Customer Churns Prediction In Telecom Using Adaptive Logitboost Learning Approach

Dr.M.Hemalatha,S.Mahalakshmi

Abstract: In telecom sector, huge amount of data is generated on daily routine owing to extensive client base. Business analysts and decision makers highlighted that including new customers seems to be costly than that of maintaining prevailing customers. Customer relationship management (CRM) and Business analysts have to recognize the cause of churn customers and to validate behavioural patterns from prevailing churn customer's data. This work anticipated a churn prediction model that utilizes clustering and classification approaches to recognize churn customers and to give factors that resides behind customer churning in telecom sectors. Feature selection is carried out with association attribute ranking model and weighted oversampling technique. The proposed approach initially classifies churn customers data with Adaptive Logitboost (ALB) classification algorithm for classifying the instance more appropriately. Generation of retention policies is a significant process of relationship management to eliminate churners. After performing classification approach, these model segments customers' data by categorization using requirement similarity matrix to acquire information gain. This model validates churn factors that are needed for determining primary cause of churn. By recognizing essential churn factors from customer data, CRM may enhance productivity by recommending appropriate promotions to set of customers based on association among behavioural patterns. Thereby, marketing campaigns are improved for the company. The anticipated churn prediction model is computed with metrics like precision, accuracy, F-measure, recall and receiving operating characteristics (ROC) area. Results depict that anticipated churn prediction provides better classification with ALB and customer profiling with requirement similarity matrix. Moreover, it provides cause of customer churning through generation of related pattern association.

Key words: Churn customer, prediction model, behavioural pattern, association rule, similarity matrix.

1.INTRODUCTION

Recently, Telecom industries have been extensively witnessed from rapid progression between subscription based enterprises and businesses over last few decade [1]. Number of mobile phone users has acquired about 8 billion around world by around 2015, which is roughly equivalent to entire world population. Certain advanced countries have huge telecom subscribers than populations [2]. Henceforth, telecom markets are confronting retaining customers and customer saturation is turning to be a top priority.However, it is extensively in appropriate literature that maintaining customer with more beneficial practice than attaining new client [3]. Retaining new customers not only consumes more time period but also more cost for maintaining customer loyalty with service providers, so as to fulfil demanded services. While, customer detainment does not include any added additional expenses in marketing, indeed it needs attention to resolution to customer concern is sufficient in more cases [4]. Moreover, long term customers produce more profitable as they are not quickly attracted by other such competitors and as well may refer new customers and turns to be much less costly to attend. Therefore, fractional improvement in customer retention may considerably influences the growth and telecom business sustainability [5 Subsequently, an appropriate churn prediction system should have superior interpreting abilities which are needed essentially to recognize churn customers and the cause of churning. In general, classification method use customer features that is expressed in billing and account information, personal demographics and call details for recognizing future characteristics of customer behaviour. However, data mining techniques for churn prediction were significantly utilized to predict telecom churners more accurately [6].

- **Dr. M. Hema Latha, Professor,School of Computing-PG*
- *Sri Ramakrishna College of Arts & Science*
- *Coimbatore*S.Mahalakshmi,*
- *Research Scholar,*
- *Bharathiar University,Coimbatore.*

For instance, neural networks and decision trees have been utilized to produce churn prediction systems. However, variants of rotboost and rotation forest have also been investigated for recognizing telecom churners with superior accuracy. Author in [7] have depicted hybrid learning system that merges rule approaches and K-means clustering with a target of acquiring prediction accuracy. Alike, genetic algorithm with neural network technique has been presented in [8] to enlarge telecom churners prediction accuracy. However, enormous data mining techniques dependent on individualor ensemble classification techniques have been utilized to acquire enhanced prediction performance for customer churn in those telecom industries as provided in [9]. Moreover, numerous churn prediction techniques use sampling or feature extraction or both before utilizing classification algorithm. In [10], author has offered comprehensive review by examining influence of sampling techniques in enhancing churn prediction performance. In their investigation, advanced and random under sampling approaches were investigated with weighted random forest and gradient boosting for acquiring enhanced prediction performance. However, it is popped up that this analysis is improved with CUBE sampling approach that utilizes sophisticated sampling techniques. It will not provide appropriate enhancement in prediction outcomes, that assist in attaining the results of [11]. Alike, in [11] author stated that some duplicating instances via oversampling will not suffer considerably with advantages in enhanced prediction outcomes. This will also assists in notations that are with duplicating minority classes via random discarding and oversampling majority classes with random under sampling that will not enhance prediction outcomes. While in [12], PSO dependent intelligent under sampling approaches helps in enhanced learning of random forest and k-NN. Therefore, that results in superior churn prediction performance. Similarly, various investigations have been concentrated with the use of meaningful characteristics for inculcating classification algorithms of telecom churn prediction. Some set of features includes bill and payment, personal demographics, call details, records, service orders, account information, Henley segmentation, line information are given to decision tress, SVM and Multi-layer perceptron based NN for recognizing telecom churn. Author in [13] anticipated a Multi-objective feature selection approaches with NSGAII for telecom churn prediction. In some other approaches, Bayesian Belief Network hauls out most essential characteristics that have to be used for customer churn prediction. Alike in [14], a hybrid two phase approaches based on feature extraction has been anticipated for recognizing telecom churns. In some approaches in [15], certain amount of features from call detail records for telecom customers is attained and some logistic regression is used for churn prediction. Moreover, in some approaches, indeed of concentrating in recognizing churners with superior accuracy, intuitiveness and comprehensibility of churn prediction system are also examined to recognize cause of customer churn characteristics. Therefore, in this investigation, an effectual churn prediction approach has been provided for investigating powerful searching ability of genetic programming support via Adaptive Logitboost based iterative approach. Features that are chosen consequently in providing generic models over diverse boosting based iterations are measured to recognize factors influencing churn characteristics of telecom customers. The ultimate objective of this approach is to exploit searching and learning abilities of anticipated integrated model for generating an effectual and intuitive churn prediction approach. As well, Churn relative management with customers is provided for handling data imbalance in anticipated approach on behalf of adaptive Logitboosting ensembles. This work is arranged as Sect. II discusses background methodologies that include the idea behind generic programming. Sect III explains proposed methodology that comprises sampling and ranking method with adaptive Logitboosting approach. Sec IV depicts comparative analysis and discussions that are related to anticipate idea. At last, conclusion is drawn in Sect. V.

2.BACKGROUND WORKS

Telecom churner's prediction has been extensively analyzed regions for investigators and hence various investigators are still under work with prediction of telecom customer churn. Author in [16], have used data mining approaches such as decision tree and neural network to recognize telecom churn. They determined that both these two approaches are directly predict customer churns more appropriately with diverse customer details such as billing information, demographics, call details and so on. This work have depicted that there are some in deficiency in data for modelling building. Owing to some unavailability of customer demographics as depicted by [17], this approach has been anticipated to recognize churners outline of customer call pattern and information regarding contracts attained from details of customers call. Here, multi-classifier class combiner techniques have been used for resolving problems of skewed distribution of customers as non-churners and churners. Outcomes recommend that these approaches may provide satisfactory outcomes when compared to other demographic based churn prediction systems. In [18], author have compared the recital of extensively used Machine learning algorithms such as Decision trees, SVM, ANN, LB and NB classifiers, then applied boosting techniques and evaluated performance of boosting approach. Classifier may perform superior outcome between various models like SVM-based poly AdaBoost with an improved accuracy of 97%. In [19], author examined customer churn prediction which is a cost sensitive crisis. In contrary to other approaches, some may considers mis-classification problems. This shows some misclassification cost of every sample which is different. Partition cost sensitive CART approaches is anticipated to reduce churn prediction cost. In [20], author anticipated novel composite algorithms to construct top logistic regression and decision trees approach for recognizing customer churn. They demonstrated that logistic regression and decision trees have certain crisis like variables handling that are interactive more and more. Logit leaf model is a novel approach to categorize data in effectual manner. This model merges power of both the base algorithms as decision model which is used to construct segments and then every leaf models are built with this. Outcomes that were acquired from this show superior performance that many other advanced models. In [21], author has anticipated a novel framework for customer churn prediction approach and that is executed with WEKA data mining software. This work evaluates performance of logistic regression and decision tree with this framework. Outcomes attained from this decision tree have projects better accuracy than in contrary to logistic regression approaches. Author in [22] demonstrates that how telecom big make have the ability to predict churners more effectually and easily. Authors depicted that huge amount of data, huge variety of features and improved incoming data enhances churn prediction performance. The anticipated system was utilized as one amongst the biggest mobile operators over various countries. This system provides prepaid customers who seek to be churn in next month, with 0.96 precision for top 50,000 predicted churmers. Author in [23] anticipated an approach that recognizes customer churn by analyzing communication patterns between customers. Telecom subscriber companies are inter-connected with one another and some network properties are also influenced by churning customers. Therefore, it is essential to use social network in recognizing churning customers. This work clearly anticipated that there are some enhancements in recognizing churners by initiating network analysis as in contrary to approaches that consider personal customer information. Author illustrated that two advanced data mining approaches are ALBA and AntMiner+. There two approached are used because: AntMiner+ shows superior performing techniques that facilitate merging of domain knowledge towards it. ALBA provides superior accuracy prediction with non-linear SVM. Results demonstrates that ALBA use outcomes in increased recital by enhancing learning of classification approaches and AntMiner+ outcome in comprehensible and accurate manner. Prediction system dependent on characteristics to recognize telecom churn is depicted in [24]. This system provides attributes from services that customer avails. Therefore, usage pattern is considered to make prediction, removing problems such as missing values, feature selection and more. As well, this approach was extremely free from crisis that conventional systems have faces numerous correlations between inputs. The anticipated system made utilize clustering approach to predict customer churn. In [24], author anticipated hybrid model that merges both unsupervised and supervised approaches to recognize customer churn. This hybrid model merges traditional rule inductive approach and modified k-means clustering algorithm. This experimentation includes validation of weighted k-means clustering leads to superior partitioning, comparison of classification outcomes with other well known approach and compares of anticipated model with other hybrid classification approach. Outcomes depicts that anticipated model had superior accuracy over benchmarking dataset utilized. In [25], author anticipated a model dependent of generic programming with AdaBoost to predict churn crisis in telecommunications. This model was verified on two standard datasets. Datasets like orange and cell 2 cell with 89% accuracy for cell 2 cell dataset and 63% for other dataset. In [26], author anticipated that crisis of customer churn in big data platform. This researcher's goal is to prove that big data significantly improve process of recognizing churn based on variety, volume and data velocity. Handling data from Business Support department and Operation Support department at largest telecommunications company required big data platform to examine fractures. Random Forest procedure was utilized and computed with AUC. Author [27], anticipated a model for churn prediction with rough set theory for telecom industries. As analyzed, Rough set classification algorithm performed other algorithms like Decision tree, Linear regression and voted perception neural network. Numerous investigators have analyzed the problem related to unbalanced datasets where churned customer classes are lesser than active customer classes, as it is a foremost crisis in churn prediction approach. In [28], author evaluated six diverse sampling techniques for oversampling concerning telecom churn prediction crisis [29]. Outcomes depicts that algorithms like rule generation and MTDF based genetic algorithms outperforms other methods like oversampling approaches [30]. In this paper, the feature engineering phase is taken into consideration to create our own features to be used in machine learning algorithms. We prepared the data using a big data platform and compared the results of four trees based machine learning algorithms. This work concentrates on feature recognizing with ranking and sampling model and depends of available telecom data of companies or features in internet.

3.PROPOSED METHODOLOGY

This section explains in detail about customer churn prediction approach. Fig. [] depicts anticipated churn prediction approach and explains it in detail. Initially, data pre-processing is carried out with filtering model to eliminate noise, balancing data features and data normalization. Some essential features are hauled out by analyzing information gain, ranking the attributed and correlating these weighted samples after ranking. In subsequent step, various classification approaches has been used for classifying customers into non-churn and churn customers. Classification algorithm considered here is adaptive Logit boosting that is extracted from boosting approaches. This step classifies the instance more appropriately. Generation of retention policies is a significant process ofrelationship management to eliminate churners. After performing classification approach, these model segments customers' data by categorization using requirement similarity matrix to acquire information gain. This model validates churn factors that are needed for determining primary cause of churn. By recognizing essential churn factors from customer data, CRM may enhance productivity by recommending appropriate promotions to set of customers based on association among behavioural patterns. Clustering ranking and ALB is dependent on patterns of customer transactional characteristics from data. Finally, this approach recommends retention strategy for all categories of churn customers.

3.1 Dataset

Here, data were collected from online available Indian Telecom industries for validation. Tables related to this are provided in section IV. Circle/state/UT-wise number of telephones in India from 2010 to 2016 is considered here. This data includes references from all over India and state wise analysis from 2010 to 2016 as in Table I and II. This includes number of telephone (both wireless and wired) in India which is released by national data sharing and accessibility policy.

3.2 Noise Subtraction

Here, data subtraction for eliminating noise is considered to be more significant, so as to project data to be more useful for indeed of degradation prediction accuracy. Degradation in prediction accuracy leads to poor outcomes. Consider telecom datasets, which comprises of more imbalanced attributes, missing values and incorrect null values. This work analyzes dataset for filtering and reducing number of features. Therefore, it holds only useful features. Some unnecessary features are removed from the dataset, which is depicted in Table I.

3.3 Feature extraction

Next is feature selection, this step is considered to be more crucial from dataset based on domain knowledge. There are various approaches that are used for feature selection based on background study in context to churn predictions. Here, weighted correlation attributes and information gain based ranking techniques for selecting features with MATLAB environment. With churn dataset, some features are chosen by eliminating other features by ranking those values. The higher dimensional dataset shows improved attribute performance measure and essential decision making procedure. Some features shows less significantfactors. Also, classification performance is also improved only in case when the dataset hold higher predictive values and resourceful variables. Henceforth, concentrating towards feature selection and diminishment of number of irrelevant attributes improves classification performance. Here, for feature selection, two machine learning approaches are utilized for selecting attributes and some essential features. Entropy based IG and weighted oversampling attribute ranking approaches are used for choosing subset of appropriate features. From these approaches, ranking of these significant subsets is chosen with lower computational cost and eliminating dimensionality crisis. Ranking these attributes are used for recognizing factors and hiding factors are measured as some significant cause of churning.

3.4 Entropy Based IG

The ultimate purpose for choosing the significant and essential features is to provide alerts patterns and enhance accuracy of correlation among features. Ranking uses Entropy based ranking model that utilizes filtering, this process attempts to rank subset of features dependent on entropy in ascending order. Extracting essential features has to analyze attributes and to establish relationship between features. Henceforth, the ultimate objective is to haul out some additional features with superior discriminative capability. This model is utilized to score variable and thresholding is utilized to eliminate variables that lies below the corresponding limits. It states that, if features are relevant to one another that may be independent to input data and not independent of class labels. This depicts that features that shows no influence towards class labels can be eliminated or removed. Therefore, it is concluded that class labels play essential role in ranking, i.e. the property that ranks features has some influence towards class labels. IG is related to entropy, that is, disordering degree of dataset is computed with entropy model. Subset computation for entropy is a primary computation to evaluate IG. Computation of IG includes computation of class label entropy for all data subset and subtracts conditional entropies of all possible feature values. All samples are chosen with certain feature value. Then, number of consecutive occurrence of all class in these instances is evaluated, and finally entropy is measured. This process is repeated for all possible feature value. Entropy of subset is evaluated more properly for generating count matrix that deals with class membership of training instances of feature value.

Algorithm 1: Entropy Base Ranking

1. Initialize $S = 0, L_C \rightarrow class\ label, E \rightarrow domain\ attribute\ value$;
2. For all $L_C \in C$ do
3. Compute $P(L_{Ci})$;
4. Entropy H_{L_C} log2 $((L_{Ci}))$;
5. Compute $S \rightarrow H_{L_C}$
6. End for
7. For all $\in E$;
8. Compute $P(L_{Ci})$;
9. Add = $S + P(e(i) * \log_2 (P(e(i)))$;
10. Compute $S \rightarrow Sum$;
11. End for
12. For all C_i do
13. For all e_j do
14. Compute feature ranking with H
15. Evaluate $FR = S + P(c[i]|e[i]) * \log_2 P(c[i]|e[j]))$;
16. Compute $S \rightarrow FR$
17. End for
18. End for
19. Entropy threshold computation
20. $(C_i|e_i) = (-1) * \quad * (-1)R$;
21. Entropy based IG = $H_C - \quad (C_i|e_i)$
22. Return value of Entropy base IG
23. Return Entropy based IG
24. End

3.5. Weighted Sampling For Imbalanced Feature

When handling huge amount of telecom data, it is essential to deal with random samples for predicting churn. Randomly sampled features should be of smaller in weight (values/size), therefore it is termed as weighted sampling approach. Every instance possess equal amount of chance to be included in dataset. This may or may not be included into the dataset. Replacing samples based on weight to provide some rejection of features. Gaining information is considered as a ratio between numbers of samples with least weight by number of instances with higher weight as in Eq. (1):

$$Weighted\ sample\ information = \frac{No\ of\ samples\ with\ minority\ weight}{No\ of\ samples\ with\ majority\ weight} \quad (1)$$

This sampling approach is used to balance data from that of imbalanced dataset. This work considers weighted oversampling. that is, this technique improves frequency of majority weights of classes. Overfitting caused due to these minority weighted classes in dataset can be eliminated. If training set size is increased then time needed to construct classifier is also higher. Weighting these oversamples may avoid some errors during training set distribution, thereby concentrating on misclassification costs. From this, weighted co-efficient computation can be performed. Thus, Pearsons' correlation co-efficient is used to compute linear correlation among two variables where values relies among [-1, 1], here, -1 specifies negative relation between sample, 1 specifies positive relation among weighted samples. Finally, 0 defines no relation among these two variables. Pearson's correlation co-efficient is provided as in Eq. (2):

$$(i) = \frac{(i, Y)}{\sqrt{var\ (f_i) var(Y)}} \quad (2)$$

Where cov is covariance and var is variance of input features and output feature Y. This technique helps todetermine relationship among every variables andconcept oriented variables. It chooses only certain variables that possess negative and positive feature relation. Flow diagram of proposed model is given in Fig 1.

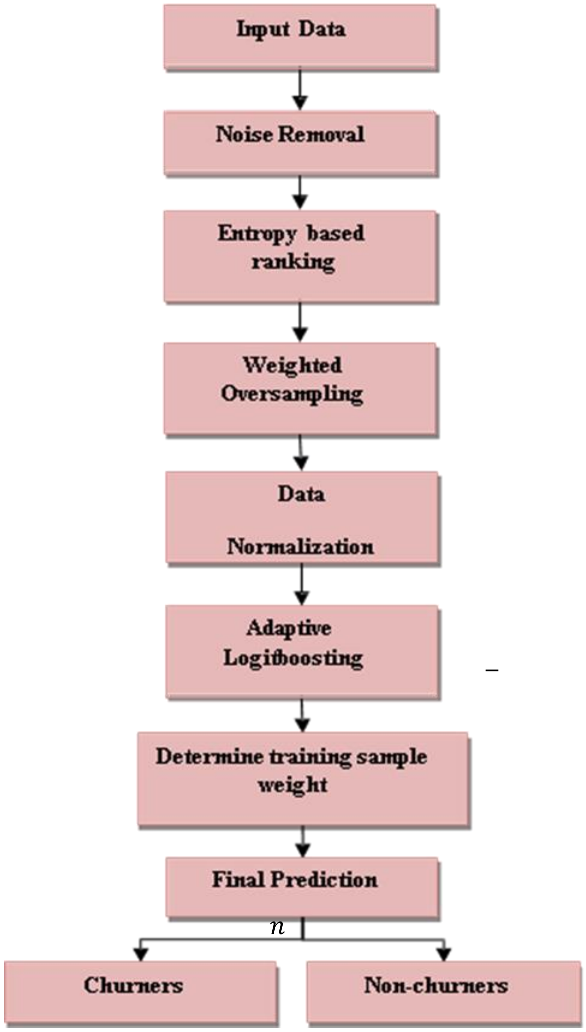

Fig 1:*Flow Diagram of ALB Approach*

Next is IG attribute computation, here information among two features are measured with entropy to handle relevance among target and attributes. This projects inconsistency priority. It measures mutual information in accordance to number of times, features considers classes, number of times feature does not considers classes and class that does not defines feature value. Information entropy is computed with Eq. (3):

$$IG(f_i) = H(T) - \sum \frac{T_i}{T} (T_i) \qquad (3)$$

Where $H(T)$entropy of is given dataset, $H(T_i)$ is entropy measure after splitting data with f_i and provided as in Eq. (4):

$$(T) = \sum_c P(c) \log(P(c)) \qquad (4)$$

3.6 Classification with Adaptive LogitBoost

There are two kinds of customers seen in telecom industrial dataset. Initially, non-churn customers, they seem to be more loyal to corresponding company, whereas seems to be more affected by other competitor concerns. Next kind of customers is churn customers. The ultimate target of this company is to recognize churn customers and to determine the cause behind their migration. Moreover, it determines and uses retention strategies to get rid of switching problems to other companies. Here, machine learning range is utilized for categorizing customer's data with labelled datasets. Here, an adaptive approach is used to find its superiority to classify customers into non-churn and churn models. Initially, boosting approach is considered for classification. It is measured as eager learning model where training data is provided to categorize newer samples. It is extensively utilized in background model in literature for data analysis and adaptive version of Logit boost model. Adaptive Logit boosting model is chosen here for building this prediction accuracy model more efficiently. This comes under the type of boosting algorithm where value of weak classifier is constrained to {−1, +1} and significant process of this method is given in Algorithm 2:

Algorithm 2: Adaptive LogitBoosting

1. Input $(a_1, b_1), (a_2, b_2) \dots (a_n, b_n)$
2. Initialize: $= \{W(i) = \frac{1}{n}; i = 1, \dots, n\}$
3. For $t = 1, 2, \dots, T$
4. Attain weak learner: $h_t; X \rightarrow [-1, +1]$
5. Compute classification error with Eq. given below:
$$E_t = \frac{1}{}\sum_i W_t(i) I[h_t(a_i) \neq b\ ]$$
6. Compute classifier weight with Eq. given below:
$$_t = \frac{1}{2} \ln [(1 - \ \ )]_E$$
7. Modify training sample weight with Eq below:
$$W_{t+1}(i) = \frac{W_t(i) e^{-a_t b_i h_i(a_i)}}{Z_t}$$
8. Classifieroutput: $= s\ (\sum^T a_t h_t(x))$

From Algorithm 2, training data is given as $(a_1, b_1), (a_2, b_2) \dots (a_n, b_n)$ where $A_i \in R^d$, $B_i \in \{-1, +1\}$, $W_t(i)$ is weighted distribution of samples at 't' iterative times .Moreover zt is a normalization factorthat guarantees that samples are provided under adistribution form. The finest benefits this adaptive Logitboosting is theoretical foundation with higher prediction accuracy and easier way of implementation. Logitboosting is an updated version of adaboosting model. Adaptive logitboosting uses exponential loss function while adaboost uses exponential loss function. The baseline idea of logitboosting is derived from adaboosting model. When it comes in churn prediction, here test indexes are available for both non-churn customers and churn customers. Here, detection results are chosen with test indexes with certain specific feature values. If customers do not come under churning, then it is test result is fixed as zero. The prediction results for these two category of classification model is given as churn and non-churn customers. While performing experimentation, various kinds of test items were considered as feature models, and churn classification were determined by adaptive logitboosting algorithm. In this experimentation, some common kinds of data mining algorithms were used to categorize churn customers. Some features of dataset have been chosen with ranking and sampling model by eliminating noise effectually. The boosting algorithm is in co-operated with churn prediction to compare its classification efficacy. From experimental validation, it is known that adaptive logitboosting shows better classification accuracy than compared to other data mining models. This ALB model depicts superior modelling time than compared to that of other approaches.

4.NUMERICAL RESULTS

Here, simulation is carried out in MATLAB environment as it is a user friendly tool. This tool establishes proper validation of classification in terms of accuracy, sensitivity, specificity, F-measure and so on. This software runs on Intel core i3 processor with certain configuration. Adaptive Logitboosting performs logistic regression. This is similar to that of additive model. This approach reduces logistic loss function indeed of exponential loss. The performance of noise removal, sampling and ranking along with Machine learning model is simulated with initial screening process. Performance has been evaluated based on prediction accuracy with 10 fold cross validation. Parameter tuning was performed with this machine learning model

Circles/States/UTs	Total Telephones - 2010	Total Telephones - 2011	Total Telephones - 2012	Total Telephones - 2013	Total Telephones - 2014	Total Telephones - 2015	Total Telephones - 2016
Andaman &	0.15	NA	NA	NA	NA	NA	NA
Andhra Pradesh	48.08	63.05	69.19	66.6	69.19	73.82	76.39
Assam	9.07	11.93	14.44	14.58	15.46	17.32	18.72
Bihar	36.63	54.74	64.09	60.7	61.97	69.67	74.84
Chhatisgarh	1.38	NA	NA	NA	NA	NA	NA

TABLEI: *DATASETPARAMETERS USED FOR 10 FOLD CROSS VALIDATION*

Jammu & Kashmir	5.78	5.97	6.51	7.04	8.13	9.46	9.95
Jharkhand	1.72	NA	NA	NA	NA	NA	NA
Karnataka	39.91	52.19	58.41	55.36	56.64	60.33	63.6
Kerala	27.65	34.66	37.21	33.76	34.01	33.94	36.61
Madhya Pradesh	32.17	47.21	52.76	53.28	56.58	61.71	66.72
Maharashtra	46.53	64.57	73.12	70.87	74.9	79.06	86.09
North-East - I	4.94	7.45	8.77	9.15	9.55	10.63	11.2
North-East - II	0.7	NA	NA	NA	NA	NA	NA
Orissa	15.88	22.99	27.08	24.98	25.47	28.19	29.37
Punjab	21.7	30.34	33.4	30.78	32.43	31.75	32.81
Rajasthan	35.27	44.39	50.32	49.61	53.57	56.03	60.88
Tamil Nadu	44.45	58.71	80.87	75.52	78.09	83.08	83.99
Uttarakhand	1.36	NA	NA	NA	NA	NA	NA
Uttar Pradesh - [E]	45.53	65.15	76.35	74.87	77.78	83.91	92.74
Uttar Pradesh - [W]	30.61	46.62	54.43	49.17	49.3	52.5	58.03
West Bengal	25.93	40.42	46.95	41.71	42.8	47.51	49.87
Kolkatta	17.86	24.61	26.17	22.4	22.15	23.56	25.93
Chennai	12.82	14.38	NA	NA	NA	NA	NA
Delhi	31.01	41.66	45.4	43.39	45.69	49.33	50.42
Mumbai	29.43	37.79	39.29	33.35	33.95	33.73	34.84

TABLE II: TOTAL TELEPHONE USERS FROM 2010-2016

Here, churn prediction is analyzed with accuracy, recall, precision, F-measure, ROC. Prediction accuracy is depicted with number of instances that are appropriately classified as in Eq. (5):

$$Acc = \frac{TP+TN}{TP-TN+FP+FN} \quad (5)$$

Where $'TN'$ is true negative, $'TP'$ is true positive,$'FN'$ is False negative and $'FP'$ is false positive. TPrate is determined as sensitivity as data is appropriately classified as positive. TP rate should be higher for anticipated ALB algorithm. TP is computed with Eq. (6):

$$TP\ rate = \frac{TP}{Actual\ positivevalues} \quad (6)$$

FP rate is depicted as inappropriately classified aspositive. FP rate has to be lower for any classifier and shown in Eq. (7):

$$Fp\ rate = \frac{FP}{Actual\ Negative\ values} \quad (7)$$

Accuracy is computed with positive prediction value (PPV). It is shown in Eq. (8)

$$Precision = \frac{TP}{(TP+FP)} \quad (8)$$

Recall is considered as a measure for completeness, that is, proper hit of algorithm. It is measured with a factor that all appropriate instances are chosen with this system. If recall is lower, then it specifies FN values. It is evaluated with Eq. (9)

$$Recall = \frac{TP}{TP+FN} \quad (9)$$

F-measure is measured as a trade off among properly classified data points to fulfil that every class comprises points of only class. It is depicted in Eq. (10):

$$F-measure = 2 * (\frac{TP}{TP+FN}) \quad (10)$$

If values are 0.5, 0.6, 0.7,0.8,and 0.9, then it is specified as random prediction model (bad, moderate, superior and good). Values that lies in other region of ROC shows some prediction is wrong.

4.1 Discussions

Various classification approaches were applied to churn dataset with MATLAB tools. The performance like properly classifies and improperly classified are considered here. Table III depicts performance metrics of proposed model with cross- validation. It is known that J48 and RF perform well when compared to NB and Adaboost model with 88.64%. When considering ensemble algorithms like RF, J48, NB and Adaboost, accuracy is considered with minimum incorrect classification. RF consumes more time to build prediction model, however, it shows higher classification accuracy compared to other models as in Fig 2 and Fig 3. J48 provides better performance in building model; however lacks in accuracy compared to RF. Computational cost of NB is lower, but prediction accuracy is extremely lower. Adaboost shows lower model building time and lower accuracy compared to other approaches. Finally, it is analyzed that RF provided better accuracy and model building time to compete with ALB model. ALB performs pre- processing like noise removal, ranking, sampling and normalization to offer higher prediction accuracy. ALB has 6.43 values for model construction and 92.65% properly classified instances. To carry out further validation process, performance of various ensemble algorithms is given in Table IV depicts TP an FP is higher for ALB classifier compared to other models. ROC is a selective performance model which is utilized by various investigators to project prediction model for evaluating accuracy. ROC is 0.95 for ALB where other ensemble values are lesser than this value as in Fig 4. F-measure and recall value of ALB is higher than other ensemble approaches that are 0.91 respectively. Churn customer factors are classified with Attribute selected classifier. This approach offers churning user patterns and behaviours. These rules are considered to be extremely valuable for decision makers for churning customer retention. This attribute provides multiple causes behind churn and validates features that depend of one another. Customer retentions is considered to be a positive factor for enhancing company profile that improves marketing performance and retention. With this retention strategy, customer behaviour is identified and provides finest packages to certain group of people. This investigation considered the characteristics of customers and classifies them into groups with use of machine learning approaches and concentrates on CRM to make higher customer retention. Fig 5 and Fig 6 depicts TP and FP rate attained with ALB and existing ensemble approaches. Fig 7, Fig 8, Fig 9 and Fig 10 depicts accuracy, Recall, F- measure and ROC computation of anticipated ALB. Similarly, Table III, Table IV and Table IV specifies performance metrics computation of proposed model

***TABLE III:** PREDICATION CLASSIFICATION COMPUTATION*

Methods	Incorrectly classified Instances (%)	Correctly classified instances (%)	Execution Time
Random Forest	11.36	88.64	108
J48	11.40	88	7
AdaBoost	16	82.90	9.23
Naive bayes	17.95	81.04	215
ALB	65.56	92.65	6.43

***TABLE IV:** TP AND FP RATE COMPUTATION*

Methods	TP rate	FP rate
Random Forest	0.88	0.22
J48	0.87	0.23
AdaBoost	0.82	0.32
Naive bayes	0.46	0.27
ALB	0.92	0.18

***TABLE IV:** PERFORMANCE METRICS COMPUTATION*

Methods	**Precision**	**Recall**	**F-measure**	**ROC area**
Random Forest	0.88	0.87	0.87	0.94
J48	0.87	0.87	0.87	0.94
AdaBoost	0.82	0.84	0.81	0.77
Naive bayes	0.70	0.46	0.44	0.60
ALB	0.92	0.91	0.91	0.95

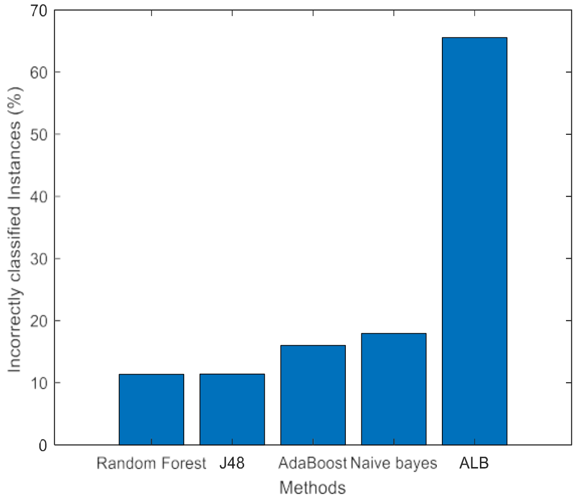

FIG 2: *INCORRECTLY CLASSIFIED INSTANCES COMPUTATION*

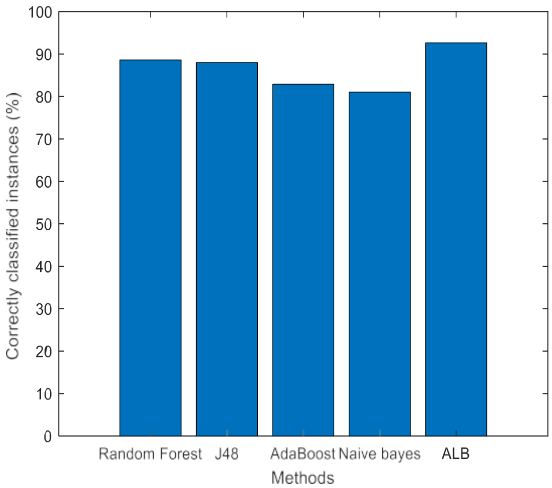

FIG 3: *CORRECTLY CLASSIFIED INSTANCES COMPUTATION*

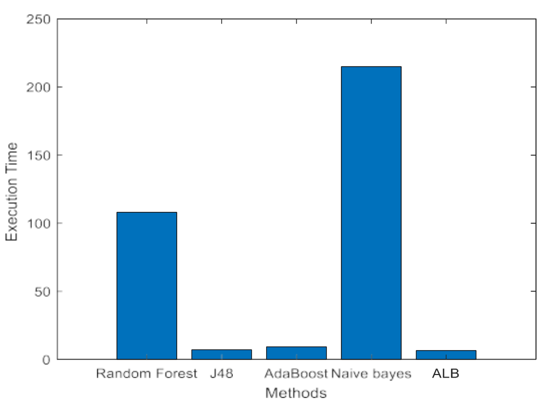

FIG 4: *EXECUTION TIME FOR BUILDING PREDICTION ACCURACY*

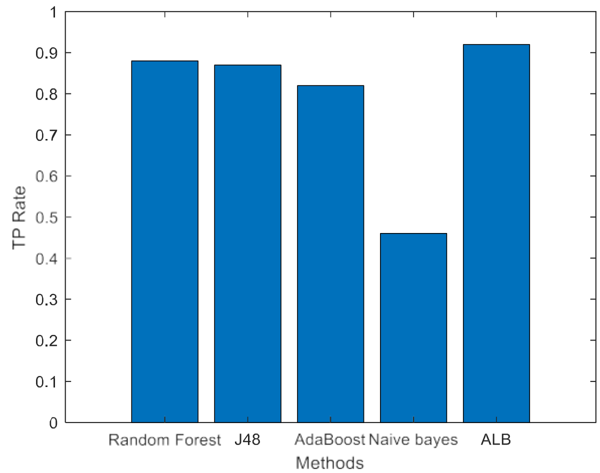

FIG 5: *TP RATE COMPUTATION*

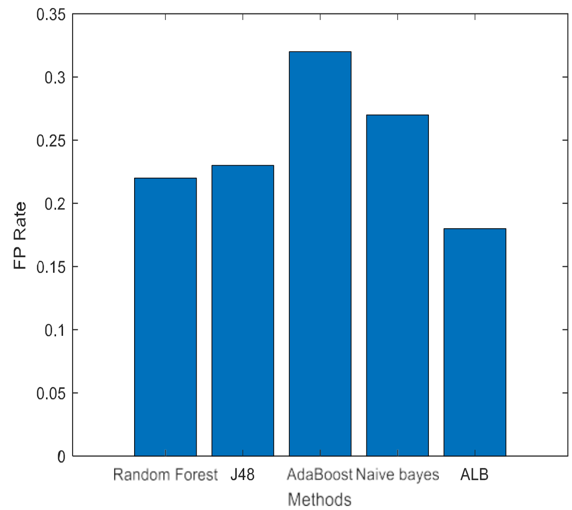

FIG 6: *FP RATE COMPUTATION*

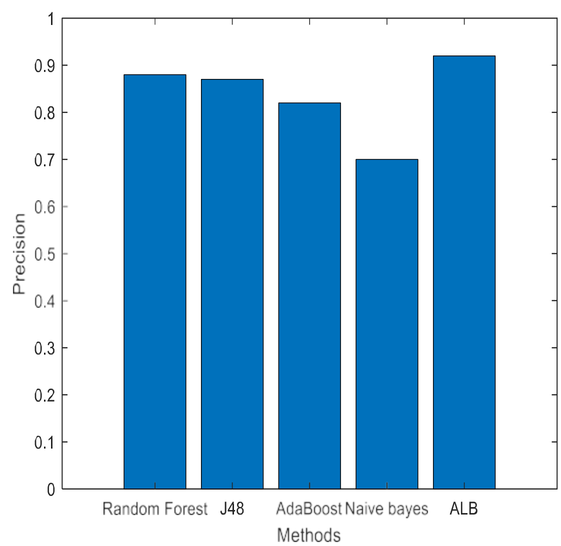

FIG 7: *PRECISION COMPUTATION*

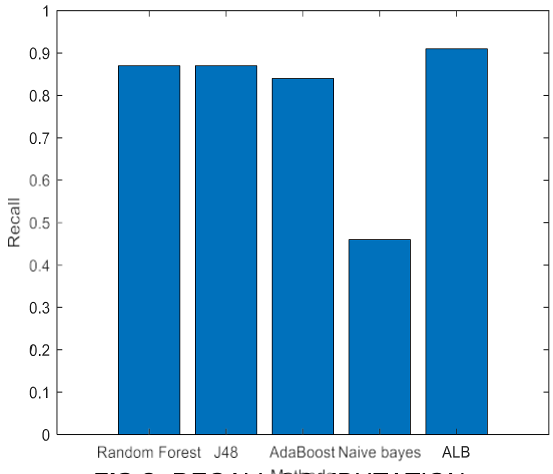

FIG 8: *RECALL COMPUTATION*

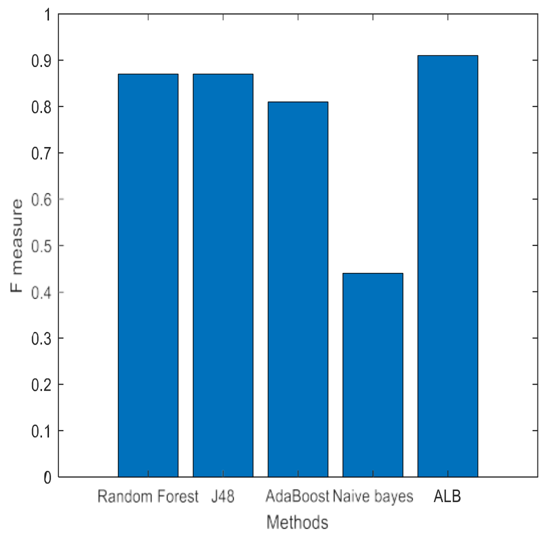

FIG 9: *F-MEASURE COMPUTATION*

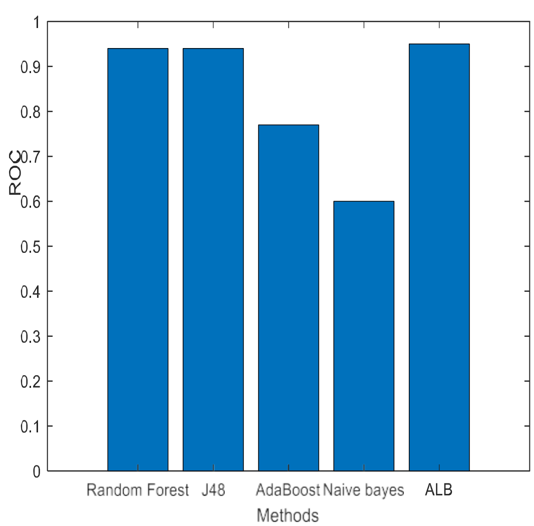

FIG 10: *ROC COMPUTATION*

5.CONCLUSION

In this competitive world of telecom industries, churn prediction is measured to be a significant cause of CRM to retain more valuable customers by recognizing similar group of customers. Here, Indian churn dataset of leading telecom industries are considered for validation. To retain some customers, more valuable packages are provided to those people to maintain a global relationship among them. Therefore, exports of telecom industries works together to retain customers by prediction the behaviour of customers and to generate better retention accuracy prediction. For this cause, this work uses some essential pre-processing steps like noise removal in dataset, ranking, sampling, normalization and classification. Here, various performance metrics has been validated to provide better significance among customers. Some ensemble algorithms are considered for measuring these performance metrics. From this, the anticipated ALB shows better trade off in contrary to other data mining approaches. ALB provides lesser time for building the model and higher prediction accuracy with least error rate. This may provide a baseline for telecom industries for offering better guideline for retaining customers. In future, this work may be extended as hybrid modelling of Machine learning approach with ANN to validate performance and improve it by another 5%. Optimization approach can also be integrated with them to acquire global optimum results.

REFERENCES

[1] Runge, P. Gao, F. Garcin, and B. Faltings, “Churn prediction for highvalue players in casual social games,” in Computational Intelligence and Games (CIG), 2014 IEEE Conference on. IEEE, 2014, pp. 1–8.

[2] Milo˘sevi´c, N. ˘ Zivi´c, and I. Andjelkovi´c, “Early churn prediction with personalized targeting in mobile social games,” Expert Systems with Applications, vol. 83, pp. 326–332, 2017.

[3] Tamassia, W. Raffe, R. Sifa, A. Drachen, F. Zambetta, and M. Hitchens, “Predicting player churn in destiny: A hidden markov models approach to predicting player departure in a major online game,” in Computational Intelligence and Games (CIG), 2016 IEEE Conference on. IEEE, 2016,

[4] Viljanen, A. Airola, T. Pahikkala, and J. Heikkonen, “Modelling user retention in mobile games,” in Computational Intelligence and Games(CIG), 2016 IEEE Conference on. IEEE, 2016

[5] Khan, I., Tariq Usman, G.U.R.: Intelligent Churn prediction for Telecommunication Industry. Int. J. Innov. Appl. Stud. 4, 165–170 (2013)

[6] Kirui, C., Hong, L., Cheruiyot, W., Kirui, H.: Predicting Customer Churn in Mobile Telephony Industry Using Probabilistic Classifiers in Data Mining. IJCSI Int. J. Comput. Sci. Issues 10, 165–172 (2013)

[7] Huang, B., Kechadi, M.T., Buckley, B.: Customer churn prediction in telecommunications. Expert Syst. Appl. 39, 1414– 1425 (2012)
[8] Verbeke, W., Martens, D., Baesens, B.: Social network analysis for customer churn prediction. Appl. Soft Comput. 14, 431–446 (2014)
[9] Farquad, M.A.H., Ravi, V., Raju, S.B.: Churn prediction using comprehensible supportvector machine: An analytical CRM application. Appl. Soft Comput. 19, 31–40 (2014)
[10] Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, "A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 42, no. 4,pp. 463–484, 2012.
[11] Ertekin, "Adaptive oversampling for imbalanced data classification," in Information Sciences and Systems 2013, Information Sciences and Systems 2013. Springer, 2013
[12] Verbeke, K. Dejaeger, D. Martens, J. Hur, and B. Baesens, "Newinsights into churn prediction in the telecommunication sector: A profit driven data mining approach," European Journal of Operational Research,vol. 218, no. 1, pp. 211–229, 2012.
[13] Long, W. Yin, L. An, H. Ni, L. Huang, Q. Luo, and Y. Chen, "Churnanalysis of online social network users using data mining techniques," in Proceedings of the International MultiConference of Engineers and Computer Scientists, vol. 1. Proceedings of the International Multi-Conference of Engineers and Computer Scientists, 2012.
[14] Prasad and S. Madhavi, "Prediction of churn behavior of bank customers using data mining tools," Business Intelligence Journal, vol. 5, no. 1, pp. 96–101, 2012.
[15] Farquad, V. Ravi, and S. B. Raju, "Churn prediction using comprehensible support vector machine: An analytical crm application," Applied Soft Computing, vol. 19, pp. 31–40, 2014.
[16] Saradhi and G. K. Palshikar, "Employee churn prediction," Expert Systems with Applications, vol. 38, no. 3, pp. 1999–2006, 2011
[17] Hawalah and M. Fasli, "Dynamic user profiles for web personalisation," Expert Systems with Applications, vol. 42, no. 5, pp. 2547–2569, 2015.
[18] Zhang, Y., Liang, R., Li, Y., Zheng, Y., & Berry, M. (2011, July). Behavior-based telecommunication churn prediction with neural network approach. In Computer Science and Society (ISCCS), 2011 International Symposium on (pp. 307-310). IEEE. July 2011
[19] Huang, Y., &Kechadi, T. (2013). An effective hybrid learning system for telecommunication churn prediction. Expert Systems with Applications, 40(14), 5635-5647.2013
[20] Dahiya, K., & Bhatia, S. (2015, September). Customer churn analysis in telecom industry. In Reliability, Infocom Technologies and Optimization (ICRITO)(Trends and Future Directions), 2015 4th International Conference on (pp. 1-6). IEEE. September 2015
[21] Qureshi, A. S. Rehman, A. M. Qamar, A. Kamal, and A. Rehman, ``Telecommunication subscribers' churn prediction model using machine learning,'' in Proc. 8th Int. Conf. Digit. Inf. Manage., Sep. 2013, pp. 131_136.
[22] Babu, D. N. Ananthanarayanan, and V. Ramesh, ``A survey on factors impacting churn in telecommunication using datamininig techniques,'' Int. J. Eng. Res. Technol., vol. 3, no. 3, pp. 1745_1748, Mar. 2014.
[23] Idris and A. Khan, ``Customer churn prediction for telecommunication: Employing various various features selection techniques and tree based ensemble classi_ers,'' in Proc. 15th Int. Multitopic Conf., Dec. 2012, pp. 23_27.
[24] Zhao, Q. Gao, X. Dong, A. Dong, and X. Dong, ``K-local maximum margin feature extraction algorithm for churn prediction in telecom,'' Cluster Comput., vol. 20, no. 2, pp. 1401_1409, Jun. 2017.
[25] J. Togelius, "How to run a successful game-based AI competition," IEEE Transactions on Computational Intelligence and AI in Games, vol. 8, no. 1, pp. 95-100, 2016