Proceedings of the International Conference on Inventive Computing and Informatics (ICICI 2017)
IEEE Xplore Compliant - Part Number: CFP17L34-ART, ISBN: 978-1-5386-4031-9

A Comparative Study of Customer Churn Prediction in Telecom Industry Using Ensemble Based Classifiers

Abinash Mishra
Research scholar
Department of Computer Applications
National Institute of Technology
Tiruchirappalli, Tamil Nadu, India
405117002@nitt.edu

U. Srinivasulu Reddy
Assistant Professor
Department of Computer Applications
National Institute of Technology
Tiruchirappalli, Tamil Nadu, India
usreddy@nitt.edu

***Abstract*— Churn Prediction plays a vital role in various domains like life insurance, banking and telecom industry. With the current advancement in Machine Learning and Artificial Intelligence, Churn Prediction is more realistic and accurate. It is very much essential for early stage detection of customers who are at high risk of leaving the company or services. In this paper, Ensemble based Classifiers namely Bagging, Boosting and Random Forest were utilized for Churn Prediction in telecom industry. The Ensemble based Classifiers were compared with the well-known classifiers namely Decision Tree, Naïve Bayes Classifier and Support Vector Machine (SVM). The experimental results shows that Random Forest has less error rate, low specificity, high sensitivity and greater accuracy of 91.66% as compared to other methods.**

Keywords—Machine Learning; Artificial Intelligence; Churn Prediction; Telecom Industry; Ensemble based Classifiers; Base Classifiers.

I. Introduction

Churn Prediction is a fundamental problem in Telecom Industry. Churn defines the customer who are moving from one Telecom service provider to another Telecom service provider. Attracting new customers into a company is more tedious, time consuming and costly affair. So a company should aim on existing customer rather than looking for new customer acquisition. ***Michael C. Mozer et al.*** [1] proved that the rate at which customer moving from one company to the other determines the rate of loss. Hence it is highly essential to identify the customer who are likely to leave the service. ***Wai-ho au et al.*** [2] defined the Churn Prediction problem as a classification problem, *i.e.*, it discovers classification rules which help to classify a record to one of the predefined class whose label is unknown**. N. Kamal raj and A. Malathi [3]** proposed application of Data Mining Techniques for Telecom Churn Prediction. In this article, the main idea of this article is to classify churn and non-churn customers in Telecom Industry by using an efficient Ensemble based Classifiers. And also to make a comparative study of these Ensemble based Classifiers with more popular based Classifiers. This process will help the service provider to detect loyal customers and to maximize the profitability.

The paper is organized as follows first section explains the notion of Churn Prediction Model, second part is describes the literature survey on Churn Prediction Models. The working methodology has been explained in the third section. In fourth section, the experimental results and analysis are discussed, and finally the conclusion and future work.

The churn prediction model with risk labels (low, medium and high) are explained by considering the following steps. In the first step one should clearly define the business objectives, after knowing the objectives. It is very much essential to identify the parameters involved in the process. Later, a model has to be built and the same has to be tested to determine the risk labels. The entire process has been schematically presented in *Fig 1.*

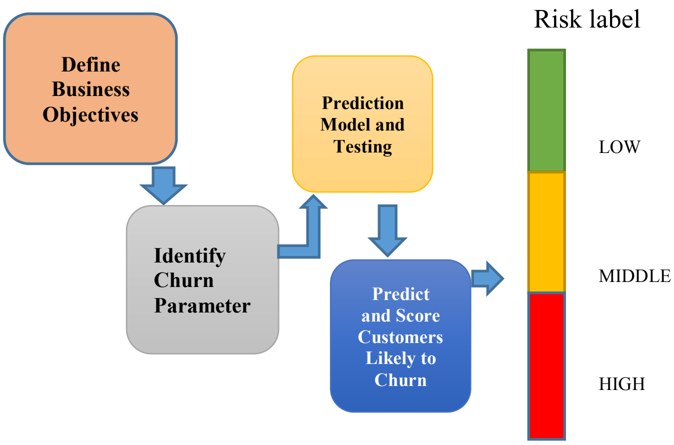

Fig. 1. Schematic of Churn Prediction Model.

The whole idea is to classify the customers into churner and non-churner group in the telecom industry. To do the same, it is essential to identify the reason for churning, based on past behaviour of customers. A company should early detect the customers who are likely to churn from the company or services. It is the most important for any company or service provider to retain the customers rather than looking for the new customers. Hence, accurate prediction of churner group helps in predicting the company's profit.

II. LITERATURE SURVEY

In the recent years, Reliance Jio attracted many customers from the other service providers in India. It is very hard for other service providers to retain the customers who are likely to leave the service. To avoid such situation, it is highly essential to predict the customer's attitude to retain them for prolonged period of time. Towards the same, various research team worked on different machine learning techniques. The following section describes the previous research work carried out in the churn prediction problem based classification models.

Chuanqi Wang et al. [4] defined Churn Prediction Model as cost sensitive classification problem. They coined the term cost sensitive because the way the model was designed to classify the customers into churner and non-churner group will decide the maximum profit earned by the company. The proposed work, the authors showed the classification performance and the misclassification rate.

Adnan Idris, and Asifullah Khan [5] designed an efficient Churn Prediction Model on Orange Dataset and Cell2Cell Dataset. First, the feature selection technique was applied on minimum redundancy and maximum relevance. The base Classifiers was combined with the Ensemble based Classifiers to obtain majority voting, so that, it can predict future instance more accurately. Random Forest, Rotation Forest and *k*-NN were utilized for the same. In this article, sensitivity, specificity, AUC (Area Under Curve) and Q-statistics measure were used as a performance metrics.

Guo-en Xia et al. [6] explained a Churn Prediction Model on the real Telecom dataset. In this research article, the weighted selective ensembles were used for churn prediction and these methods were compared with the Base Classifiers namely Decision Tree, Naïve Bayes, SVM and Artificial Neural Network (ANN). The experimental results shows that the proposed method performs better than the Base Classifiers.

Ning Lu et al. [7] proposed a novel Churn Prediction Model using Boosting algorithm. In this approach, customers were divided into two clusters based on the weight assigned by the Boosting algorithm. The Logistic Regression was used on each Base Classifier to predict churn customer. The results depicted that Boosting was good classifier for the Churn Prediction analysis.

III. WORKING METHODOLOGY

This section explains how the Ensemble based Classifiers and well-known Base Classifiers were used in the Churn Prediction Model.

A. Decision Tree

Decision Tree (C4.5) [8] was developed to overcome the drawbacks of ID3 algorithm [10]. C4.5 utilizes the benefits of greedy approach and uses a series of rules for classification. Although this approach gives a high classification accuracy rate it fails to respond to noisy data. Gain is the main metric used in the decision tree to decide the root node attribute.

B. Naïve Bayes

Naïve Bayes [8] is a brute-force method for training the model. The underlying principle behind Naïve Bayesian classifier is Bayes Theorem [10]. For the classification problem, each predictor attribute was consider separately with class label for model construction using training dataset. Naïve Bayes algorithm is represented as follows:

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)} \quad (1)$$

A: represents the predictor attributes.
B: represents the response variable *i.e*., churn or non-churn.

Predictor attribute includes the area, service calls, evening calls night calls etc. Apply the conditional probability for each attribute belongs to all the predictor attributes given that class label represents churn. The disadvantage of this methods is, it is not suited for the large dataset.

C. Support Vector Machine

SVM [9] algorithm was proposed by ***Boser, Guyon, and Vapnik.*** It was very well used for both classification and regression problem [9]. SVM maps all the data points to a higher dimensional plane to make the data points linear separable. The plane which divides data points is known as hyper plane. It can be used for small dataset to give an optimal solution. SVM cannot be more effective for noisy data.

SVM [10] model tries to find out the churn and non-churn customer. In order to divide the dataset into churner and non-churner group, first it will take all the data points in *n-dimensional* plane and divide the data points into churner and non-churner group based on maximum marginal hyper plane. Based on the maximum marginal hyper plane it will divide the data points into churner and non-churner group. Here *n* represents the number of predictor variable associated with the dataset.

D. Bagging

Bagging [10] (or Bootstrap aggregation) is one of the Ensemble based Classifiers which consist of bag of similar type or dissimilar type base classifiers. Bagging algorithm helps to reduce the variance of the classifier used for the Churn Prediction Model in order to increase the performance.

The steps in designing of Churn Prediction Model using Bagging algorithm are as follows, First, it is required to divide the input dataset into *k* subset with replacement, then it requires to train the model by using the *(k-1)* subset and test the model using the dataset which has not been used for training model. The experimental results showed that, Bagging is effective, because it predicts the test instances using the classifier which has more accuracy from the bag of classifier, Bagging requires heavier computational resource for the Model construction.

E. Boosting

Boosting [10] Ensemble technique is designed in such a way that it will maintain a weight for each training tuple. After a classifier is learned from the training tuple, weights are updated for the subsequent classifier. The final Boosted Classifier combines the vote of each individual classifier for prediction to improve the performance of the classifier. In similar to SVM, this model is also not suited for noisy data.

The key idea for the customer Churn Prediction using Boosting algorithm is to train a series of classifier simultaneously and keep updating the model accuracy for improving the performance of the classifier.

F. Random Forest

Random forest [10] (***Breiman 2001)*** works based on the random subspace method. The designed strategy used in Random Forest is divide and conquer. It forms number of Decision Trees and each Decision Tree is trained by selecting any random subset of attribute from the whole predictor attribute set. Each tree will grow up to maximum extent based on the attribute present in the subset. Then after, based on average or weighted average method, the final Decision Tree will be constructed for the prediction of the test dataset.

Random forest runs efficiently in large dataset. It can handle thousands of input variables without variable deletion. It also handle the missing values inside the dataset for training the model. It is difficult to handle the unbalanced dataset by using Random Forest.

IV. Proposed Work

In this section, the proposed work has been explained by considering dataset, performance measures used in this article.

A. Dataset

The design of churn prediction model in Telecom Industry requires the past history or past behaviour of customer during a specific period of time to predict their behavior in near future. Therefore, it is the usual practice to gather information of all the customers as much as possible. The dataset which is used in this research papers has been collected from the web link *http://www.ics.uci.edu~mlearn/MLRepository.html.*

The dataset consists of *3333* records and each record is described by the following attributes *i.e.* Churn as class label, area, service calls, evening calls, evening charge, minutes spend in the evening, day calls, day charge, minutes spend in day time, international calls, international charge, minutes spend in international calls and finally it includes night calls, night charge, minutes spend in nights. The target attribute is the churn *i.e.* a customer is going to churn or not, and all the remaining attributes are the predictor attributes, which will have some impact to determine the target attribute based on the information obtained by the attribute.

B. Performace Metrics

The performance of the any classification model can be evaluated with the help following measures:

- *Accuracy*
- *Error rate*
- *Specificity*
- *Sensitivity*

To evaluate the performance characteristic confusion matrix model was chosen. A **confusion matrix** is a matrix used to designate the efficiency of a classification model on a group of test data for which the true values are known. The values for different classifiers was obtained using R studio tool. From the confusion matrices the value of *True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN)* can be evaluated then applying values in the equation given below gives the performance measure of the model.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

$$Error\ rate = 1 - Accuracy \tag{3}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{4}$$

$$Specificity = \frac{TN}{TN + FP} \tag{5}$$

It is shown that all the parameter of confusion matrices for Ensemble based classifiers and basic classifiers for churn prediction model in terms of *True Positive, True Negative, False Positive, False Negative,* which are shown in the table I. In this paper, *TP* represents they are churn customer and predicted as churn, *TN* represents they are non-churn customer and predicted as non-churn, *FP* represents predicted as churn but actual value is non-churn, *FN* represents predicted value is churn but actual value is non-churn.

TABLE I. CONFUSION MATRIX FOR CHURN PREDICTION MODEL.

Classifier Name	TP	TN	FP	FN
Bagging	796	86	69	20
Boosting	799	78	17	77
Random Forest	807	83	9	72
Decision Tree	879	58	72	21
SVM	880	59	98	5
Naïve Bayes	806	74	81	56

TABLE II. PERFORMANCE METRICS FOR BASE AND ENSEMBLE BASED CLASSIFIERS

Classifier Name	Accuracy	Error rate	Specificity	Sensitivity
Bagging	90.83	9.17	81.13	92.02
Boosting	90.32	9.68	50.32	97.91
Random Forest	**91.66**	**8.34**	**53.54**	**98.89**
Decision Tree	90.97	9.03	73.41	92.42

SVM	90.12	9.88	92.18	89.98
Naïve Bayes	86.53	13.47	56.92	90.86

The performance of the Churn Prediction Models are compared in the table II. Specificity represents the portion of negative cases that were classified correctly and sensitivity represents the portion of positive cases that were correctly identified. From table II one can understand that the Random Forest performed better than other methods with the following results: more accurate with 91.66%, low error rate of 8.34%, less specificity 53.54 and high sensitivity 98.89. The same experimental results are visualized from following Fig. 2, 3, 4 & Fig 5.

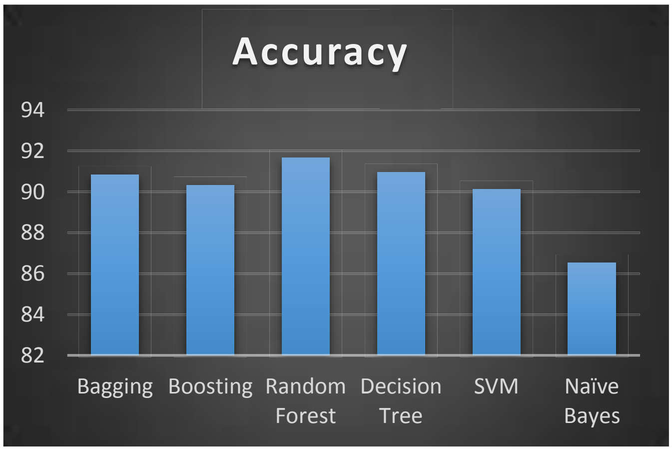

Fig. 2. Measure of Accuracy

Fig. 3. Measure of Error rate.

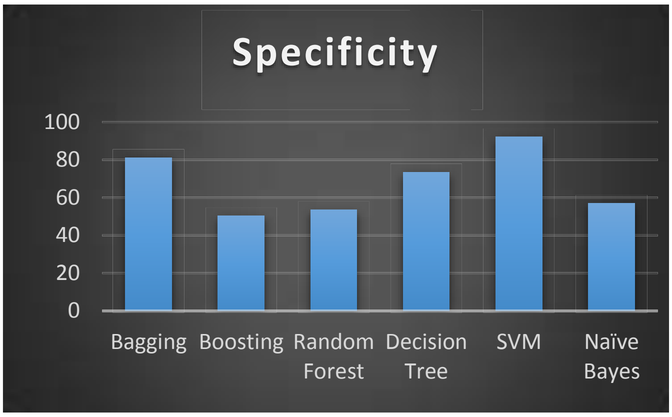

Fig. 4. Measure of False Positive Rate.

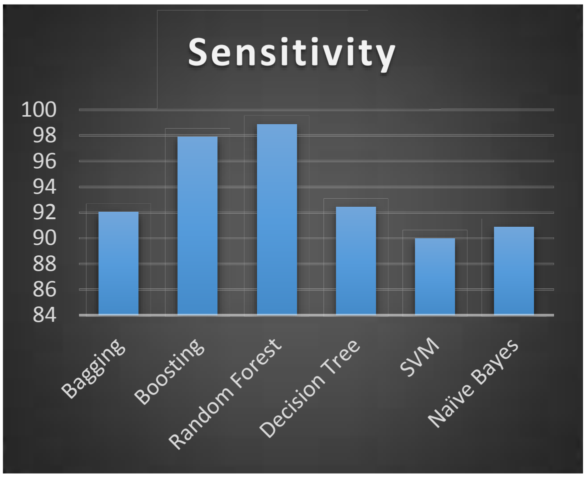

Fig. 5. Measure of True positive rate

Experimental results of Random Forest was compared with existing Churn Prediction Model namely Classification and Regression Tree (CART) and its variants (Table III) proposed by ***Chuanqi Wang et al.*** [3]. The results clearly depicted that the Random Forest gives better results when compared to the CART and its types, based on Precision, Recall and F-score.

TABLE III. COMPARISION OF CART AND ITS VARIANTS WITH RANDOM FOREST

Model	Precision	Recall	F-score
CART	62.21	75.01	34.01
CARTCS-A	34.22	88.91	27.21
CARTCS-M	36.28	93.16	26.11
CARTCS-B	36.28	93.16	26.11
P-CARTCS	67.56	70.41	34.48
Random Forest	**83.11**	**98.89**	**95.22**

The experimental results of Random Forest are also compared with base classifiers and weighted selective ensembles proposed by ***Guo-en Xia et al*** [5]. Which are shown in the table IV. From the results, the Random Forest is highly accurate than the other models.

TABLE IV. COMPARISON OF BASE CLASSIFIERS AND WEIGHTED ENSEMBLES WITH RANDOM FOREST.

Model	Accuracy
Bayes	45.73
C4.5	56.68
ANN	55.61
SVM	69.95
LIBSVM	68.24
Probability Weighted Integration	74.52
Gaussian Weighted Integration	74.80
Random Forest	91.66

V. CONCLUSION AND FUTURE WORK

In the current digital world, the usage of the smart mobile phones are very much essential for every human life. Due to this, many service providers would like to give values added services to retain their customers. Many Telecom companies are facing difficulty to predict the customer who is likely to leave the services. In Telecom Industry, Churn Prediction is a fundamental problem which is gaining attention of many researches in the recent years.

In this research paper, the authors proposed a Comparative Study of Customer Churn Prediction in Telecom Industry using Ensemble Based Classifiers and also compared with the existing well-known base Classifiers. The experimental results shows that the Random Forest is the best Classifier for the Churn Prediction Problem when compared to others models, in terms of all the performance measures like accuracy, sensitivity, specificity and error rate.

Every customer is expecting good service or reward points from the service providers. Enabling the prompt services for the valid customers is more tedious task, because it is very difficult to predict the genuine customers for the company. The early churn prediction can prevent the company loss by predicting the customer behaviour. In future, the Reinforcement Learning or Deep Learning are the two prominent techniques to addresses the Churn Prediction Problem.

REFERENCES

[1] Michael C. Mozer, Richard Wolniewicz, David B. Grimes,"Predicting Subscriber Dissatisfaction and Improving Retention in the Wireless Telecommunications Industry," IEEE Transactions On Neural Networks, Vol. 11, pp. 690-696, September 2000.

[2] Wai-Ho Au , K. C.C. Chan , Xin Yao, "A novel evolutionary data mining algorithm with applications to churn prediction," IEEE Transactions on Evolutionary Computation, Vol. 7 (6) pp. 532-545, December 2003.

[3] N. Kamalraj, A.Malathi,"Applying Data Mining Techniques in Telecom Churn Prediction, " in proc. International Journal of Advanced Research in Computer Science and Software Engineering, 10, October 2013.

[4] Chuanqi Wang, Ruiqi Li, Peng Wang, Zonghai Chen,"Partition cost-sensitive CART based on customer value for Telecom customer Churn Prediction, Control Conference (CCC), September 2017.

[5] Adnan Idris, Asifullah Khan,"Ensemble based Efficient Churn Prediction Model for Telecom", Frontiers of Information Technology (FIT), International Conference, pp.5680-5684,June 2015.

[6] Guo-en Xia, Hui Wang, Yilin Jiang,"Application of Customer Churn Prediction Based on Weighted Selective Ensembles," International Conference on Systems and Informatics (ICSAI 2016), pp. 513-519, November 2016.

[7] Ning Lu, Hua lin, Jie Lu, Guangquan Zhang,"A Customer Churn Prediction in Telecom Industry using Boosting. Customer Behaviour in Telecommunications," IEEE Transaction on Industrial Informatics, Vol. 10,(2), May 2014.

[8] D. Lopez and L. Sigrist, "A Centalized UFLS Scheme Using Decision Tree For Small Isolated Power System," IEEE Latin America Transactions, Vol. 15, (10), October 2017.

[9] Ali Dehghan, Theodore B. Trafalis, "Examining Churn and Loyalty Using Support Vector Machine", Scienceedu Press, Vol. 1, (4), pp. 153-161, December 2012.

[10] Jiawei Han and Micheline Kamber, "Data mining, concept and techniques" [http://www.cs.sfu.ca].

[11] http://www.ics.uci.edu~mlearn/MLRepository.html

[12] http://www.kloutix.com/wp-content/uploads/2016/02/churn_new.png