International Journal of Engineering Innovation & Research
Volume 11, Issue 4, ISSN: 2277 – 5668

Telecom Customer Churn Prediction Using Adaboost Classifier and Neural Network

Shivani Vaidya[1*] and Rajesh Kumar Nigam[2]
[1]M.Tech Scholar, Technocrats Institute of Technology, Bhopal, Madhya Pradesh, India.
[2]Associate Professor, Technocrats Institute of Technology, Bhopal Madhya Pradesh, India.
[*]Corresponding author email id: vaidyashivani08@gmail.com

Date of publication (dd/mm/yyyy): 22/07/2022

Abstract **– Churn is normally occurring today. Churn prediction is very tuff process in all industry. The Telecommunications (telecom) industry is saturated day by day and marketing strategies are focusing on customer retention and churn prevention. Churning is when a customer stops using a company's service thereby opting for the next available service provider. This churn threat has led to various Customer Churn Prediction (CCP) studies and many models have been developed to predict possible churn cases for timely retention strategies. Customer churn is always a grievous issue for the Telecom industry as customers do not hesitate to leave if they don't find what they are looking for. They certainly want competitive pricing, value for money and above all, high quality service. Customer churning is directly related to customer satisfaction. It's a known fact that the cost of customer acquisition is far greater than cost of customer retention, that makes retention a crucial business prototype. Data preprocessing, data normalization and feature selection have shown to be prominently influential. Monthly data volumes have not shown much decision power. Average Quality, Churn Risk and to some extent, Annoyance scores may point out a probable churner. Weekly data volumes with customer's recent history and necessary attributes like age, gender, tenure, bill, contract, data plan, etc., are pivotal for churn prediction. Data preparation methods and churn prediction challenges have also been explored. This study reveals that Support Vector Machines, Naïve Bayes, Decision Trees and Neural Networks are the mostly used CCP techniques. Feature selection is the mostly used data preparation method followed by Normalization and Noise removal. Under sampling is the mostly preferred technique for handling telecom data class imbalances. Imbalanced, large and high dimensional datasets are the key challenges in telecom churn prediction.**

Keywords **– Big Data, Churn Prediction, Decision Tree, Quality of Experience.**

I. INTRODUCTION

In the telecom industry, the biggest loss of revenue is happening because of increasing customer churns. Such customers who are not loyal to the company result in a financial burden on the company. This fact is very well known that the cost of finding new customers is far more than retaining the old ones. So, detecting the "going to be churner" customers beforehand is the objective of the telecom companies. This poses a serious threat among companies because customers have many options to switch to (Fei, Shuan, & Yan, 2017), which is termed as customer churn. Churning has attracted a lot of research from various scholars to aid timely detection of churners before they practically use their switching intentions (Umayaparvathi & Iyakutti, 2016). Timely detection of potential churners saves telecom companies from persuading costs that would be incurred to attract new customers or win back already churned customers. Such costs are 5 to 6 times higher than customer retention costs (Idris, Iftikhar, & Rehman, 2017; Zhu, Baesens, & vanden Broucke, 2017). Many studies have revealed that high prices are among the top factors causing customer churn. Akmal (2017)'s qualitative study found out that high rates and bills are key factors causing churn. Mehwish, Zaffar and Sumaira (2017) also confirmed that network charges are among the top factors causing churn. Churning is also accelerated with good Mobile Number Portability (MNP) services (Adebiyi, Oyatoye and Amole 2016). Companies therefore have to try as much as possible to avoid customer churn cases because customers who leave a company have capabilities to influence members of their social groups to do the same (De Caigny, Coussement, & De Bock, 2018). Telecom companies deal with two kinds of customers namely Business to Consumer (B2C) and Business to Business (B2B) customers. B2C transactions provide telecom services to individual consumers while B2B transactions provide telecom services to other businesses.

In competitive Telecom market, the customers want competitive pricing, value for money and high quality service. Today's customers won't hesitate to switch providers if they don't find what they are looking for. This phenomenon is called churning. Customer churning is directly related to customer satisfaction. Since the cost of winning a new customer is far greater than cost of retaining an existing one, mobile carriers have now shifted their focus from customer acquisition to customer retention. After substantial research in the field of churn prediction over many years, Big Data analytics with Machine Learning was found to be an efficient way for identifying churn. These achieve results more efficiently and receive insights that sets alarm bells ringing before any damage could happen, giving companies an opportunity to take precautionary measures. These techniques are usually applied to predict customer churn by building models and learning from historical data. However, most of these techniques provide a result that customers might churn or not, but only few tell us why they churn. Conducting experiments with end users' perspective, gathering their opinions on network, data normalization, preprocessing data sets, employing feature selection, eliminating class imbalance and missing values, replacing existing variables with derive n their customers more efficiently. Comparatively, a smaller study was done on user's perspective, taking d variables improves the accuracy of churn prediction which assists Telecom industries to retai into consideration their quality of experience. In fact, no study was done taking into consideration only user's data volumes. Estimation of Quality of Experience by finding relationships between QoE and traffic characteristics could help the service providers to continuously monitor the user satisfaction level, react timely and appropriately to rectify the performance problems and reduce the churn. In order to classify customers into potential churners and non-churners, Mitrovic et al. (2017) confirmed that various Customer Churn Prediction (CCP) studies have been carried out and many models and techniques provided over time. However, many of these CCP studies initially did not consider the profit maximization objective which is the ultimate objective of any profit organization. Coussement, Lessmann and Verstraeten (2017) stated that companies should clearly distinguish non-profitable customers from those that are profitable. Profitable customers who are also potential churners are then given special attention with retention measures. As a result, retention campaigns-related costs are minimized because resources are utilized on the right and targeted customer group.

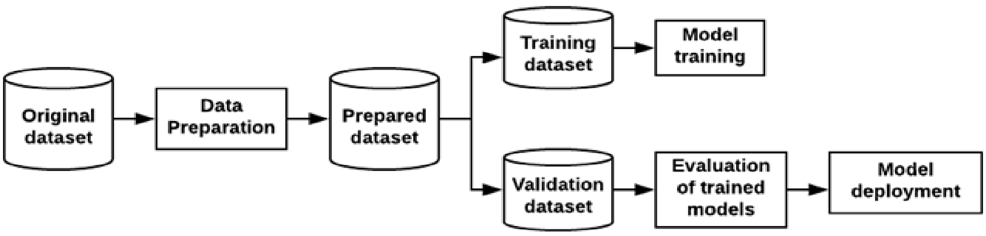


Fig. 1. Churn prediction modeling process.

CCP relies on machine learning algorithms to develop models that classify telecom customers into churners and non-churners. Churn modeling includes three major steps after data collection and before model deployment; data preparation, model training and model evaluation (Umayaparvathi & Iyakutti, 2016). Data preparation is aimed at making data suitable for machine learning algorithms and model training. 50-80% of the data mining effort is put to data preparation because the quality of data affects the model performance results (Zhang, Zhang, & Yang, 2010). Data preparation is also responsible for removing any bias in the data through class balances and other randomization procedures. Missing value imputation, data cleaning, transformation and general exploration is done in this phase (Federico, 2014).

A. *Data Mining*

Data Mining can be defined as "the process of searching large stores of data to discover patterns, associations and trends to dig out useful structures from large amounts of data stored in different databases or other information repositories." [w1] There are many organizations which are using data mining techniques for managing their customer relationships, including getting new customers, increasing revenue from existing customers, and retaining high valued and loyal customers.

B. *Data Mining in Telecom*

According to [1], data mining in the field of telecommunication can be used for the following purposes:

- Churn prediction: - The process of predicting the customers who are at a risk of leaving the company is known as churn prediction in telecommunication. These customers should be focused upon, and efforts should be made to retain them. This is very important because retaining a customer is less expensive than acquire a new one.
- Insolvency prediction: - Hike in the number of due bills are becoming an important area of concern for all telecom companies. In such a competitive environment, companies cannot bear the burden of insolvency. To find such insolvent customers, data mining technique can be used. Customers who will turn defaulters, can be predicted beforehand.
- Fraud Detection: - Fraud is an expensive affair for the telecom industry, so the companies should try to predict fraudulent users by identifying their usage patterns.

II. PROPOSED METHOD

A. *Churn Prediction Techniques*

A variety of churn prediction techniques were used. In certain studies, a combination of two or more techniques was used for model development and comparison purposes. Support vector machines emerged the mostly used technique, similar to Umayaparvathi and Iyakutti (2016)'s findings but contrary to Hashmi, Butt and Iqbal (2013)'s. The reason behind this rise in Support Vector Machines' usage is mainly due to their applicability in both regression and classification (Kumar & Chandrakala, 2017). Neural Networks have also been used with the same frequency as Support Vector Machines. This is due to their performance consistence even with large datasets. In this section, Support Vector Machines, Neural networks, Decision tree and Naïve Bayes are discussed considering the fact that they were used the most as evidenced by their frequency of usage.

B. *AdaBoost Classifier*

Ada-boost or Adaptive Boosting is one of ensemble boosting classifier proposed by Yoav Freund and Robert

Schapire in 1996. It combines multiple classifiers to increase the accuracy of classifiers. AdaBoost is an iterative ensemble method. AdaBoost classifier builds a strong classifier by combining multiple poorly performing classifiers so that you will get high accuracy strong classifier. The basic concept behind Adaboost is to set the weights of classifiers and training the data sample in each iteration such that it ensures the accurate predictions of unusual observations. Any machine learning algorithm can be used as base classifier if it accepts weights on the training set. Adaboost should meet two conditions:

1. The classifier should be trained interactively on various weighed training examples.
2. In each iteration, it tries to provide an excellent fit for these examples by minimizing training error.

AdaBoost Algorithm

It works in the following steps:

1. Initially, Adaboost selects a training subset randomly.
2. It iteratively trains the AdaBoost machine learning model by selecting the training set based on the accurate prediction of the last training.
3. It assigns the higher weight to wrong classified observations so that in the next iteration these observations will get the high probability for classification.
4. Also, it assigns the weight to the trained classifier in each iteration according to the accuracy of the classifier. The more accurate classifier will get high weight.
5. This process iterate until the complete training data fits without any error or until reached to the specified maximum number of estimators.
6. To classify, perform a "vote" across all of the learning algorithms you built.

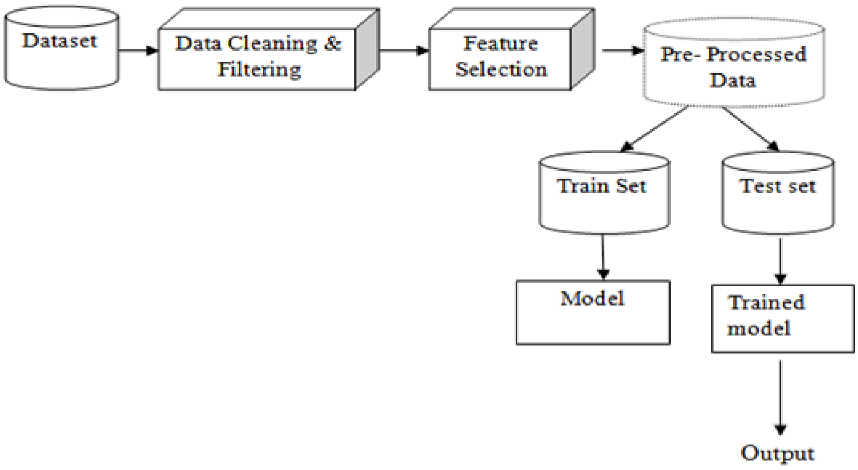


Fig. 2. System Architecture.

C. *Support Vector Machines (SVM)*

Khodabandehlou and Zivari Rahman (2017) described SVM as a machine learning technique for solving linear and non-linear classification problems. The authors asserted that SVM aims at minimizing the distance between hyper planes and classes to separate the classes as much as possible. In telecom CCP, SVM have been widely used and registered success (Yu et al., 2016). However, the technique was disregarded by Coussement, Lessmann and Verstraeten (2017) who claimed that it requires additional parameter tuning and often fails to give straightforward predictions. In this study, SVM registered the highest frequency of usage.

D. *Neural Networks (NN)*

Neural Networks work in a way that converts data into a brain neuron system (Monani et al., 2016), and are used in speech and pattern recognition as well as computer vision. The authors also noted its easy parallelization, continued learning even after application of the training set but to also observed that it is not easy to understand. In their survey on CCP techniques, the authors credited NN as better than most other methods. NN are also more capable for large datasets compared to other techniques. This justifies their increased adoption in the telecom churn prediction where large datasets are the order of the day. NN are however complex to comprehend and interpret.

III. SIMULATION RESULT

This section gives a brief overview of the statistical and decision tree analytical results for three different data sources.

A. *Churn Case Study*

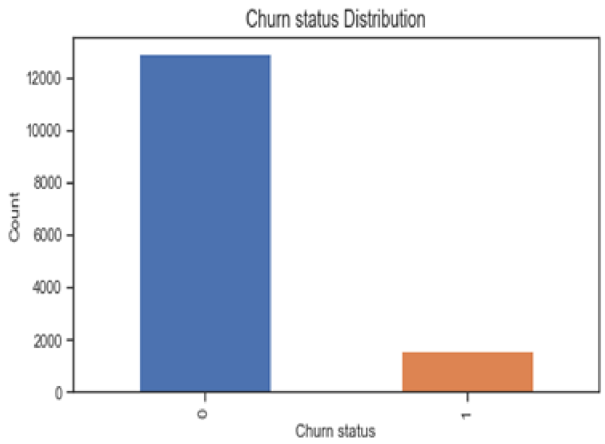
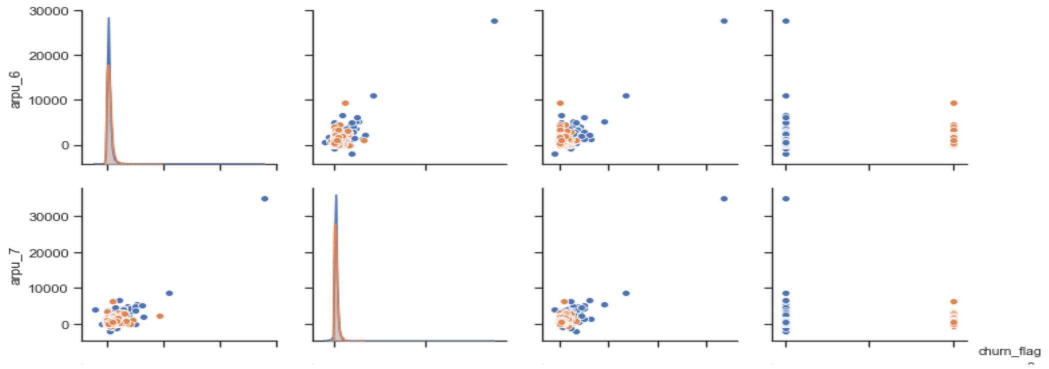


Fig. 3. Churn Status.

Figure 3 is showing churn status of user level 0 is showing not churning user and level 1 is showing churn user. In Telecom Manthan case study, when churn is distributed, it is seen in its graph that user is showing churn status of level 0 user is not churn and level 1 user is showing churn. show in Figure 3.



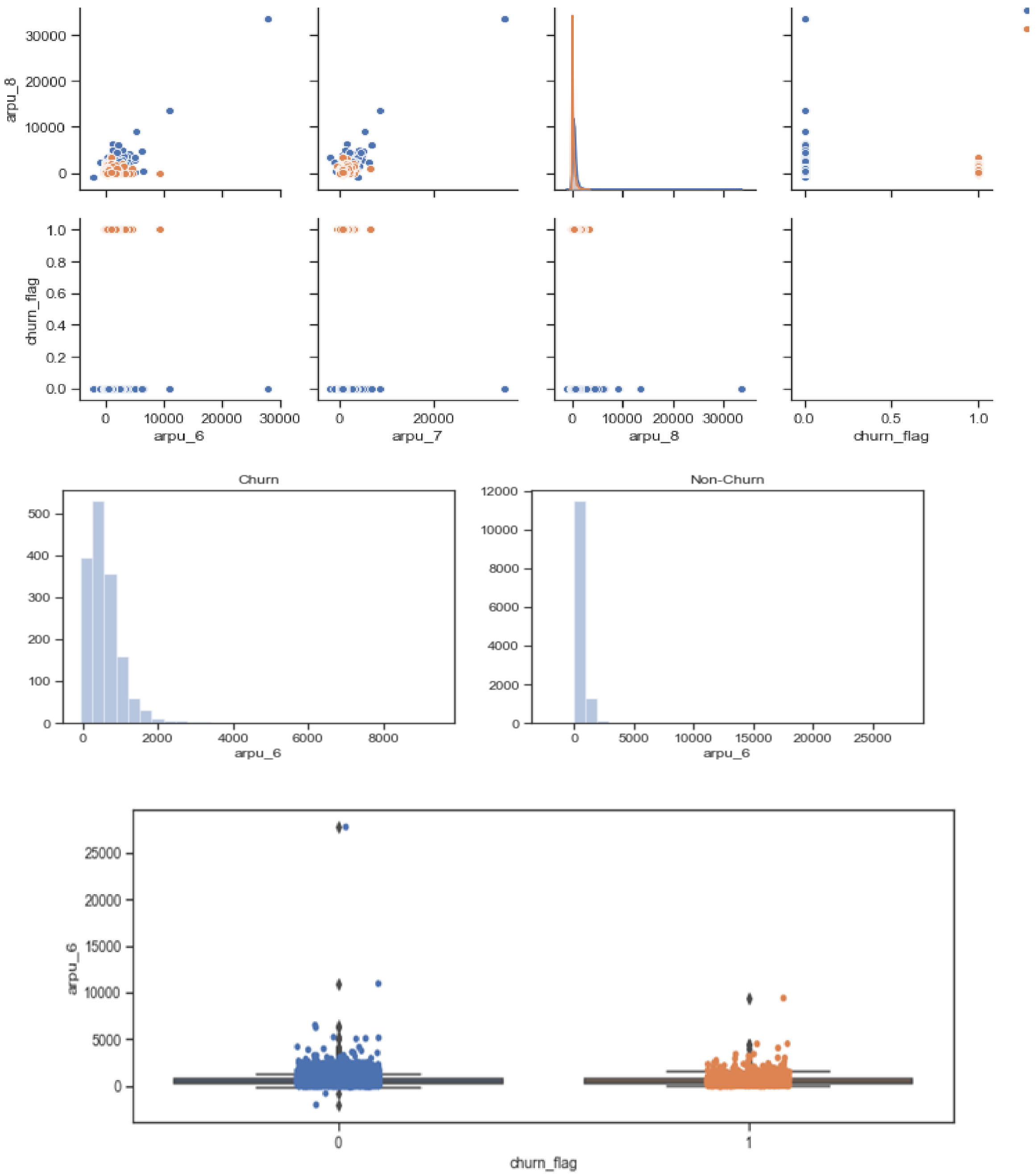


Fig. 4. Churning and Training Model.

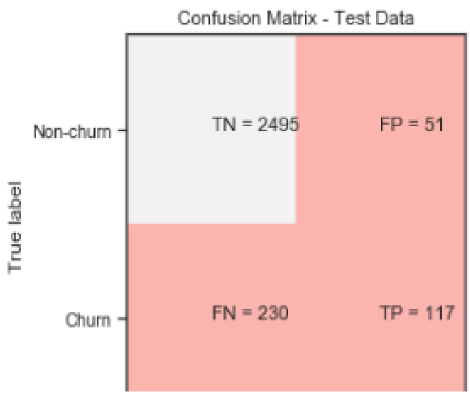


Fig. 5. Confusion Matrix.

International Journal of Engineering Innovation & Research
Volume 11, Issue 4, ISSN: 2277 – 5668

B. *Random Forest Classifier*

	precision	recall	f1-score	support
0	0.96	0.86	0.91	2546
1	0.41	0.75	0.53	347
micro avg	0.84	0.84	0.84	2893
macro avg	0.69	0.80	0.72	2893
weighted avg	0.90	0.84	0.86	2893

```
Accuracy for the test dataset 84.2%
ROC for the test dataset 87.2%
```

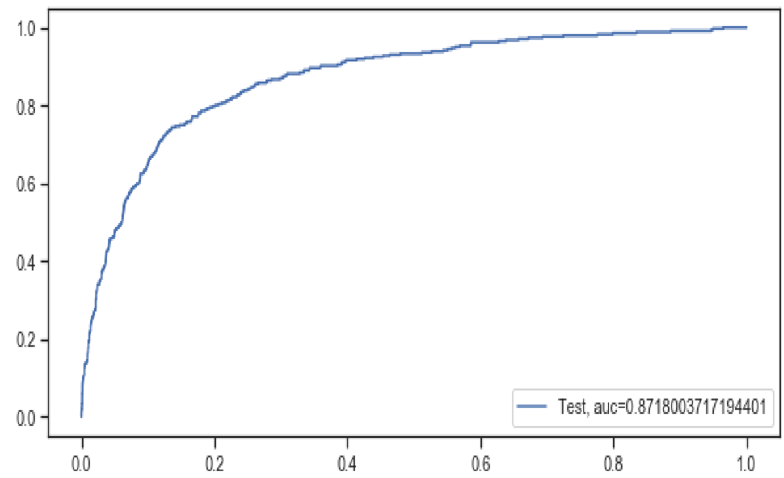


Fig. 6. Test set random forest classifier.

```
In [128]: plotLiftChart(y_test.values,preds_probs_RFC,"Random Forest")
```

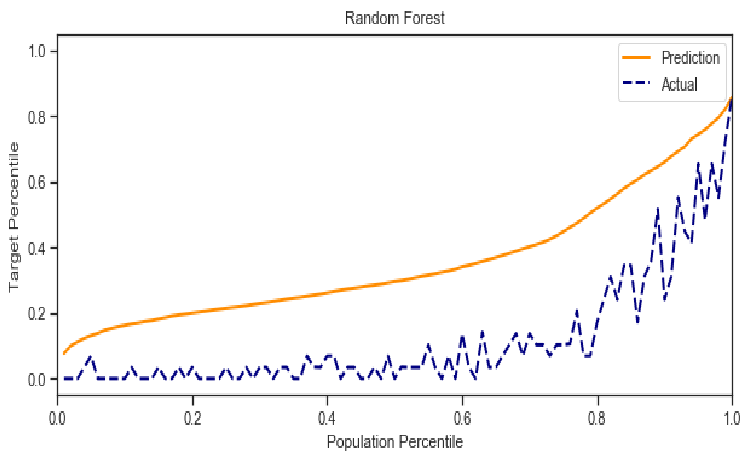


Fig. 7. Prediction random forest classifier.

C. *Gradient Boosting Classifier*

```
              precision    recall  f1-score   support

           0       0.91      0.99      0.95      2546
           1       0.74      0.31      0.43       347

   micro avg       0.90      0.90      0.90      2893
   macro avg       0.83      0.65      0.69      2893
weighted avg       0.89      0.90      0.89      2893

Accuracy for the test dataset 90.4%
ROC for the test dataset 88.6%
```

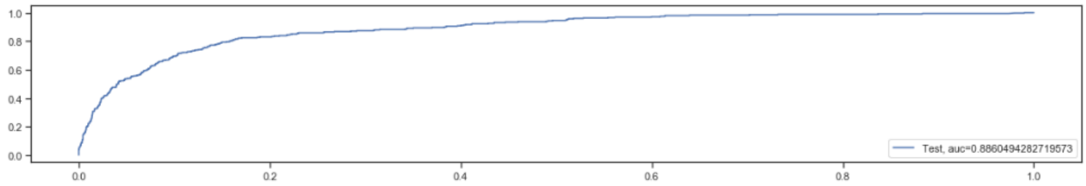


Fig. 8. Test set gradient boosting classifier.

```
In [135]: plotLiftChart(y_test.values,preds_probs_GBC,"Gradient Boosting")
```

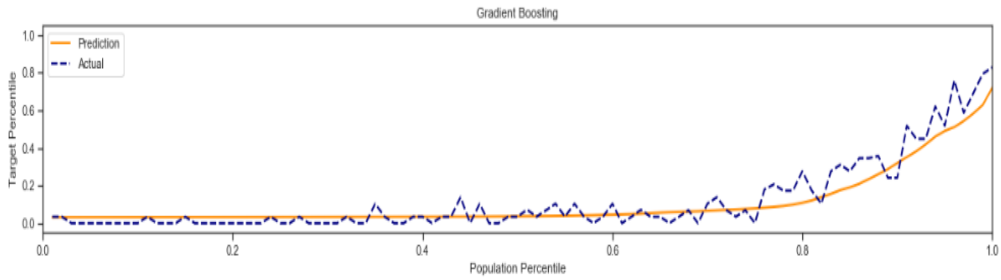


Fig. 9. Prediction gradient boosting classifier.

D. *Gradient Boosting with Hyper Parameter Tun*

```
              precision    recall  f1-score   support

           0       0.92      0.97      0.95      2546
           1       0.64      0.41      0.50       347

   micro avg       0.90      0.90      0.90      2893
   macro avg       0.78      0.69      0.72      2893
weighted avg       0.89      0.90      0.89      2893

Accuracy for the test dataset 90.1%
ROC for the test dataset 88.7%
```

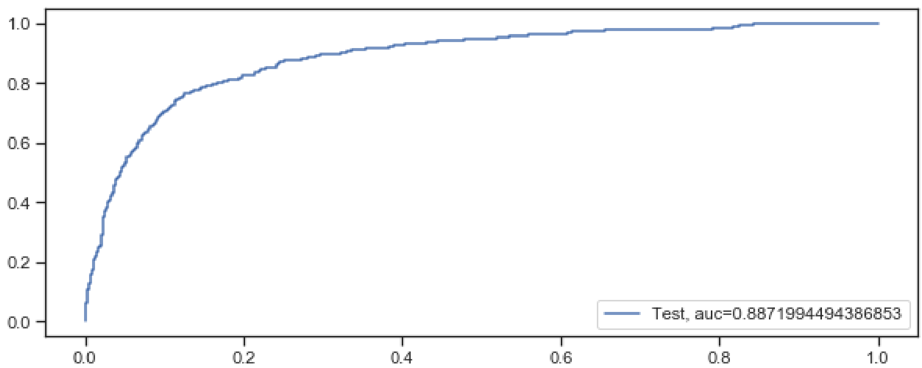


Fig. 10. Test set gradient boosting with hyper parameter.

```
In [145]: plotLiftChart(y_test.values,preds_GBC_probs_HT,"Gradient Boosting with hyper parameter tuning")
```

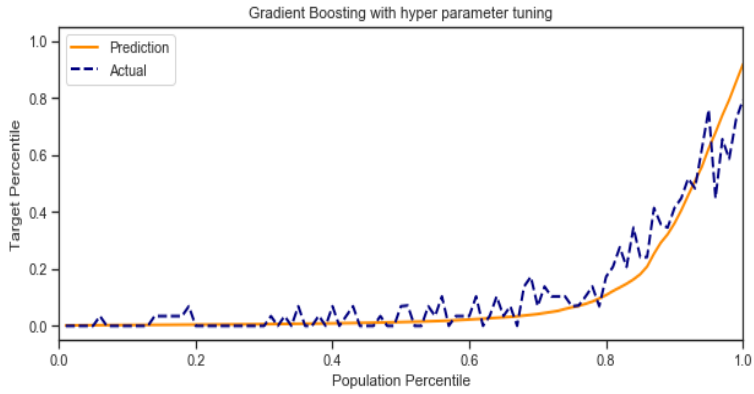


Fig. 11. Prediction gradient boosting with hyper parameter.

E. *Ada Boost Classifier*

```
Accuracy for the test dataset 91.3%
              precision    recall  f1-score   support

           0       0.93      0.97      0.95      2546
           1       0.69      0.50      0.58       347

   micro avg       0.91      0.91      0.91      2893
   macro avg       0.81      0.74      0.77      2893
weighted avg       0.90      0.91      0.91      2893

ROC for the test dataset 73.5%
```

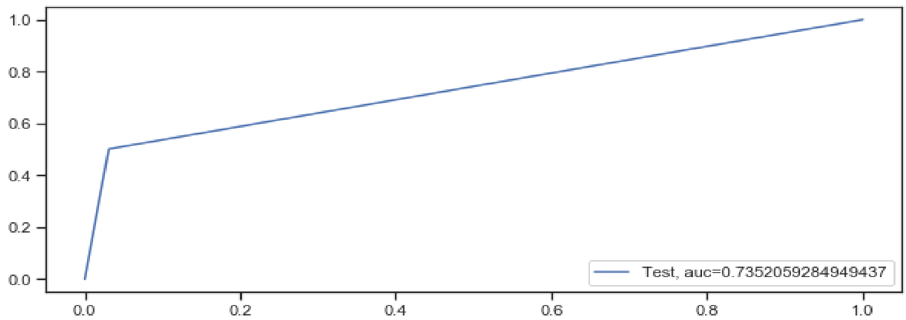


Fig. 11. Test set ada boost classifier.

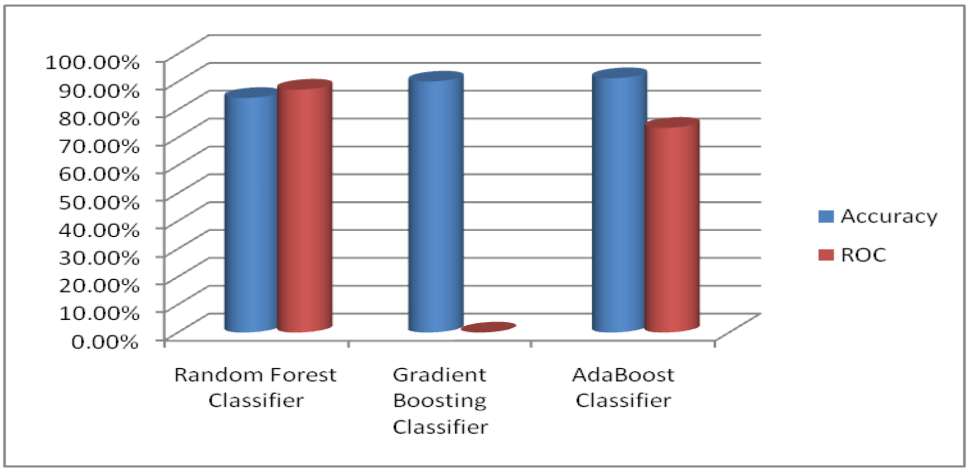


Chart 1. Compression different accuracy.

The graph above the accuracy and ROC of Random Forest Classifier, Gradient Boosting Classifier and AdaBoost Classifier has been made from the compression different accuracy graph.

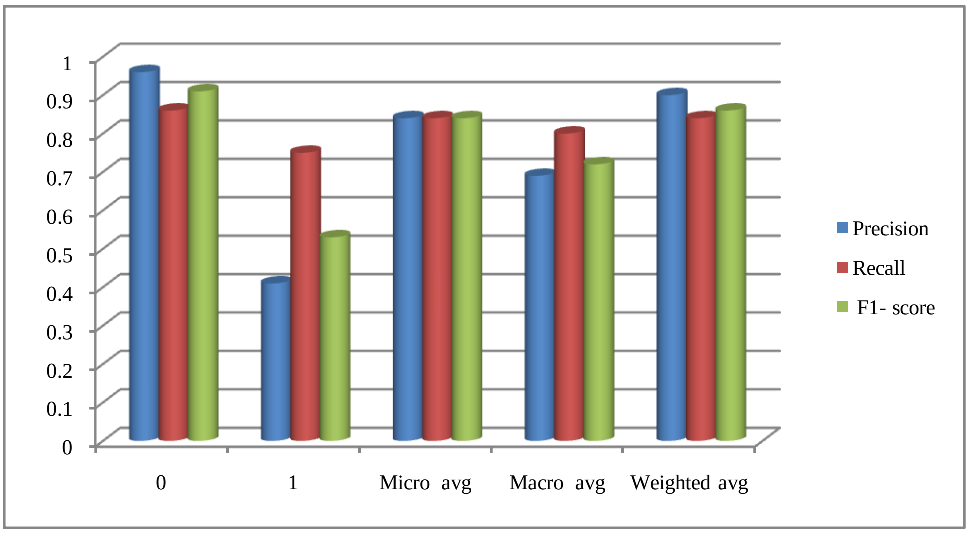


Chart 2. Random Forest Classifier Precision and recall

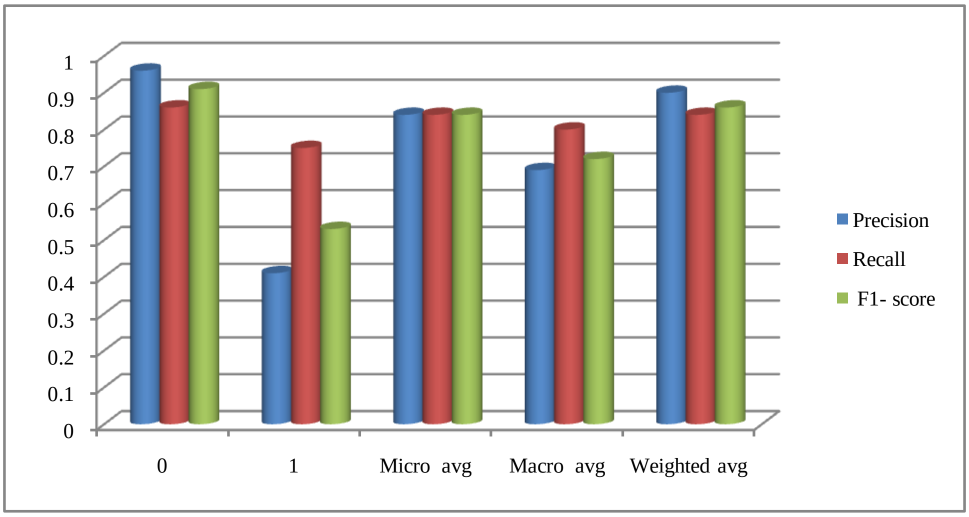


Chart 3. Gradient Boosting Classifier Precision and recall.

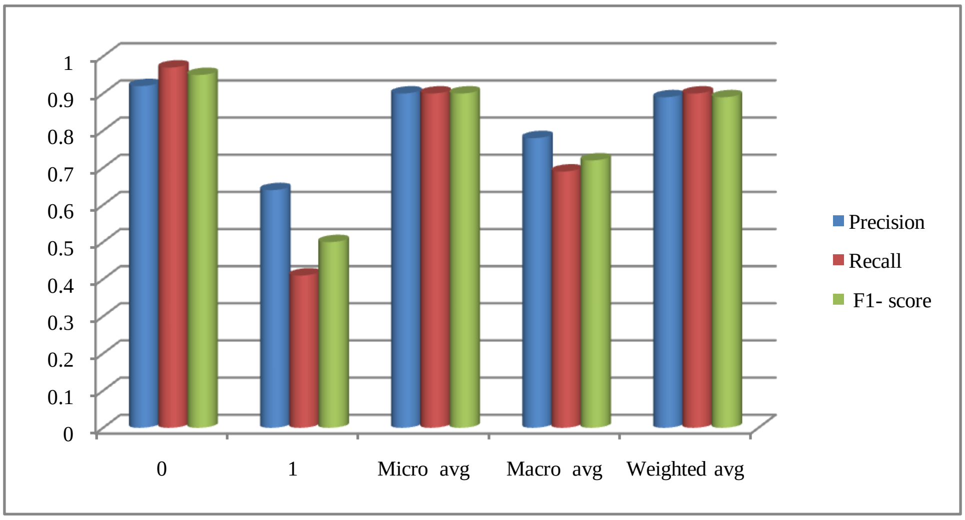


Chart 4. Gradient boosting with hyper parameter tun precision and recall.

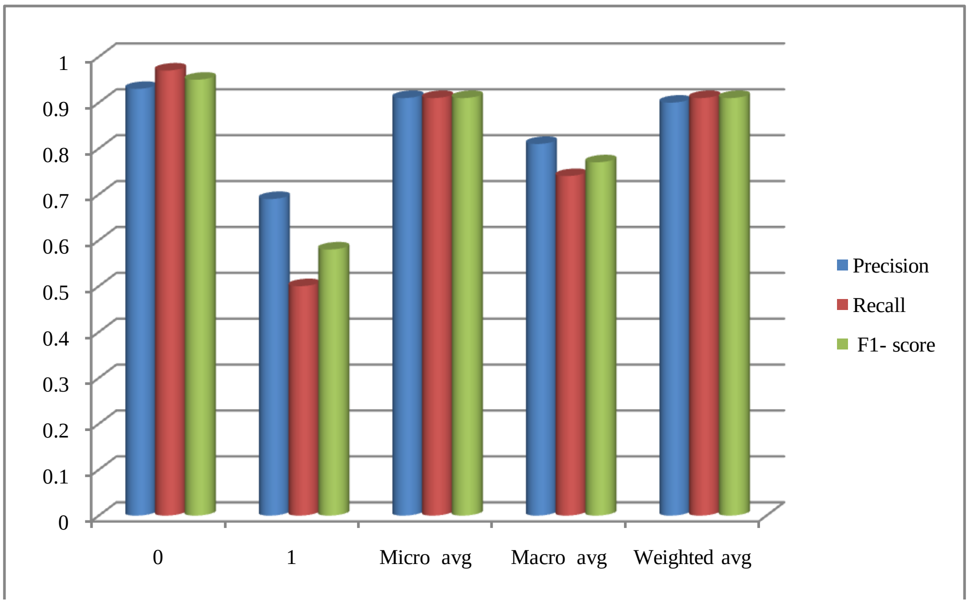


Chart 5. AdaBoost Classifier Precision and recall.

IV. Conclusion

Customer churn is always a grievous issue for the Telecom industry as customers do not hesitate to leave if they don't find what they are looking for. Customer churning is directly related to customer satisfaction. There is no standard model which addresses the issues of global telecom service providers accurately. Keeping all such things into consideration, a research thesis on customer churn prediction based on mobile data usage volumes with respect to QoE and users' perspective was studied. Statistical and Decision trees from Big Data analytics were proposed for analysis. We were provided with three sources of data. Firstly, the monthly data volumes for churned and active users by anonymous Telecom provider for a four-month period starting from June to Sep. 2014. Initially only data about the churners was given, which was later augmented with the active users. The acquired datasets from the anonymous Telecom provider could not be directly applied to the churn prediction models, here decision tree. Data preprocessing along with normalization are extremely indispensable for better comparability of usage trends between months. From the statistics of normalized volumes, confidence intervals were overlapping and close by, therefore no much significance could be noticed. Though autocorrelations were small owing to reliable confidence intervals, no strong trends could be observed. From decision tree analytics, a decision tree with just 72% accuracy could be achieved. Secondly, the results of surveyed data by accompanying thesis. Preprocessed details of a total of 770 customers with about 45 Telecom providers from various countries were tabulated. 339 customers have churned from one Telecom provider to another. 271 customers are in a plan to churn in near future. Considering the reasons from already churned users in order to predict the probable churners, similar reasons were grouped together and allocated an alphabet to carry out decision tree analysis. Without clear distinction between the reasons for problems with call and data, we get an accuracy of 41.69%. With clear distinction between their classes, like Ddata for disturbance problems with respect to data and Dcall for disturbances with respect to calls, we get an accuracy of 89.74%. Thirdly, weekly data volumes of 22 android users from three different countries along with quality, annoyance and churn risk scores for a period of eight weeks were noted. Different relationships between Quality-Annoyance, Quality-Churn risk, Annoyance-Churn risk were analyzed. Surprisingly three customers turned out to churn. Based on the decision tree analysis, 95.23%, 71.42%, 76.19% accuracies were achieved with respect to average churn risk, annoyance and quality scores respectively when dealt with churn prediction. Though the number of users being limited, these percentages are quite appreciable. Confirmed trends observed through correlations: As Quality increases, Volume increases, accordingly Annoyance and Churn risk decreases. Data preprocessing, data normalization and feature selection have shown to be prominently influential. Average Quality, Churn Risk and to some extent, Annoyance scores may point out a probable churner. The bigger the screen, higher the data consumption. Weekly data volumes with customer's recent history and necessary attributes like age, gender, tenure, bill, contract, data plan, etc., are pivotal for churn prediction.

REFERENCES

[1] Jose Saias, Luís Rato and Teresa Gonçalves "An approach to churn prediction for cloud service recommendation and user retention" 2022 Journals Information, Volume 13 Issue 5 10.3390/info13050227.

[2] IMMB Illangasinghe "Importance of customer churn prediction using machine learning and providing recommendation plans in Mobile Telecommunication Industry in Sri Lanka" JSPUI 2022.

[3] Shawni Dutta1 , Samir Kumar Bandyopadhyay "Employee attrition prediction using neural network cross validation method" International Journal of Commerce and Management Research Volume 6; Issue 3; 2020; Page No. 80-85

[4] Sadgi Sharma "A Comparative Study on Machine Learning Algorithms for Customer Churn Analytics with Power BI Dashboard" Simulation Modelling Practice and Theory, Volume 55, June 2015, Pages 1-9 2020.

[5] Nikhil Vilas Raut "A study of ensemble machine learning to improve telecommunication customer churn prediction" Information & Communications Technology 2020.

[6] Sarangi Choudhari, Nisha Balani, Parul Jha "Churn prediction using neural networks for improved accuracy" Towards Data Science 2019.

[7] Carolina Andrea Martinez Troncoso "Predicting customer churn using voice of the customer. A Text Mining Approach" Journal of Research Manchester 2019.

[8] Helder Martins "Predicting user churn on streaming services using recurrent neural networks" International Conference on Service Systems and Service Management October 19, 2017

[9] Tom Siemes " Churn prediction models tested and evaluated in the Dutch indemnity industry" International Conference on Computer Science and Information Engineering 2016.

[10] Hemanth Madabushi " Predicting the nature of service disruptions in cellular networks" Journal of the Operational Research Society 2016.

[11] Ionuț Brândușoiu, Gavril Toderean, Horia Beleiu " Methods for churn prediction in the pre-paid mobile telecommunication industry" 2016. International Conference on Advanced Computing and Communication Systems ICACCS -2017), Jan. 06 – 07, 2017, Coimbatore, INDI

[12] Kiran Dahiya, Surbhi Bhatia "Customer churn analysis in telecom industry" 978-1-4673-7231-2/15 IEEE 2015.

[13] Naren Naga Pavan Prithvi Tanneedi " Customer Churn Prediction Using Big Data Analytics" 978-1-4673-7231-2/15/$31.00 ©2015 IEEE 2015.

[14] Utku Yabas, Hakki C. Cankaya "Churn prediction in subscriber management for mobile and wireless communications services' 2013 IEEE Globecom Workshops.

AUTHOR'S PROFILE

First Author
Shivani Vaidya, M.Tech Scholar, Technocrats Institute of Technology, Bhopal, Madhya Pradesh, India.

Second Author
Rajesh Kumar Nigam, Associate Professor, Technocrats Institute of Technology, Bhopal Madhya Pradesh, India.