Customer Churn Prediction and Upselling using MRF (Modified Random Forest) technique

Swetha P, Dayananda R B

Abstract: ***Customer Churn Prediction has become one of the eminent topic in the telecom industry, it has gained a lot of attention in the research industry due to fierce competition from the various, and hence companies have focused on the larger size of the data for churning and upselling prediction. The model of customer churn prediction detects and identify the customer who are willing to terminate the subscription, customer churn prediction and upselling can be done through the data mining process. Hence, In this paper we have introduce a model Named MRF(Modified Random Forest), this model helps in enhancing the accuracy and also helps in ignoring the regression issue. Our methodology has been performed on the provided orange Datasets. For the evaluation of our algorithm comparative analysis between the existing and proposed methodology is done considering the two scenario i.e. churn and upselling. Later our model is compared with the various existing churn prediction model, the result of the analysis indicates that our model outperforms the existing method including the standard random forest in terms of AUC and classification accuracy.***

Keywords: Churn, prediction, upselling, MRF, customer churn.

I. INTRODUCTION

Customer churn is a term, which is used for denoting the customer movement from oneservice provider to the other. There might be several reason for customer to churn, to mention few of them are comparatively high service cost, quality of the service etc[1]. In the real scenario, gaining the new customer is 5 to 6 times higher than the keeping existing customer. Hence, this scenario creates the pressure on the company to minimize the customer churn. Customer Churn is one of the essential aspects and has gained lot of interest from various industries such as telecommunication operators[2]. Moreover, customers are considered to be the source material for the telecommunication companies, hence it is very essential to develop technique to detect the customer churn for survival as well as development of telecom companies.In the last few years, there has been a huge generation of data; however having the large amount of data makes it more essential to extract the information that is hidden over the given raw data. From this large amountof data the meaningful data can be extracted in order to help the industry grow. One of the method used to extract the information is known as data mining[3]. Data mining is nothing but the process of discovering patterns in given hugedatato identify the patterns as well as tries to establish the relationships to resolve the issue of data analytics. In the last two decades, the growth in data volume has been noticed due to day-by-day improvisation in the IT(Information Technology). In parallel deep research has taken place in data mining. Several methods as well as novel technique is added in order to gather the required information and process the data[4]. Data, which is collected from the various resources, is the raw data where the many information is hidden. In recent the main issue of the telecom industry has been the customer churn, the model of customer churn helps to analyze and detect the customers that has the high probability of exchanging the service provider. Hence, the given database who is trying to churn allows the organization to focus on the particular customer and develop the retention strategy; this in terms reduces the customer churning percentage. Moreover retaining the old customer has been always preferable option for the company since the analytics of the new customer shows that the getting new clients is way more expensive this includes the attractive offer, marketing and several discounts. A marginal step towards retaining the existing customer might lead to marginal increase in the profits as well as revenues. In recent days, it is observed that the Telecommunication companies has been losing the valuable customers these results in losing the revenues. Over the last few years, the telecom industry has gone through the vast changes in terms of adding novel services and technological advancements. The above discussed scenarios in customer prediction churn makes very essential to protect the loyal customer base, growth of an organization and improvisation in CRM[Customer Relationship Management [5]. Retention of the customer whose churn is very much high is one of the challenging task in telecommunication industry. Nowadays, due to the many number of service providers and intense competition customers have number of alternatives to churn.There are various factors which helps in influencing the customer to churn. Prepaid customers are not bound to any service contracts like post-paid customers, hence it is very difficult to predict the churn rate. Other issues that includes reception quality, network issue and product quality also causes to migrate from one service provider to other service providers. An existing telecommunication might just work fine, if it can take care of its own customer rather than grabbing the others. Through the analysis, it is found that globally the churn rate is approximately 2%, which makes a whopping loss of 1oo billion dollar. Moreover the advancement in technologies has helped companies to analyze the competitive strategies, this ensure the increase in customer

Revised Manuscript Received on January 05, 2020

Swetha P, Asst. Professor, Department of CSE, Rajarajeswari College of Engineering Bangalore.

Dayananda R B, Professor, Dept. of CSE, K.S Institute of Technology, Bangalore,

retention rate in order to survive in the industry. Hence, a marginal amount of research has been done in this area. Due to the competition and the current scenario situation has become worse due to competition. Moreover,improvisation of 1 percent in CRR(Customer Retention Rate) could easily increase the share price by 5% [6]. Hence, in this research work we have proposed a model namely MRF(Modified Random forest) , our method provides several layer for improvising the accuracy and ignores the overfitting. This model helps in discarding the regression issue, regression problem is one of mediocre problem that arises while churn and upsell prediction. Our models provides the minimum error rate and better classification rate than the other state of art method including the existing method (Modified Random Forest).

This particular research work is organized as follows such that first section presents the background of customer churn, second section discuss the various existing methodology for customer churn prediction , it is parted based on the various technology. Section 3 presents the proposed work along with the pictorial representation of the same and algorithm and mathematical notation. Fifth section of this paper shows the evaluation of work by showing the comparative analysis along with graph and tabular representation. Last but not the least section presents the conclusion of our work.

II. LITERATURE SURVEY

In the past, there has been enormous research in the customer churn prediction, in this section, we have discussed various existing methodology. This survey is parted based on the various method such as Neural Network, Decision trees and covering algorithm, these brief study about the existing work has led us to design our model they are:

1.1 Neural Networks

Neural networks is nothing but one of the data mining technique that is capable of learning from the given errors, they are mainly motivated by the brain. As brain learns from the various new things later it will be transmitted via the neurons. Similarly, the neural network neuron along with the learning algorithm is capable of learning from the trained data; this in terms makes it to be referred as the ANN(Artificial Neural Network)[7]. The outcome of Capota and lazarov shows that ANNs gives the better performance when compared to the other algorithm. Moreover, Au et al.[8] concluded in his research that the limitation of the neural networks is that they arent able to uncover the patterns in the flexible manner, in that research it was also observed that the NN(Neural Networks) performs better than the decision trees. To contradict that MoZer et al.[9] presented the NNN(Non-linear Neural Network) outperforms the logistic regression as well as decision trees. Moreover, proposed the NN-based methodology for the prediction f customer churn. Here, the churn dataset of the UCI repository clearly shows that the NN based approach performs better than the other.Statistical based methods are the collection of several methods that are applied in the data mining and which is used for processing the huge volume data,basically they are used as the link between the independent and dependent attributes. Scheme based on the regression provides the efficient outcomes in predicting and estimating the churn rate. Several methods of regression tree were considered along with the method like decision trees , neural networks and rule based learning. It was concluded that the ideal prediction model needs to be developed constantly. [7] Used the logistic regression technique on the telecom industry for identifying the churners. However they failed drastically since the number of churners identified was on the lower side. However, the logistic regression performed better than the above. Hence, to overcome with that problem implied the two algorithm i.e. logistic regression as well as decsion trees which helps in constructing the churn prediction model. [10]Proposed the prediction model which is mainly based on the naïve Bayes algorithm and this provided the better accuracy in the given first pass which was Bayesian model based.

K-nearest algorithm is known as the basic classification method, hence [11] presented the hybrid scheme of logistic regression and k-nearest algorithm in order to construct the classifier named KNN-R which is binary based. Here brief comparison is presented between the RBF (Radial Basis Function) network, C4.5 along with the KNN-LR with the logistic regression and it was concluded that the KNN-LR easily outperforms the RBF method based on the several benchmark datasets. However, it wasoutlasted by C4.5. In paper [12] a modified algorithm of K-Mean clustering is presented along with the classic rule of inductive method for predicting the behavior of Customer churn. Here comparison has been done based on the six methodologies that are oneR, KNN,SVM, PART, logistic regression, decision tree and original K-means, the method presented seems to be better than the above model.

1.2 Decision Trees

Decision trees are said to be one of the common methodology, which are used in predicting as well as estimating the customer churn issue. Based on the divide and conquer method, the decision are developed. However decision tree does have some boundaries such as they can not be used for the non-linear and complex relationships among the attributes. However, it has been observed that the decision tree method does try to improvise the classification accuracy. In paper[13] ANN along with the decision trees are used for performing the predictionof customer churn and it is noticed that decision trees does outperforms the neural networks when it comes to the accuracy. [14] Presented a methodology of application classification DT in order to analyze the churn rate in the telecommunication industry. Here, ID3 decision tree is used and it is observed that subscriber area was one of the main classification feature, the other given two results in the customer to churn.[7] used the several methods such as K-means clustering, QUEST, CART, Logistic Regression, neural network, exhaustive CHAId. In here, it is observed that CHAID performs much better than the other mentioned methodology.Itis noticed that the accuracy was approximately around 60%, which was way better than theother methodology. Moreover, the

other decision trees did not stand in front for the Exhaustive CHAID.

1.3 COVERING ALGORITHMS

Several covering algorithms such as RULES, RIPPER, CN2 and AQ, here the rules are induced directly from the given training sets. [1] Presented the application of the two given novel data mining schemes to predict the customer churn. Here, they integrated AntMiner+ and ALBA to in order to achieve the accurate rules for the classification.Hence, in here it is observed that in order to achieve the highest accuracy integration of C4.5 with ALBA is required. RULE Extraction System has been separated from the several covering algorithms due to its simplicity. RULES-1 were the first member of the RULES family[15]. Second version of the RULES has been developed and applied variousdomains [16]. [17] presented the RULE-3 i.e. Inductive Learning Algorithm is efficient for data mining, here they have used RULE-3 on the various real life data sets for the data mining. Several methodologies has been presented in the above literature survey, the survey of this research can be concluded that the methods based on the decision trees neural networks and algorithm does provide the flexibility for estimating the customer churn rate. However, several issue apart from the problem of churning such as class balancing problem, data segmentation problem that are associated with these methodologies and might affect drastically on the performance. Hence, in order to get rid of these problems we proposed the MRF(Modified Random Forest) algorithm for minimizing as less as possible, the MRF (Modified Random Forest) algorithm for estimation of customer churn rate in the telecom industry. The main motive behind this research is to estimate the customer churn rate; motivation is that it is vigorous towards overfitting for achieving the better accuracy. Our methodology is more flexible and the error rate is comparatively less than the other methods that are described in the literature survey of the same research.

III. PROPOSED METHODOLOGY

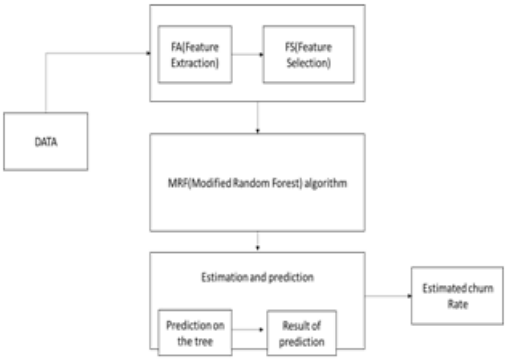

Figure 1: propopsed Modified Random Forest Technique

The above diagram in figure 1 represents the block diagram of proposed MRF (Modified Random Forest) ,Basically the flow of diagram shows such that it is parted into the five blocks, first blocks represents the data, in second block the operation takes place, i.e. feature extraction and feature selection. Once the extraction and selection is done, our algorithm is applied on the given data. Next block shows the estimation and prediction. In here, the first the tree prediction is done and sent to he result of prediction, at last the churn rate is estimated.

1.4 Modified Random Forest Preliminaries

In this section, the proposed Modified RF (Random Forest) technique is utilized. Several conventional algorithm such as Decision Tree, Genetic algorithms, neural networks and classification tree has been proposed. These abovementioned algorithm is capable of estimating the churn rate. However, they have the several issue such as Decision tree lacks with the issue of same class probability that can drastically reduce the performance. Similarly, in case of the genetic algorithm it is highly improbable to recognize the likelihood linked with the estimation that results in the low performance and in case of technique such as state-of-art produces the several error.Hence, the above discussion can be concluded that conventional RF technique do not produce the efficient outcome in case of the huge datasets and also it performs very bad when they are unbalance. Our proposed MRF(Modified Random Forest) method performs better in various parameter such as accuracy robustness and others. Our Scheme primarily has the extra layer of RV(Random Variable) that helps to perform the model much better along with that our method helps in minimizing the Gaussian noise and also helps in decreasing the problems of regression and classification. Our method helps in constructing the various different trees from the particular training datasets. Proposed method helps in focusing e estimationof consumer churn rate in the telecom services.

1.4.1 Tree prediction

At first, the predicted value is assigned for the given each terminal node. For example consider a terminal node TN_m contains the a_m training casesY_1,.....,Y_{am} along with Y_i =(A_i, B_i), later the prediction for particular model is weighted average of given response in the particular node, $B_1, \ldots, B_{a_m,}$

$$\widehat{B_m} = \frac{1}{\sum_{i=1}^{a_m} W_i} \sum_{i=1}^{a_m} Z_i . A_i$$

Here, $Z_1, i = 1, \ldots, a_m$ are the3 given positive numbers.

In the given regression issue,

$$Z_i = \frac{1}{a_m} \quad (2)$$

1.4.2 Classification problems

Here indicator function is denoted by I, this clearly indicates that the regression problem is nothing but the mathematical average of all the provided training cases that responses in $Node_k$, whereas the classification is defined as the receiving category by all the given training cases in the given node. Moreover, $CART$ is one of the way for

partitioning the space of regression to disjoint hyperactiverectangles. All the test cases that falls under the terminal nodeTN_m, and this is done using the B_m.

$$Z_i = \begin{cases} 1 \; if \; \sum_{i=1}^{a_m} (B_l = B_i) \geq \sum_{l=1}^{a_m} I(B_l = B_i) \forall B_i \neq B_j, \\ 0 \end{cases} \quad (3)$$

Pruning a Tree

Moreover, the largest performs drastically for the new test dataset prediction, since prediction variance gets maximized as the tree exceeds its optimal. Moreover, pruning is conducted in order to get rid of the problem of overfitting. For any tree T with terminal nodes $K, S(N)$ represents the overall training sample error. Let complexity parameter be denoted by α, this quantifies the penalty for maximizing the tree size and this can be achieved through splitting the single node,

$S_\alpha(N)$is computed using the penalty on the three complexity to the generated training error $S(N)$.

$$S_\alpha(N) = S(N) + \alpha M \quad (4)$$

1.5 Modified RF Algorithm

Let us consider any training set $Y = \{Y_1, \ldots\ldots, Y_n\}$ along with $Y_i = (A_i, B_i)$, any independent test case Y_0 with the predictor A_0.

$$\widehat{D_i} = \frac{1}{L_i} \sum_{\{b \,|\, i \in y_b^{oob}\}} \widehat{B_i^b} \quad (7)$$

Step1: In order to generate bootstrap resamples $D_1, \ldots\ldots, D_M$, training set C is sampled.

Step2: grow a regression tree N_m or classification

Step3: The predicted value by whole is obtained by integrating the results of the individual trees.

$$\begin{cases} \frac{1}{L} \sum_{l=1}^{L} \widehat{c_m^*}(A_0) & for\ regression\ problems \\ argmax_p \left\{ \sum_{m=1}^{L} K[\widehat{c_m^*}(A_0) = p] \right\} & for\ classification\ p \end{cases} \quad (5)$$

1.5.1 MRF(Modified Random Forest) based on test cases by weighted bootstrap

Once the Random forest is constructed, rather than creating the trees from given bootstrap samples which is based on the uniform sampling, we tend to imply the method of resampling which assigns the maximum probabilityweights only to the provided data cases in A_o close proximity. The standard version of Randdom Forest provides the proximity measure in order ot assign the distances between the observations. Proximity measure is defined such that the given any training datasets $Y = \{Y_1, \ldots\ldots, Y_N\}$ and the desired location of regressorA_0 , by resampling from R, B trees are grown. In case of $Y_i = (A_i, B_i)$, we consider E_t which denotes the number of trees that contains both A_0 and A_1 . proximity weight for any training cases Y_i relative to the A_o that can be described as in given below equation.

$$O_i = \frac{F_i}{\sum_{j=1}^{R} F_j}, i = 1, \ldots\ldots, R. \quad (6)$$

1.5.2 MRF Algorithm based on Test Cases

Step1: standard Random Forest, $STRF^{wf}$ (wf for "weight defining") is developed based on the described $Y = \{Y_1, \ldots\ldots, Y_N\}$, here every tree have the same amount of MTN (Maximal Terminal Node) size and represented as MTN^w.

Step2: Proximity weights on the given cases Y are defined as $(O_1, \ldots\ldots, O_N)$ from $STRF^{wf}$ and A_0 .

Step3: second $RF, SECRF^p$ is developed and here the resample of every bootstrap $BTR_j^* = \{Y_1^*, \ldots\ldots, Y_N^*\}$, this in terms provides the corresponding treesN_1^*,…..,$N_{B_p}^*$ which defines RF^p.Here the number of bootstrap resamples which defines the $SECRF^p$ is denoted by $SECRF^p$.

Step4: A_0 is submitted to $SECRF^P$ and the prediction is achieved as $\widehat{B}(A_0)$.

1.5.3 Variable Importance Based on Test Cases

Consider any training set Y= $\{Y_i, i = 1, \ldots\ldots, \ldots R\}$ of given Rcases $Y_i = (A_i, B_i)$.

Step1: based on Y a standard RF is developed and it consists of B trees which is generated from the bootstrap resample C and this contains the three basic scenario.

Scenario1: for any dthtree $d = 1$,…..,D and let B_b^{in} and B_b^{oob} be any subsetof B.

Scenario 2: predicting the out of bag training cases by using the dth tree in order to obtain $\widehat{D_i^b}$ for $i \in B_b^{oob}$.

Scenario 3: permuting the jth predictor variable among the various cases in Y_b^{oob} and repeating the same by u times.

Step2: In case if there are L_t trees developed without any cases then the bag prediction by the forest is:

$$\widehat{B_i^{u,j}} = \frac{1}{L_i} \sum_{\{b \,|\, i \in c_b^{oob}\}} \widehat{Y_i^{b,u,j}}, u = 1, \ldots\ldots, U \quad (8)$$

Step3: the prediction MSE(Mean Square Error) for the ith case is given below.

$$MSE_i^j = \frac{1}{U} \sum_{u=1}^{U} (B_i^{u,j} - B_i)^2 \quad (9)$$

Step4: variable importance in case of jth predictor is given as

$$X_j = \frac{1}{R}\sum_{i \in c} \Delta MSE_i^j \quad (10)$$

Later the variable importance values are normalized as given below equation.

Moreover, a modified version of variable importance(XI) scales the modified RF which is shown below.

$$XI_j = \frac{V_j}{\sum_k V_k} \quad (11)$$

$$TSXI_j = \sum O_i . \Delta MSE_i^j \quad (12)$$

Here D_i is the proximity measure between the X_0 along with the ith training in C, and it is given in below equation.

$$TSXI_j = \frac{TSXI_j}{\sum_k TSXI_k} \quad (13)$$

1.5.4 Prediction estimation based on the Test Cases

Here, for the training dataset C which is discussed earlier of the size n bootstrap resamples B_1,B_2,……, B_m. Here each datasets is achieved by resampling the n cases from C independently with the replacement. O_{-i}represents the set of indices that do not contain the ith training C_i. Prediction loss is given in the below equation.

$$\widehat{MSE} = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{||o_{-i}||} \sum_{m \in O_{-i}} L(B_i, \widehat{c_m^*}(a_i)) \quad (14)$$

Here, $\widehat{C_m^*}$ is nothing but the fitted rule fitted which is based on the bootstrap resample D_m and at last $\widehat{MSE}$ is calculated using the below equation, it tends to estimate the expected loss.

IV. SIMULATION AND RESULT ANALYSIS

In this section performance, our proposed MRF(Modified Random Forest) technique is evaluated. The current phase in the telecom industry shows that customer churn has become one of the big issue. Hence, it is essential to estimate the churn rate so that telecom industry can identify the user who wants to port from one company to the other and this in terms helps the company to retain the customer. In this research, we have developed and proposed the MRF(Modified Random Forest Technique) for estimating the customer churn rate. Proposed technique provides robustness towards the error rates as well as overfitting than the other existing state-of- art technique such as SPA(Sequence Pattern Analysis), BM(Bayesian Multi-net) Classifier.Proposed method MRF(Modified Random Forest) is tested on the orange dataset, knowledge recovery and data mining information is obtained through KDD Cup 2009, the datasets In our research the dataset, which is used, is taken from the orange, orange provides the huge dataset of the customer data, which consists of two types i.e. training and test. Here the three target variables are given namely churn, appetency and upselling. The two version of datasets were made publicly available, i.e. small and large small datasets are available with the 15,000 variables and 230 variables respectively. This 230 features are parted into the two category. First category shows the numerical value which contains the 190 variables, the remaining 40 variables comes under the category of categorical features. Our algorithm has been adopted for estimating the churn rate since it has better classification accuracy and high computation power. In order to give in real time scenario, the datasets of churn, upselling and appetency were available.Data pre-processing is said to be one of the essential steps in data mining, it involves the transformation of raw data into the normal format. The data from the real worlds are inconsistent and incomplete and has high chance of containing the several errors. In data processing, the raw data are prepared for the further processing. Our model is simulated on the windows platform with windows 10 packed with 16 GB Ram, the processor used is INTEL core I5- 4460 processor, the CPU has speed of 3.2GHz, code is deployed in R language and IDE used is Visual Studio 2017.

1.6 Comparative Analysis with various state- of-Art technique

Moreover, In order to evaluate our algorithm we have compared the proposed algorithm (MRF) with the state art of technique, which I depicted in Table1. The comparison is one based on the AUC (Area under Curve), when the table 1 is observed we see that various algorithm has various AUC, i.e. CSS LRM has AUC of 0.7354, whereas ANN possesses 0.7352 of the curve. Similarly, in case of standard RF model is good and it possesses 0.9929 of AUC. Moreover, our model possesses the 0.9959 AUC which is slightly better than the standard RF.

Table 1: Comparison of Customer Churn with State of Art Techniques

Algorithm	Area Under Curve
Area Under Curve	0.7354
CSS LRM	0.7354
ANN	0.7352
Standard RF	0.9929
Modified RF (proposed)	0.9959

Accuracy is one of the parameter in order to evaluate the classification model; accuracy is the number of correct prediction divided by total number of prediction. In here we see that our model i.e. MRF is compared with the various methodology in terms of Classification Accuracy. Different model gives the different classification accuracy, ALH, KLMM [18], Firefly, Hybrid Firefly; SVM has the classification accuracy of 74.8, 79, 86.36, 86.38 and 92.55 (in percentage) respectively. Similarly, K-means + Boosted C5.0, Ensemble of classifier (Boosting), K-means+ KNN [19] has the classification accuracy of 92.84, 94.17, and 94.72. The random forest technique itself has the classification accuracy of 97.59, however; our proposed model excels and possesses the

classification accuracy of 99.26. However, not much work has been done in upsell prediction in past.

Table 2: Classification Accuracy

Algorithm	Classification Accuracy (in percentage)
ALH	74.8
KLMM	79
Firefly	86.36
Hybrid Firefly	86.38
SVM	92.55
K-means + Boosted C5.0	92.84
Ensemble of classifier (Boosting)	94.17
K-means+ KNN	94.72
Standard RF	97.59
Modified RF	99.26

1.7 ROC based evaluation

In here, in order to prove the efficiency of our algorithm we have plotted the ROC graph for churn prediction and upsell prediction. Moreover in case of churn and upsell, different scenario has been considered and plotted which is discussed in the sub-section below.

1.7.1 Churn Prediction

The below graph (figure 2) shows the comparison of existing RF technique and MRF through ROC curve. Roc curve is one of the graphical plot which is generated by plotting the TPR (True Positive Rate) against FPR(False Positive Rate).Through the below graph we observe that TPR of MRF is slightly better than the existing RF

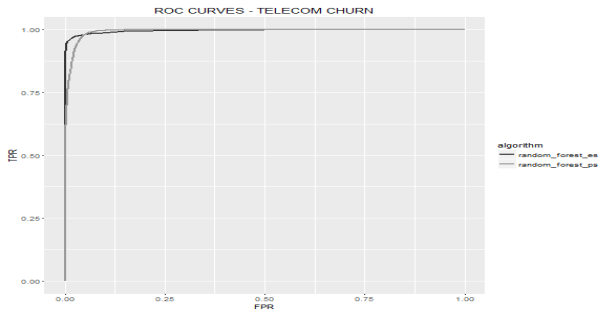

Figure 2 ROC Curve for Churn

1.7.1.1 Accuracy Vs Cutoff and Error Rate vs Cutoff

The below graph (Figure 3) shows the comparative analysis of Existing and proposed Random forest technique, in here the accuracy vs graph has been plotted to prove the efficiency of our algorithm. We observe that when compared to the Existing Random Forest technique, our model i.e. Modified Random Technique performs better. Generally, the classification models generates the probability, which lies between 0 and 1. Moreover, the cutoff determines the occurrence as event or non-event. Various cutoff shows the various performance measures(here in terms of accuracy). The below graph shows the accuracy vs cutoff, here we observe that as the accuracy also increases and from the below graph it is clear that our model possesses better accuracy than the existing one. In Figure 4 Thegraph illustrate the given trade-off among the error types as cutoff varies, we observe that in proposed methodology the error rate is reduced. Moreover, change in cutoff improvises the performance we see that our model has better error rate

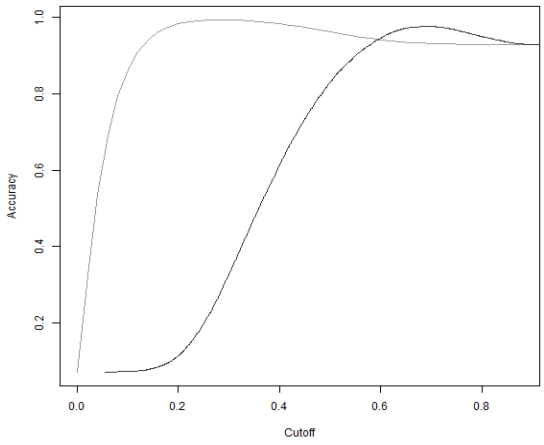

Figure 3 Accuracy vs Cutoff

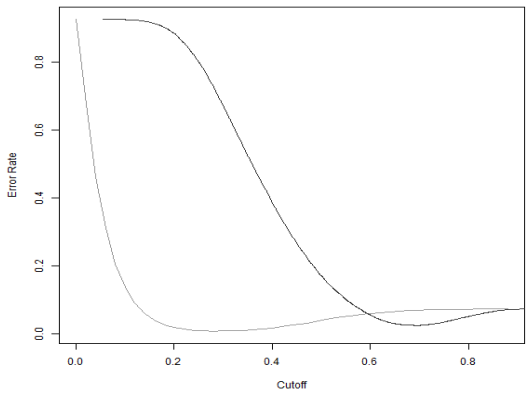

Figure 4 Error vs Cutoff

1.7.2 Upsell prediction

The below figure5 shows the RoC curve in terms of upsell, upselling is method which is used for upgrading the customer, in below graph we observe that in case of our methodology ranks all the positive above all the given negative.

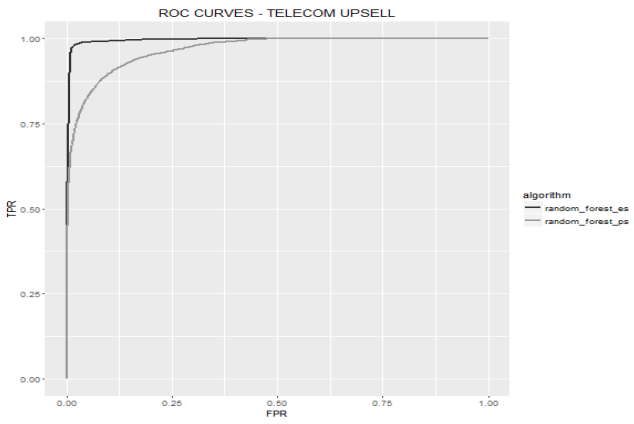

Figure 5Roc Curve for upsell

1.7.2.1 Accuracy vs Cutoff and Error Rate vs Cutoff

Accuracy is one of the parameter in order to evaluate the classification model; accuracy is the number of correct prediction divided by total number of prediction. Moreover, in case of binary classification the accuracy is calculate by using the TP, FP,TN, FN. Moreover, in below graph (Figure 6) we see that our methodology possesses the better accuracy when the cutoff is considered. Similarly, in the figure7 , the Error Rate graph is plotted against the Cutoff, and we observe that the Error Rate generated by the proposed methodology is better than the existing model(Standard Ransom Forest).

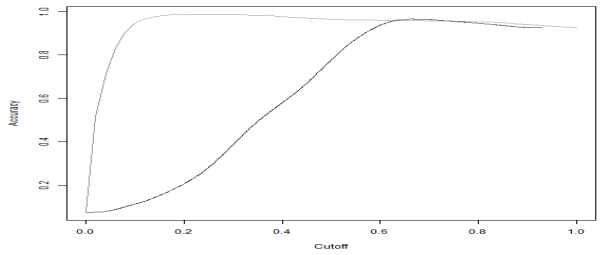

Figure 6 Accuracy Vs cutoff (upsell)

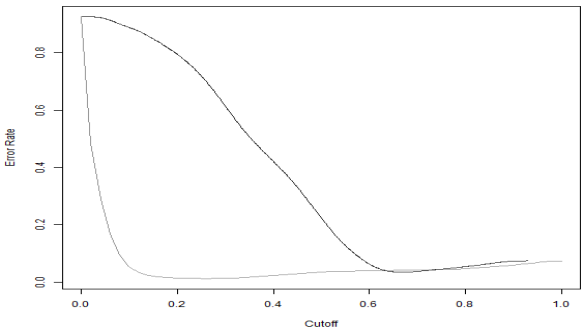

Figure 7 Error Rate vs Cutoff(upsell)

V. CONCLUSION

Customer attrition,which is firmly known as customer churn is losing customers, In recent Telecom industry has focused on analyzing the customer churn rate. In this research work, we have developed a model namely MRF(Modified Random Forest) for estimating the churn rate in the telecom industry. Our methodology provides the high classification accuracy, better AUC and our model can be adopted because of the same. In order to evaluate our algorithm has been compared with the existing methodology. In here we have considered two scenario i.e. churn and upsell, upselling provides the flexibility to encourage the customer to get the high-end product. In this paper, we take both churn as well as churn and compare it with the existing methodology i.e. standard Random Forest technique. The comparison takes place based on the two parameter i.e. accuracy vs cutoff and error rate vs cutoff, the graph clearly indicates that our algorithm outperforms the existing algorithm. Moreover, In ROC graph we observe that in case of our methodology ranks all the positive above all the given negative in both scenario i.e. churn as well as upsell. Later for more evaluation, our algorithm is compared in terms of AUC with the state of art technique. CSS LRM has AUC of 0.7354, whereas ANN possesses 0.7352 of the curve. Similarly, in case of standard RF model is good and it possesses 0.9929 of AU. However our model performs better than all these, Later the comparative analysis is done based on the classification accuracy and our model MRF simply outperforms the other existing methodology of churn prediction including the standard Random Forest.

ACKNOWLEDGMENT

The author would like to thank the head of department, Principal and management of Rajarajeswari college of Engineering and K.S Institute of Technology, Bangalore for their generosity towards supporting the research work.

REFERENCES

1. Halim, Joseph & Vucetic, Jelena. (2016). Causes of Churn in the Wireless Telecommunication Industry in Kenya. International Journal of Management Science and Engineering Research. 3. 1. 10.14355/ijmser.2016.0301.01.
2. K. Dahiya and S. Bhatia, "Customer churn analysis in telecom industry," *2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions)*, Noida, 2015, pp. 1-6.
I. M. M. Mitkees, S. M. Badr and A. I. B. ElSeddawy, "Customer churn prediction model using data mining techniques," *2017 13th International Computer Engineering Conference (ICENCO)*, Cairo, 2017, pp. 262-268.
3. H. Harb, A. Makhoul and C. Abou Jaoude, "A Real-Time Massive Data Processing Technique for Densely Distributed Sensor Networks," in *IEEE Access*, vol. 6, pp. 56551-56561, 2018.
4. S. U. Natchiar and S. Baulkani, "Customer relationship management classification using data mining techniques," *2014 International Conference on Science Engineering and Management Research (ICSEMR)*, Chennai, 2014, pp. 1-5.
5. S. Gupta, D. Hanssens, B. Hardie, W. Kahn, V. Kumar, N. Lin, N. Ravishanker, and N. R. S. Sriram, "Modeling Customer Lifetime Value," J. Serv. Res., vol. 9, no. 2, pp. 139–155, Nov. 2006.
6. 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.
7. 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.
8. M. C. Mozer, R. Wolniewicz, D. B. Grimes, E. Johnson, and H. Kaushansky, "Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry," Neural Networks, IEEE Trans., vol. 11, no. 3, pp. 690–696, 2000.
9. S. V Nath and R. S. Behara, "Customer churn analysis in the wireless industry: A data mining approach," in Proceedings-Annual Meeting of the Decision Sciences Institute, 2003.
10. 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.
11. Y. Huang and T. Kechadi, "An effective hybrid learning system for telecommunication churn prediction," Expert Syst. Appl., vol. 40, no. 14, pp. 5635–5647, Oct. 2013.
12. V. Umayaparvathi and K. Iyakutti, "Applications of Data Mining Techniques in Telecom Churn Prediction," Int. J. Comput. Appl., vol. 42, no. 12, 2012.
13. K. binti Oseman, 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.
14. D. T. Pham and M. S. Aksoy, "RULES: A simple rule extraction system," Expert Syst. Appl., vol. 8, no. 1, pp. 59–65, Jan. 1995.

15. A.M. AlMana and M. S. Aksoy, “An Overview of Inductive Learning Algorithms,” Int. J. Comput. Appl., vol. 88, no. 4, pp. 20–28, 2014.
16. M. S. Aksoy, H. Mathkour, and B. A. Alasoos, “Performance evaluation of rules-3 induction system on data mining,” Int. J. Innov. Comput. Inf. Control, vol. 6, no. 8, pp. 1–8, 2010.
17. Zhao, L., Gao, Q., Dong, X. et al. "K- local maximum margin feature extraction algorithm for churn prediction in telecom" Cluster Comput (2017) 20: 1401.
18. Vijaya, J., Sivasankar, E. (2018). Improved Churn Prediction Based on Supervised and Unsupervised Hybrid Data Mining System. ICT4SD2016, pp. 485–499. Springer, Singapore.

AUTHORS PROFILE

Author-1

MrsSwetha P is a BE, M. Tech graduate from VTU. She is currently pursuing PhD in the domain of Machine Learning from VTU, Karnataka. She is an active member of LMISTE, IEEE, IAENG. She has guided many students’ projects and were also funded by IEEE and KSCST. She also received best student project award from IEAE for guiding them. She published many papers in various National and International conferences and journals with Scopus indexing. Currently she is working as an Assistant Professor in Computer Science and engineering department, Rajarajeswaricollege of engineering, Bangalore, Karnataka.

Author-2

Dr. Dayananda R. B is a **Professor** in Department of Computer Science and Engineering, KSIT, Bengaluru. He is also the Director of IQAC, KSIT. He has academic experience of 17 years, holding various posts as HOD and Vice Principal at Prestigious Engineering Institutions.

His research mainly focuses on Design and Development of a Cloud Computing Architecture for data security. He has been felicitated with **Governor’s Award**-a National award for Excellence in Research and Development. His other **achievements** include:

1. **SHIKSHA RATAN** Award for talented personalities presented on Saturday 24th September 2016 at New Delhi.
2. **Research supervisor at VTU GUIDING 7 STUDENTS AT PRESENT**
3. **Certificate of appreciation** in rolling out Infosys campus connect Foundation program
4. Research committee member **at Global Research Academy**, a center of excellence, approved by ministry of Science and Technology, Govt of India
5. Represented SBIT, Tiptur at **‘sankara channel’** in TURNING POINT Program
6. Organized 29th CSI State level student convention at GSSSIETW, Mysuru.
7. Organized in association with CSI, the **International conference CONSEG-2011** on Software engineering at Chancery pavilion, residency road, Bangalore-25
8. He is reviewer at Springer Journal (Cluster Computing) and also reviewed many papers of Conferences and symposiums.