed as the following:

$$f_L(x) = \sum_{i=1}^{L} \beta_i h_i(x) = h(x)\beta \quad (11)$$

Where $\beta = \beta_1, \beta_2, \ldots, \beta_L$ is the vector of the output weights between the hidden layer of L customer behaviour attributes and the output customer churn prediction and $h(x) = [h_1(x), h_2(x), \ldots, h_L(x)]$ is the output vector of the hidden layer with respect to the input x, which maps the data from input space to the ELM feature space.

For decreasing the training error and improving the generalization performance of churn prediction, the training error and the output weights should be minimized at the same time, that is,

$$minimize: \|H\beta - T\|, \|\beta\| \quad (12)$$

The least squares solution of (2) based on Karush–Kuhn–Tucker (KKT) conditions can be written as

$$\beta = H^T \left(\frac{1}{C} + HH^T\right)^{-1} T \quad (13)$$

Where H is the hidden layer output matrix, C is the regulation coefficient, and T is the expected output matrix of customer attributes. Then, the output function of the ELM learning algorithm is

$$f(x) = h(x) H^T \left(\frac{1}{C} + HH^T\right)^{-1} T \quad (14)$$

If the feature mapping $h(x)$ is unknown and the kernel matrix of ELM based on Mercer's conditions can be defined as follows:

$$M = HH^T : m_{ij} = h(x_i)h(x_j) = k(x_i, x_j) \quad (15)$$

Thus, the output function (x) of the kernel based extreme learning machine (KELM) can be written compactly as

$$f(x) = [k(x, x_1), \dots, k(x, x_N)] \left(\frac{1}{C} + M\right)^{-1} T \quad (16)$$

where$M = HH\ T$ and (x, y) is the kernel function of hidden neurons of single hidden layer feed-forward.

In this paper, Gaussian kernel function is used for simulation and performance analysis and the chosen kernel functions are as follows.

$$k(x, y) = exp\,(-a\|x - y\|) \quad (17)$$

Moreover, the KELM learning algorithm achieves better churn prediction performance and is more stable compared to ELM.

7. Experimental Results and Discussion

Through this section, on the way to evaluate the functionality of KELM churn prediction scheme applied a training set of customer information collected over a six-month period. Each and every customer is classified into one of two pre-programmed groups and his/her churn propensity is supervised and modified in accord with his/her most recent three-month information. In this way, it will be simple to represent the real-world setting of churn prediction. A churn prediction system is assumed to be calculated through its capacity to recognize churners for marketing purpose, and in this work use the Receiver Operating Characteristic (ROC) curve and top-quantile-lift values to produce an entire evaluation of KELM prediction scheme. Eventually, the results with a HSVM and logistic regression model and Ada boost model without customer separation is compared.

8. Data Collection

For this research work, a data set is chosen from a telecommunication organization which encompass a segment of mobile customers. The dataset is IBM Watson Analytics Telco Customer Churn data from https://community.watsonanalytics.com/predictive-insights-in-the-telco-customer-churn-data-set/. This data set presents information of behavior to retain customers. A telecommunications company is concerned with the number of customers leaving their landline business for cable competitors. They have to recognize who is leaving. So they have to analyst at this organization and identify who will be leaving and why. The data set includes information about: Customers who left within the last month – the column is called Churn. Services that each customer has registered for – phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies. Customer account information – how long they've been a customer, contract, payment method, paperless billing, monthly charges, and total charges.

9. Evaluation Criteria

In this study, the Area Under receiver Curve (AUC), sensitivity, and specificity [19] are used to quantify the accuracy of the predictive models.

If True Positives (TP), False Positives (FP), True Negatives (TN) and False Negatives (FN) are the TP, FP, TN and FN in the confusion matrix, then the sensitivity is $(TP/(TP + FN))$: The proportion of positive cases which are predicted to be positive.

The specificity is $(TN/((TN + FP))$: The proportion of negative cases which are predicted to be negative [19]. To assess the accuracy of a classifier independent of any threshold, ROC analysis can be used. The horizontal axis and the vertical axis of an ROC curve are defined by Equations 18 and 19 respectively [19].

$$x = 1 - \text{specificity}\ (t) \quad (18)$$

$$y = \text{sensitivity}\ (t) \quad (19)$$

To measure the accuracy of a model, the AUC can be measured [19, 20].

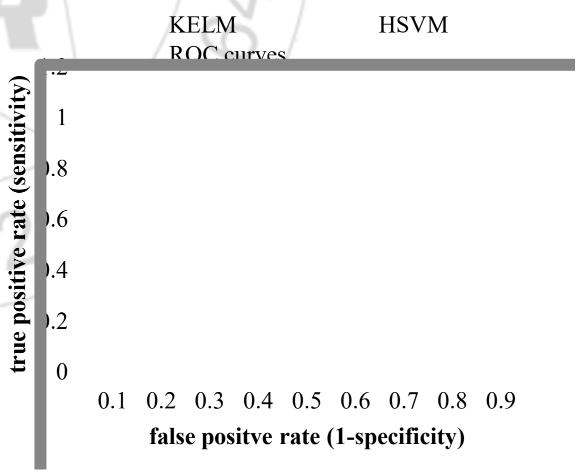

Figure 2: ROC curves of the separation

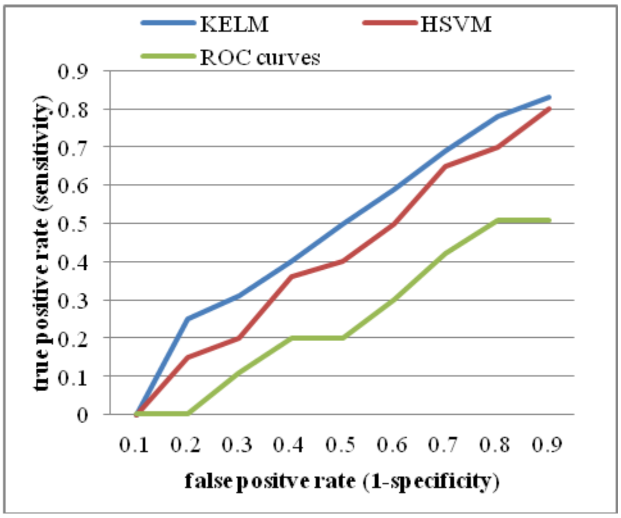

Figure 3: ROC curves of predictions on samples

Fig. 3 shows that, for customers who belong to the training samples, the ROC curve of proposed churn prediction model is located above but close to the curve of the HSVM and AdaBoost learning models, with the Curve increased from 80.13 % to 83%. KELM model achieves a much better result than existing HSVM and Adaboost methods.

9.1 Processing Time Comparison

The KELM prediction model discovers frequent churn prediction and much greater efficiency than the existing prediction model of HSVM and AdaBoost shown in Fig.4 .The proposed KELM prediction model takes less computation time to predict the customer churn when compared to existing system. The KELM is reportedly working efficiently and in many cases, it's much faster than HSVM and AdaBoost.

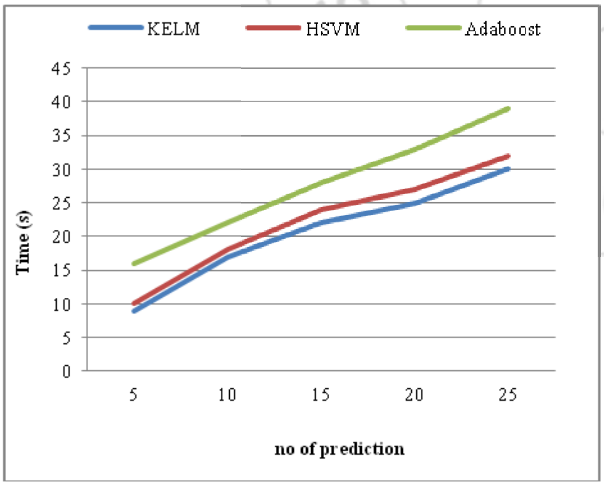

Figure 4: Processing time comparison

9.2 Accuracy Comparison

The KELM prediction model discovers frequent churn prediction and much greater accuracy results than existing prediction model of AdaBoost and logical regression shown in Fig 5. When the number of prediction is increases the accuracy of the result is increases. The proposed KELM produces high accuracy rate when compared to existing system. With the Area Under receiver Curve increased from 98.13 % to 98.3%. The KELM is reportedly working efficiently and in many cases produces high accuracy rate than HSVM and AdaBoost.

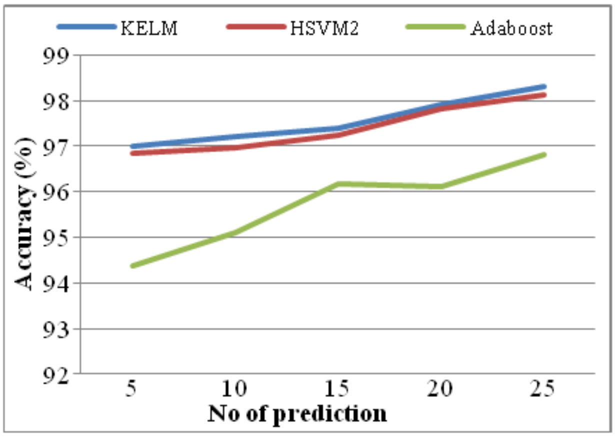

Figure 5: Accuracy comparison

10. Conclusion

Customer churn prediction 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. 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 research, Kernelized Extreme Learning Machine (KELM) algorithm is presented. The data is cleaned using preprocessing with Expectation Maximization (EM) clustering algorithm. Then customer churn behavior is analyzed by using Naive Bayes Classifier (NBC). The attributes are evaluated using BAT algorithm and KELM algorithm used for churn prediction. The experimental results show that proposed model is better than AdaBoost and Hybrid Support Vector Machine (HSVM) models in terms of the performance of ROC, sensitivity, specificity, accuracy and processing time. The future work, enhance new methods for data separation and attribute selection. And other classification methods are used to predict the churn efficiently.

References

[1] Y. Huang, B. Huang, and M. T. Kechadi, "A rule-based method for customer churn prediction in telecommunication services," in Advances in Knowledge Discovery and Data Mining, Springer, 2011, pp. 411–422.

[2] A. Idris and A. Khan, "Customer churn prediction for telecommunication: Employing various various features selection techniques and tree based ensemble classifiers," in Multitopic Conference (INMIC), 2012 15th International, 2012, pp. 23–27.

[3] M. Kaur, K. Singh, and N. Sharma, "Data Mining as a tool to Predict the Churn Behaviour among Indian bank

customers," Int. J. Recent Innov. Trends Comput. Commun., vol. 1, no. 9, pp. 720–725, 2013.
[4] V. L. Miguéis, D. Van den Poel, a. S.Camanho, and J. Falcão e Cunha, "Modeling partial customer churn: On the value of first product-category purchase sequences," Expert Syst. Appl., vol. 39, no. 12, pp. 11250–11256, Sep. 2012.
[5] N.Kamalraj and Dr.A.Malathi, ―Applying Data Mining Techniques in Telecom Churn Prediction, Volume 3, Issue 10, October 2013.
[6] Amal M. Almana , Mehmet Sabih Aksoy and Rasheed Alzahrani,Data Mining Techniques In Customer Churn Analysis For Mobile industry Amal M. Almana et al Int. Journal of Engineering Research and Applications ISSN : 2248-9622, Vol. 4, Issue 5(Version 6), May 2014, pp.165-171.
[7] K. bintiOseman, S. binti M. Shukor, N. Abu Haris, and F. bin Abu Bakar, ―Data Mining in Churn Analysis Model for Telecommunication Industry,‖ J. Stat. Model. Anal., vol. 1, no. 19–27, 2010.
[8] C.-P.Wei and I. T. Chiu, ―Turning telecommunications call details to churn prediction: a data mining approach,‖ Expert Syst. Appl., vol. 23, no. 2, pp. 103–112, 2002.
[9] A.T. Jahromi, M. Moeini, I. Akbari, and A. Akbarzadeh, ―A Dual-Step Multi-Algorithm Approach For Churn Prediction in Pre-Paid Telecommunications Service Providers,‖ RISUS. J. Innov. Sustain, vol. 1, no. 2, 2010.
[10] V. Yeshwanth, V. V. Raj, and M. Saravanan, ―Evolutionary Churn Prediction in Mobile Networks Using Hybrid Learning,‖ in Twenty-Fourth International FLAIRS Conference, 2011, pp. 471–476.
[11] M. Kaur, K. Singh, and N. Sharma, ―Data Mining as a tool to Predict the Churn Behaviour among Indian bank customers,‖ Int. J. Recent Innov. Trends Comput. Commun. vol. 1, no. 9, pp. 720–725, 2013.
[12] R. A. Soeini and K. V. Rodpysh, ―Evaluations of Data Mining Methods in Order to Provide the Optimum Method for Customer Churn Prediction: Case Study Insurance Industry,‖ 2012 Int. Conf. Inf. Comput. Appl. (ICICA 2012), vol. 24, pp. 290–297, 2012.
[13] J. Hadden, A. Tiwari, R. Roy, and D. Ruta, ―Churn prediction: does technology matter,‖ Int. J. Intell. Technol., vol. 1, no. 2, 2006.
[14] W. H. Au, K. C. C. Chan, and Y. Xin, "A novel evolutionary data mining algorithm with applications to churn prediction," Evol. Comput. IEEE Trans., vol. 7, no. 6, pp. 532– 545, 2003.
[15] S. A. Qureshi, A. S. Rehman, A. M. Qamar, A. Kamal, and A. Rehman, ―Telecommunication subscribers' churn prediction model using machine learning,‖ in Digital Information Management (ICDIM), 2013 Eighth International Conference on, 2013, pp. 131–136.
[16] S.A. Qureshi, A. S. Rehman, A. M. Qamar, A. Kamal, and A. Rehman, "Telecommunication subscribers' churn prediction model using machine learning," in Digital Information Management (ICDIM), 2013 Eighth International Conference on, 2013, pp. 131–136.
[17] V. Lazarov and M. Capota, "Churn Prediction," TUM Comput. Sci., 2007.
[18] Y. Zhang, J. Qi, H. Shu, and J. Cao, "A hybrid KNN-LR classifier and its application in customer churn prediction," in 2007 IEEE International Conference on Systems, Man and Cybernetics, 2007, pp.3265–3269.
[19] Han J. and Kamber M., Data Mining Concepts and Techniques, Morgan Kaufmann, San Francisco, USA, 2006.
[20] Burez J. and Van D., "Handling Class Imbalance in Customer Churn Prediction," Expert Systems with Applications, vol. 36, no. 3, pp. 4626-4636, 2009.

Author Profile

S. Induja received her **M.Sc(IT),** degree from Bharathidasan University, **M.Phil** degree from Annamalai University. She is having 10 years of teaching experience and presently working as Assistant Professor in the Dept. of Computer Science, Tiruppur Kumaran college for Women, Tirupur. She is pursuing her PhD at Periyar Univerrsity, Coimbatore. Her area of interest includes Data Mining, Computer Networks and Cryptography and Network security, Software engineering.

Dr. V. P. Eswaramurthy received his **Doctrate** from Anna University ,Chennai **, M.C.A., M.Phil** degree from Bharathiar University, Coimbatore. He is having 22 years of teaching experience and presently working as Assistant Professor in the Dept. of Computer Science, Government Arts and Science College, Komarapalayam, Tamilnadu. His area of interest includes Data Mining, Computer Networks and Cryptography and Network security, Software engineering , Neural Networks.