Hindusthan International Journal of Multidisciplinary Research (HIJOMR)
Volume 1 Issue 1, September 2024, pp.68-80

CUSTOMER CHURN PREDICTION IN TELECOMMUNICATIONS INDUSTRY USING MACHINE LEARNING TECHNIQUES

Dr.K.Mythili[1], Dr.S.Santhana Megala[2], B.Dhineshkumar[3]

[1] Professor & Head, Department of Computer Technology,
Hindusthan College of Arts and Science, Coimbatore.
mythili.k@hicas.ac.in

[2] Assistant Professor, Department of Computer Technology,
Hindusthan College of Arts and Science, Coimbatore.
Santhanamegala.s@hicas.ac.in

[3] Student, Department of Computer Technology,
Hindusthan College of Arts and Science, Coimbatore.
dhineshkumarb6@gmail.com

ABSTRACT

Data mining techniques have been developed recently to address the challenging problems of customer attrition in the telecommunications service industry This paper introduces a novel feature set for predicting land-line customer attrition. This paper takes the features such as two six-month segmentations, accurate four month call details, line data, account data, bill and payment details, demographic profiles, service orders, complaint data, etc. Afterwards, based on the new features, the three prediction techniquesLinear Classifications, Naive Bayes, Support Vector Machines Algorithm were used to predict customer turnover. Ultimately, the three modelling methodologies and the new feature set were assessed through comparative tests in order to predict client attrition. The experimental findings indicate that for customer churn prediction in the telecom services industry, the new features utilising the six modelling methodologies are more successful than the current ones.

Keywords: *Churn Prediction, Linear Classifications, Naive Bayes, Support Vector Machine.*

1. INTRODUCTION

A phenomenon known as customer churn affects service providers, particularly those in the telecommunications industry, as they lose important clients to rivals. The telecommunications sector has seen numerous changes in recent years, including new services, new technologies, and market liberalisation that increased competition. Customer attrition is a major issue that results in a significant loss of telecom services.

As one of the most important customer retention tactics, churn prediction has caused concern in the telecom industry and research[17].Over the past ten years, the majority of research on churn prediction has focused on voice services that are available over mobile and fixed-line networks. Fewer researchers have examined the churn prediction of land-line services than those of mobile ones.

Figure1: Customer Churn in Telecom

In the mobile telecommunications industry, churn prediction is typically achieved by using customer demographics, contractual data, customer service logs, call details, complaint data, bill and payment information, and other data Compared to mobile services, land-line service providers have fewer qualifying data available. Data types used by landline and mobile communication services differ. Data from landline communication service providers is sometimes inaccurate, missing, or incomplete. Problem reports, client ages, complaint information, and call specifics for the preceding few months are all absent in some instances.[7]

This work introduces a novel collection of features with three modelling methodologies to increase the precision of customer churn prediction in the telecom service industry. The two six-month Henley segments, exact four-month call details, grant information, line information, bill and payment information, account information, demographic profiles, and service orders all of which are gleaned from the scant information that is now availableare the new features. Logic Regression, Naive Bayes, Support Vector Machines are the modelling techniques. Lastly, the comparative studies are completed..According to the experimental results, the six modelling methodologies and characteristics that have been provided are more successful than the current features for predicting customer churn in land-line communication services.

The remainder of this essay is structured as follows: The evaluation standards for churn prediction systems are presented in the following section. Our methodology, which includes feature extraction, normalisation, and prediction algorithms, presents the experimental data along with a commentary, concludes this paper and makes recommendations for further work.

2. LITERATURE REVIEW

There are no publicly available datasets for churn prediction because to privacy and commercial confidentiality concerns.[17]presented a set of features for churn prediction in the field of land-line telecommunication services. These features include the length of service use, the type of payment, the amount and structure of monthly service fees, proportions variables, consumption level rates variables, and the growth rates of the second three months. A collection of features, including a six-month segmentation, line information, bill and payment information, account information, call details, and service log data, among other things, was recently given in [10]. Furthermore, the majority of the literature [23]demonstrates the characteristics that the combined call information are crucial for predicting customer attrition. By adding together the length of time, costs, and the quantity of calls of any kind made throughout each period, these properties are obtained. Nonetheless, the call information can be further subdivided into more specific categories based on the various call kinds. For churn prediction, this more accurate data may be more helpful than the current call detail features.

3. METHODOLOGY

This work typically consists of three phases: Data Preparation, Classification/Prediction, And Data Sampling. By definition, data sampling chooses a group of customers at random with relevant information. Data cleaning, feature extraction, and normalisation procedures are all part of the data preparation phase, which is also known as data pre-processing. Data cleaning eliminates unnecessary information, such as duplicated data, missing values, special mathematical symbols, human error-caused misspellings, strings that end in "NULL," and so forth. A collection of features that represent customers are extracted through feature extraction. The normalisation process puts feature values into a range, such as between 0 and 1. The projection phase makes predictions about how customers could behave in the near future.The focus of this paper is mostly on feature extraction. The next subsections include descriptions of the feature/variables extraction, normalisation, and prediction/classification processes.

3.1. Feature/variable Extraction

Feature extraction is one of the most significant aspects that might affect how well predictive models function in terms of prediction rates (high TP and low FP). The prediction rates of TP and FP can be greatly increased and decreased, respectively, if a strong set of characteristics can be extracted during this phase. Nevertheless, obtaining such a good set of traits is not simple. Up to now, the majority of feature sets have been introduced for churn prediction in fixed-line telecommunication [10]and mobile telecom business.[7]These current feature sets, meanwhile, still have room for improvement. The following is a description of the new features we propose in this research for customer churn prediction in telecommunication service fields:

- **Demographic profiles:** these depict a market segment or grouping based on demographic data that includes predicted consumer behaviour. Typically, this data consists of gender, age, and social class groups. Gender and county data are chosen as two new features and are offered.

- **Grants information:** a few clients received unique grants that allowed a third party to pay all or a portion of their costs (note: only one new feature is selected). Customers who are over 80 or have a disability, for instance, are more likely to stick with the services.

- **Customer account information:** This includes details about the various service packages available, credit controller and junk mail indicators, the date the account was created, the frequency of bills, the amount of money in the account, the types of payments made, and the summarised attributes (call duration, number of calls, standard prices and fees, current outstanding charges, and charges paid). The customer accounts are described in this data quite simply (we think that one account number corresponds to one customer). Furthermore, the account information wraps together the bitstream access networks, Local Loop Unbundling (LLU), and broadband indicators.These indicators show whether a customer uses LLU services, is a bit stream user, and has access to broadband service. These three indicators might not be easily located in a single file, unlike other account information (e.g. a table in the database). They are typically located by merging several related files (such as tables) in a database. For example, by combining a few files (tables) of broadband services with the account file (or table), one can find the broadband indicator. Consequently, using a classification model, all of this account data may be used to distinguish between churners and nonchurners.

Customers may receive bills at different intervals. For a predetermined amount of time, the parameters pertaining to call duration, call volume, standard fees, and paid fees must be recalculated. The chosen duration is 60 days because the majority of clients purchased the bills on a monthly basis. Let's say that "Dur," "Ncall," "Standfees," and "Paidfees" stand for the length of the call, the number of calls, the standard fees, and the actual fees paid, in that order. Eq. (4) can be used to recalculate them.

$$Ncall' = \frac{Ncall_M}{nDays} * 60 \tag{1}$$

$$Dur' = \frac{Dur\_M}{nDays} * 60 \tag{2}$$

$$Standfees' = \frac{Standfees_M}{nDays} * 60 \tag{3}$$

$$Paidfees' = \frac{Paidfees_M}{nDays} * 60 \tag{4}$$

where, which may be found using Equation (5), "Ncall_M" is the number of calls in the most recent bill, "Dur_M" is the bill's duration, "Standfees_M" is the bill's fees, "Paidfees_M" is the bill's customer fees, and "nDays" is the bill's day count.

$$nDays = ENDDate - startDate \tag{5}$$

where "startDate" and "endDate" denote the beginning and ending dates of the bill, respectively.

• **Service orders:** list the services that the client has requested. We only choose the data from the most recent three service orders because an order list may be quite lengthy. The amount of services purchased, the rental cost, the deadline, and the date that services were approved as new features are all included in this information.

• **Henley segments:** based on traits, requirements, and commercial value, the Henley segmentation algorithm splits consumers and potential customers into several categories or levels. The individual and discriminant segments are the two different kinds of Henley segments. Ambitious Techno Enthusiast (ATE) and Communications Intense Families (CIF) Henley segments are included in the individual segment. Mutually exclusive segments (DS) are discriminant segments that might symbolise client loyalty. The most recent two six-month Henley segments are chosen as new input features. Similar to this, neutral data is used to fill in the gaps in the Henley segments. [8]

• **Telephone line information:** this includes details about the number, kind and district codes of telephone lines as well as if voice mail service is offered. Customers with many phone lines may find the services more appealing and be more inclined to stick with them. A prediction model cannot benefit from this information. As a result, the voice mail service indication, district codes, and phone line count are chosen as new features.

• **Complaint Information:** Customers who have voiced complaints about the telecommunications services have provided this information. Consumers can file complaints about the services via email, letter, or, more commonly, phone call to several departments. Most of the complaint information's specifics are absent from the database. On most occasions, nevertheless, it is possible to locate the indicator that shows the identity of the complainant on a given date. As a result, this paper solely chooses this indicator as a new feature from complaint data.

• **The historical billing and payment data:** this includes the billing details for every client and service over a predetermined period of time. Only our prediction system uses the call data from the previous four months, which are available. The last four months' worth of bill information are chosen so as to include the call specifics for that four months.

As was already indicated, different clients might get bills at different times. As a result, client bills may have varying lengths. We divide the bill whose duration is longer than one month into several monthly bills, each covering one month, in order to ensure that bill durations are consistent. For instance, we must divide a customer's one two-month charge into two separate monthly bills. The total standard charge fees, total rental charges, total value added tax (VAT), total call time, and total fees paid are all detailed in each monthly bill. The eight call kinds local calls, international calls, other sorts of calls, and mobile phone calls are also detailed in each bill. A monthly statement includes a summary of the length of the call, normal fees, actual charged fees, and fees paid for each type of call.

As additional features, this paper extracts the total of the following: total of the standard charge fees, total of the rental charges, total of the value added tax (or total VAT), total of the call duration, total of the fees paid, total of the call duration of each type of call, the standard fees of each type of call, the actual charged fees of each type of call, and the fees paid of each type of call. Let us assume that for the ith month, the total standard charge fees, total rental charges, total VAT, total call length, and total fees paid are, respectively, "TotSFi," "TotRCi," "TotVATi," "TotDi," and "TotFPi." Equation (6) can be used to compute the total of the following: the total of all standard charge fees ('SSF'), the total of all rental charges ('SRC'), the total of all VAT ('SVAT'), the total of all call duration ('SD'), and the total of all fees paid ('SFP').

$$SSF = \sum_{i-1}^{4} TotSF_i \tag{6}$$

$$SRC = \sum_{i=1}^{4} TotRC_i \tag{7}$$

$$SVAT = \sum_{i=1}^{4} TotVAT_i \tag{8}$$

$$SD = \sum_{i=1}^{4} TotD_i \tag{9}$$

$$SFP = \sum_{i=1}^{4} TotFP_i \tag{10}$$

Let the type of calls be k. Consider that the call duration, the standard fees, the actual charged fees and the fees paid for the I th month are MD_k^i, MSF_k^i, MAF_k^i and MPF_k^i, respectively. Let the change of the call duration, and the change of the standard fees, the change of the actual charged fees and the change of the fees paid between the two consecutive months I th and i - 1th for the type of calls k be $ch\_MD_k^{i,i-1}$, $ch\_MSF_k^{i,i-1}$, $ch\_MAF_k^{i,i-1}$, and $ch\_MPF_k^{i,i-1}$, respectively. This subset of information is extracted as new features by Eq. (13).

$$ch\_MD_k^{i,i-1} = \frac{|MD_k^i - MD_k^{i-1}|}{\sum_{i=2}^{4}|MD_k^i - MD_k^{i-1}|} \tag{11}$$

$$ch\_MSF_k^{i,i-1} = \frac{|MSF_k^i - MSF_k^{i-1}|}{\sum_{i=2}^{4}|MSF_k^i - MSF_k^{i-1}|} \tag{12}$$

$$ch\_MAF_k^{i,i-1} = \frac{|MAF_k^i - MAF_k^{i-1}|}{\sum_{i=2}^{4}|MAF_k^i - MAF_k^{i-1}|} \tag{13}$$

$$ch\_MPF_k^{i,i-1} = \frac{|MPF_k^i - MPF_k^{i-1}|}{\sum_{i=2}^{4}|MPF_k^i - MPF_k^{i-1}|} \tag{14}$$

Call details: these include the length of the call, the cost, and the kind of call (local or international, for example) for each and every call. As a result, keeping track of every call's information for every consumer, every month, is challenging. The majority of telecom providers retain call logs dating back a few months. Therefore, the restricted call details are only useful for forecasting client attrition. However, they can still show the frequency of service usage by comparing prices, length of call, and other factors. For instance, the short call length suggests that the clients weren't using the services frequently; In the future, the clients may decide to stop using the services. Should there be an abrupt change in the cost of the services, the client may decide to stop using them earlier. The literature reports on the application of call information to churn prediction. But the majority of researchers simply chose as features the total number of calls, duration, and fees from all different kinds of phone calls.The number of calls, duration, and fees were combined by the authors and were not further separated into categories for international, local, national, mobile phone, free, and whole sale line calls. On the other hand, this split total number of calls, length, and fees might better represent the state of service usage and be more useful for churn prediction. From the call information of the previous four months, this research extracts the number, duration, and fees as new features. The number of calls, duration, and fees that are changed are also regarded as new features based on these features. The following is a description of the process used to extract these additional features the call details into several distinct periods totalise the duration, fees, and calls for each period for all calls made by each customer, including local, national, international, and mobile calls the sum of these total calls, duration, and fees represents the total calls, duration, and fees for the previous four months the variation in calls, duration, and fees can be obtained. According to literature, calling every 15 or 20 days is effective. Thus, in this article, the defined term is 15 days[23].

Let "NCALL," "DUR," and "FEES" stand for the number of calls, duration, and fees, respectively. Additionally, let x stand for any kind of call—local, national, international, mobile, or otherwise. Assume that for the kind of calls x, the total number of calls, duration, and fees on segment i are "$NCALL_x^i$," "DUR_x^i," and "$FEES_x^i$," respectively. Determine the altered call count, duration, and costs between two successive segments i and i 1 of the call type x.

$$ch\_DUR_x^i = \frac{|DUR_x^i - DUR_x^{i-1}|}{\sum_{j=2}^{M'} |DUR_x^i - DUR_x^{j-1}|} \tag{15}$$

$$ch\_NCALL_x^{i,i-1} = \frac{|NCALL_x^i - NCALL_x^{i-1}|}{\sum_{j=2}^{M'} |NCALL_x^j - NCALL_x^{j-1}|} \tag{16}$$

$$ch\_FEES_x^{i,i-1} = \frac{|FEES_x^i - FEES_x^{i-1}|}{\sum_{j=2}^{M'} |FEES_x^j - FEES_x^{j-1}|} \tag{17}$$

swhere the variables ‘‘$ch\_DUR_x^{i,i-1}$’’,‘‘$ch\_NCALL_x^{i,i-1}$ ’’and‘‘$ch\_FEES_x^{i,i-1}$’’denote the altered duration, number of calls, and fees between two consecutive segments I and I - 1 of the call type x, respectively (note: the features of the free calls exclude the features "ch_FEES" and FEESi).

• **Details on incoming calls:** these refer to calls that have been received. Details on incoming calls include the length of the calls, the quantity of calls, and the associated costs. On the other hand, the majority of calls that are received are free. Incoming call fees are not included in the list of new features. The new features from the incoming call details include the number and duration of calls received, the number of calls changed, and the duration of calls changed every 15 days. In a similar veincan be used to determine changes in the quantity of calls made and the length of calls received.

3.2. Normalisation

Certain predictors or classifiers find it challenging to accept the string values of characteristics in the retrieved features. Furthermore, certain numerical features (e.g., the number of lines, duration, calls, fees) lie in different dynamical ranges, even though the values of the features (e.g., change duration, change fees, change fees paid) are in the range between 0 and 1 and these features will not be normalised. The cost functions are more impacted by these features' large values than by their little ones. It cannot, however, convey the idea that large values are more crucial to the predictor or classifier's design.

The values of the characteristics must be normalised for some predictors (such as artificial neural networks) in order to tackle the aforementioned concerns. The following is a description of the normalisation process: First, the values of string characteristics must be converted into binary strings. Secondly, the values of numerical features can be normalised using Eq. (9), putting them within a comparable range.

$$\bar{x_j} = \frac{1}{N} \sum_{i=1}^{N} x_{ij}, \quad j = 1,2 \dots, l \tag{18}$$

$$\sigma_j^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_{ij} - \bar{x_j}) \tag{19}$$

$$y = \frac{x_{ij} - \overline{x_j}}{r\sigma_j} \tag{20}$$

$$\tilde{x}_{ij} = \frac{1}{1+e^{-y}} \tag{21}$$

so that the characteristic isx_j. N is the number of occurrences or patterns, j th, i is the number of features, and r is a user-defined constant parameter. R is set to one in this investigation.

3.3. Prediction/classification

Many methods have been put forth for telecom churn prediction. Just three modelling approaches are chosen in this article to serve as churn prediction predictors. The following is an outline of these three modelling techniques:

3.3.1. Logistic Regressions (LR)

An extensively used statistical modelling method for discriminative probabilistic classification is logistic regression[9]. Logistic regression calculates the likelihood that an event will occur. The model is expressed as follows:

$$prob(y=1) = \frac{e^{\beta_0 + \sum_{k=1}^{K} \beta_k x_k}}{1 - e^{\beta_0 + \sum_{k=1}^{K} \beta_k x_k}} \tag{22}$$

where x1, x2,...,x_k are the independent inputs and Y is a binary dependent variable that indicates whether the event occurred (e.g., y = 1 if event takes place, y = 0 otherwise). Based on the given training data, the maximum likelihood technique can estimate the regression co-efficient b0, b1,...,b_k has the logistic regression models' specifics.

We require a predictive model that can do binary classification, or forecast an output variable of type 1/0 or Yes/No, based on the problem statement. Logistic regression is a popular predictive model used for binary categorisation and outcome prediction. A binary classification procedure that is a part of the generalised linear regression model is called logistic regression. More than two class problems can also be resolved with it. Using the customer turnover data, logistic regression may be used to build a model that predicts whether a specific customer or a group of customers will stop using the service.

One of the variables in the data, for instance, may be "annual income." The customer's "gender" is another variable. The logistic regression function's result will indicate how a customer's likelihood of discontinuing service is influenced by their income and/or gender.

Logistic Regression

The function for logistic regression is shown below:

P (Y=1|X)= 1 / (1+ e^(-1x(β0+β1X1+β2X2+β3X3+β4X4+⋯+βnXn)))

where β0 through βn different coefficients The independent variables influencing the dependent variable are X0 through Xn. Furthermore, the probability of a favourable result is P (Y = 1 | X).Take note of the function's exponent. This is the part where (β0+β1X1+β2X2+β3X3+β4X4+⋯+βnXn) uses linear regression.According to the graph above, the logistic regression function is a sigmoid function. The function produces values between 0 and 1, with a transition between the levels, as the graph illustrates. This feature of the function facilitates binary outcome prediction.

The output can be at level 1 or 0, depending on the values of the variables. This indicates the likelihood that the customer will leave the business or stay with it.

3.3.2. Naive Bayes (NB)

The likelihood that a given input sample belongs to a particular class is determined using a Naive Bayes classifier. To find the probability for the class $y_j$given an X sample that consists of a feature/variable vector $\{x_1,...,x_n\}$, use Eq. (23):

$$p(y_j|X = p(X|y_j)p(y_j) = p(x_1,x_2,.\dots\dots X_n|y_j)p(y_j) \tag{23}$$

where$y_j$prior probability is represented by p(y_j). Nonetheless, Naive Bayes makes the assumption that the independent variables' conditional probabilities are statistically independent. The likelihood can be expressed as follows:

$$p(X|y_j) = \prod_{i=1}^{n} p(x_i|y_j) \tag{24}$$

Thus, the posterior can be written as

$$p(y_j|X) = p(y_j)\prod_{i=1}^{n} p(x_i|y_j) \tag{25}$$

Assume that Y = $\{y_1,y_2,...,y_k\}$ is a set of classes. Therefore, a Naive Bayes classifier decides how to classify an unknown sample X by:

$$c = argmaxp(y_j|X)y_j\epsilon Y \tag{26}$$

3.3.3. Support Vector Machines (SVM)

By determining a maximal margin hyper-plane in terms of a linear combination of training set subsets (support vectors), an SVM classifier can be trained. Using the kernel approach, SVM first maps the data into a high dimensional feature space if the input feature vectors are nonlinearly separable.[2] It next classifies the data according to the maximal margin hyper-plane as follows:

$$f(\vec{x}) = sgn(\textstyle\sum_i^M y_i x_i \phi(\vec{x_i}, \vec{x} + \delta) \tag{27}$$

where / is a kernel function, ~x is an unknown sample feature vector, d is a threshold, and M is the number of samples in the training set. $x\sim_i$is a support vector with a_i> 0.

By applying linear constraints to a convex quadratic programming problem, the parameters $\{a_i\}$ can be found. In practical application, kernel functions are typically implemented using Gaussian radial basis functions (RBF) and polynomial kernels. By considering the Karush–Kuhn–Tucker condition and selecting any i for whicha_i> 0 (i.e., support vectors), d can be found. In actuality, though, it is safer to take the mean value of d across all support vectors.[4]

Because SVM can handle high-dimensional data and is especially helpful for datasets with a lot of features or when the data cannot be linearly separated, it is thought to be successful in forecasting churn. Furthermore, SVMs can deal with imbalanced datasets, which are frequently seen in churn prediction scenarios where the proportion of churning customers is significantly lower than that of non-churning customers. Therefore, anyone who is interested in learning more about SVM, classification (unsupervised learning), etc., should also work on this project.

We must first collect and prepare the data on your clients before we can utilise SVM for churn prediction. Information on their behaviour, demographics, and other pertinent factors may be included. The data must then be divided into a test set and a training set.

4. EXPERIMENTS

A classifier/predictor will be utilised to forecast the future actions of clients. Evaluating the predictive churn model's performance is one of the crucial elements in making sure the model generalises successfully. In other words, one must take a predictor's prediction rates into account. True churn rate (TP) and false churn rate (FP) are the prediction rates used in this work. The application's goal is to obtain high TP at cheap FP. TP is defined as the percentage of churn cases that were correctly identified based on the confusion matrix, as determined by

$$TP = \frac{a_{11}}{a_{11}+a_{12}}. \tag{28}$$

and FP, which is expressed as the percentage of non-churn cases that were mistakenly categorised as churn,

$$FP = \frac{a_{21}}{a_{21}+a_{22}}. \tag{29}$$

The Receive Operating Curves (ROC) approach can be used to determine the expected pair of prediction rates (TP and FP) from these pairings of TP and FP.

However, sets of pairs (TP and FP) from various prediction modelling approaches or feature subsets of data are typically difficult to analyse using the ROC technique. In this work, models and feature sets for the churn prediction are assessed using the area under a ROC curve (AUC) method as a solution to the issue. The following formula can be used to get the area under a ROC curve:

$$AUC = \frac{S_0 - n_0 \times (n_0+1) \times 0.5}{n_0 n_1} \tag{30}$$

where n_0 is the number of patterns in the test set that are class 0 and SS_0 is the sum of the ranks of the class 0 (churn) test patterns.

4.1. Experiment set-up

The following parameters were used to train each model in the experiments: There are 300 generations in the population, 30 mutations, a Z-value of 1.96 for identifying interesting traits, and 60% and 0.1%, respectively, for crossover and mutation probability.

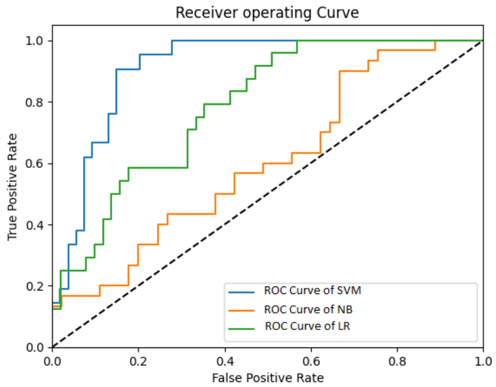

Figure2. Prediction results when using call details.

The experimental outcomes of the first series of trials, which assess the feature subsets of the call detail data for the prediction, are displayed in Fig. 1. This image has two subfigures: the relative AUC values are displayed in the other, while ROC curves are displayed in the one on the left. Each pair of the average FP and average TP from each of the six modelling techniquesfor a particular sampling rate is represented by a point in Figure 1.ThisROC figure consists of three curves with varying colours, each of which represents a particular feature subset for the various sampling rates. In Figure 1, the AUC values and various named detail kinds are shown by the y- and x-axes, respectively. This subfigure shows three distinct coloured curves: the red, grey, and black curves were plotted when FP was 50%, FP was greater than 50%, and FP was zero, respectively.

- The false churn rate may drop when the sampling rate $\frac{n\_churners}{non\_churners}$ drops (or as the number of non churners rises). However, the actual turnover rate can drop.
- TP is extremely low for any feature subset or modelling technique when FP is less than or equal to 5%. Any ROC curve's AUC values are roughly equal when FP is greater than 50%. However, the AUC values differ significantly when FP is between 50% and 100%. As a result, whether FP is 50% or 100%, most decision-makers may be more likely to take the AUC values into account.
- The prediction rates from all feature subsets are comparable (high actual churn rates and high false churn rates) when the sampling rate$\frac{n_churners}{non_churners}$ is extremely high.

The ROC curves based on the prediction rate pairs (FP and TP) obtained from the second set of trials using the six prediction modelling methodologies and various subsets. In other words, based on the eight subsets of the new features, displays the ROC curves when the prediction modelling approaches LR, NB, SVM were applied, respectively.The information of current bills, the information of the six bills with payments, the information of incoming calls (received calls), the fixed features that are typically not changed on a monthly or daily basis, the features of free calls, the features of call-details, the features of whole sale line calls, and the fullset of new features are these eight subsets of features. The AUC values were computed and shown into based on these ROC curves.The curve from the fullset of the new features that is closest to the top left corner is displayed in each subfigure. This illustrates that, when the same prediction modelling technique was applied, the fullset represents the best prediction outcomes in terms of the highest TP rates with the lowest FP.

5. CONCLUSIONS

A new set of features, such as Aggregated Call Details, Account Information, Bill Information, Dial Types, Line Information, Payment Information, Complaint Information, Service Information, and so forth, were presented in this paper for the purpose of predicting customer churn in the telecom industry. In this research, three modelling techniques LR, NB, SVM were employed as predictors. Lastly, the comparative tests were conducted using the new feature set, the old feature sets, and the three modelling methodologies.Every subset of the new feature was assessed and examined in the experiments. The studies also revealed the following: (1) the relative efficacy of the three modelling strategies; and the relative efficacy of the new feature set in comparison to the pre-existing feature sets. The experimental results demonstrated that: (1) the newly proposed feature set is more predictively effective than the pre-existing feature sets; the decision makers' objectives determine which modelling technique is best for customer churn prediction NB and SVM with a low ratio should be used if interested in the truechurn rate and false churn rate; If one is looking for the churn probability, one may use the Logistic Regressions. The call detail combination feature set, which consists of the feature sets of Local, National, International, Mobile Phone, and the sum of all call details, can obtain higher true churn rates with lower false churn rates than any individual subset. Nevertheless, our suggested methods have certain drawbacks. Future updates to the new feature set should incorporate more data in a way that enhances features. In the future, feature extraction and selection strategies will be examined in order to further minimise the dimensions of input characteristics. Furthermore, this application presents an imbalance classification problem, which we attempted to tackle solely through the use of the sampling technique. Thus, future research should concentrate on developing other techniques for imbalance classifications.

REFERENCES

[1] Au, W., Chan, C., & Yao, X. (2003). A novel evolutionary data mining algorithm with applications to churn prediction. IEEE Transactions on Evolutionary Computation, 7, 532–545.

[2] Boser, B., Guyon, I., &Vapnik, V. (1992). A training algorithm for optimal margin classifiers. In Proceedings the 5th annual ACM workshop on computational learning theory (pp. 144–152). Pittsburgh, PA: ACM Press.

[3] Bradley, A. P. (1997). The use of the area under the roc curve in the evaluation of machine learning algorithms. Pattern Recognition, 30, 1145–1159.

[4] Burges, C. J. C. (1998). A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery, 2(2), 121–167.

[5] Coussement, K., & den Poe, D. V. (2008). Churn prediction in subscription services: An application of support vector machines while comparing two parameterselection techniques. Expert Systems with Applications, 34, 313–327.

[6] Domingos, P., &Pazzani, M. J. (1997). On the optimality of the simple bayesian classifier under zero-one loss. Machine Learning, 29(2–3), 103–130.

[7] Hadden, J., Tiwari, A., Roy, R., &Ruta, D. (2006). Churn prediction: Does technology matter? International Journal of Intelligent Technology, 1(2).

[8] Hosmer, D., &Lemeshow, S. (1989). Applied logistic regression. New York: Wiley.

[9] Huang, B., Kechadi, M.-T., & Buckley, B. (2010). A new feature set with new window techniques for customer churn prediction in land-line telecommunication. Expert Systems with Applications, 37(5), 3657–3665.

[10] Hung, S.-Y., Yen, D. C., & Wang, H.-Y. (2006). Applying data mining to telecom churn management. Expert Systems with Applications, 31, 515–524.

[11] Japkowicz, N. (2000). Learning from imbalanced data sets: A comparison of various strategies (pp. 10–15). AAAI Press.

[12] Japkowicz, N. (2006). Why question machine learning evaluation methods? In AAAI Workshop. Boston.

[13] John, H., Ashutosh, T., Rajkumar, R., Dymitr, R. (2007). Computer assisted customer churn management: State-of-the-art and future trends.

[14] Jolliffe, I. T. (1986). Principal component analysis. New York: Springer.

[15] Langley, P., Iba, W., & Thompson, K. (1992). An analysis of Bayesian classifiers. In Proceedings of the 10th national conference on ARTI CIAL intelligence (pp. 223–228). MIT Press.
[16] Luo, B., Shao, P., Liu, J. (2007). Customer churn prediction based on the decision tree in personal handyphone system service. In International conference on service systems and service management (pp. 1–5).
[17] Quinlan, J. R. (1996). Improved use of continuous attributes in c4.5. Journal of Artificial Intelligence Research, 4, 77–90.
[18] Rumelhart, D., Hinton, G., & Williams, R. (1986). Learning internal representations by error propagation (Vol. 1). MA: MIT Press.
[19] Vapnik, V.N. (1998). The nature of statistical learning theory (2nd ed., pp. 23–57).
[20] Wei, C., & Chiu, I. (2002). Turning telecommunications call details to churn prediction: A data mining approach. Expert Systems with Applications, 23, 103–112.
[21] Yan, L., Wolniewicz, R., &Dodier, R. (2004). Customer behavior prediction – it's all in the timing. Potentials IEEE, 23(4), 20–25.
[22] Zhang, Y., Qi, J., Shu, H., Li, Y. (2006). Case study on crm: Detecting likely churners with limited information of fixed-line subscriber. In 2006 International conference on service systems and service management (Vol. 2, pp. 1495–1500).

AUTHOR PROFILE:

Author – 1

Dr.K.Mythili is currently working as Professor & Head of Computer Technology in Hindusthan College of Arts & Science. She has published more than 20 papers in National and International journals. Her area of interest are Data Mining, Machine Learning and Computer Networks. She has guided 10 M.Phil scholars and has a teaching experience of 26 years.

Author – 2

Dr. S. Santhana Megala is currently working as an Assistant Professor in Department of Computer Technology in Hindusthan College of Arts & Science. She has published 23 Papers in National and International Journals. Her area of Interest is Data Mining, Machine Learning and Data Analytics. She has guided 2 M.Phil scholars, 1 Ph.D scholars and guiding 3 Ph.D Scholars.

Author – 3

B. Dhineshkumar is a second year student pursuing B.Sc Computer Technology.