

Customer Churn Prediction in Telecommunication System

Mr. P. Rajasekhar Reddy

*Department of Computer Science and Engineering
Anurag Group of Institutions, Ghatkesar, Hyderabad, Telangana, India
Email- rajasekharreddycse@cvsr.ac.in*

Ch. Roshan Giri Reddy

*Department of Computer Science and Engineering
Anurag Group of Institutions, Ghatkesar, Hyderabad, Telangana, India
Email- 16h61a0510@cvsr.ac.in*

M. Sahithya

*Department of Computer Science and Engineering
Anurag Group of Institutions, Ghatkesar, Hyderabad, Telangana, India
Email- 16h61a0529@cvsr.ac.in*

R. Shiny Olivia

*Department of Computer Science and Engineering
Anurag Group of Institutions, Ghatkesar, Hyderabad, Telangana, India
Email- 16h61a0548@cvsr.ac.in*

Abstract- Churn Prediction is one of the world wide used analysis on Subscription Oriented Industries to analyze customer behaviors to predict the customers which are about to leave the service agreement from a company. We are using churn prediction model that uses classification to identify the churn customers. It is based on Machine Learning methods and algorithms and become so important for companies in today's commercial conditions as gaining a new customer's cost is more than retaining the existing ones. It is evaluated using metrics, such as accuracy, precision, recall, f-measure, and receiving operating characteristics (ROC) area. We are producing results using machine learning techniques like Random Forest, Decision Tree, Logistic Regression, XGBoost, Adaboost algorithms; out of all we got highest results in XGBoost. Churn Analysis on Telecommunication Industry in literature helps to present general information to readers about the frequently used data mining methods used, results and performance of the methods and shedding a light to further studies. To keep the review up to date, studies published in last five years and mainly last two years have been included.

Keywords – Churn prediction, XGBoost

I. INTRODUCTION

Customer churn is a considerable concern in service sectors with highly competitive services. On the other hand, predicting the customers who are likely to leave the company will represent potentially large additional revenue source if it is done in the early phase. The telecommunications sector has become one of the main industries in developed countries. The technical progress and the increasing number of operators raised the level of competition. Companies are working hard to survive in this competitive market depending on multiple strategies. Companies are working hard to survive in this competitive market depending on multiple strategies. Three main strategies have been proposed to generate more revenues: (1) acquire new customers (2) up sell the existing customers and (3) increase the retention period of customers. In a business environment, the term, customer attrition simply refers to the customers leaving one business service to another. Customer churn or subscriber churn is also similar to attrition, which is the process of customers switching from one service provider to another anonymously. From a machine learning perspective, churn prediction is a supervised (i.e. labelled) problem defined as follows: Given a predefined forecast horizon, the goal is to predict the future churners over that horizon, given the data associated with each subscriber in the network. Churn Prediction is a phenomenon which is used to identify the possible churners in

advance before they leave the network. This helps the CRM department to prevent subscribers who are likely to churn in future by taking the required retention policies to attract the likely churners and to retain them. Thereby, the potential loss of the company could be avoided. A limitation of current research is that other studies have focussed almost exclusively on churn capture, neglecting the issue of misclassification of non-churn as churn. A further limitation of current research is that it is usually based on a single output in the form of 0 for non-churn and 1 for churn. This has been recognized as a limitation because it restricts analysis possibilities. In order to address the problems mentioned above, a profile-based analysis methodology is identified as a possible solution. It is anticipated that profile-based analysis will enable future prediction, through the ability to match customers to profile clusters that are identified as most suitable for capturing future churn. It is also anticipated that profile analysis will provide a method for controlling misclassification levels through eliminating the profile clusters that statistically hold the smallest future churn capture accuracy.

II. PROPOSED WORK

Our proposed system is customer churn prediction in telecommunication sector using XGBoost Algorithm that helps the telecom sector to predict the customers who are likely to churn and to improve their services to retain the customers. Reviewing the relevant studies about churn analysis on telecom industry presented in the last five years, particularly in the last two years, it was not possible to extract the useful information hidden in the datasets, unless they are processed properly. In order to find out the hidden data, various analysis should be performed using data mining, which consists of numerous methods. Unlike, the existing system, to overcome this cons, the proposed churn prediction model has being evaluated using metrics, such as accuracy, precision, Recall, F- measure and receiving operating characteristics (ROC) area. The results reveal that our proposed churn prediction model produced better churn classification. For classification purpose we used multiple machine learning algorithms like Decision tree, Random Forest algorithm, Linear regression (LR), Support Vector Machine(SVM), XGBoost, Adaboost classifiers and the we compared the results with highest accuracy classifier. The proposed work of predicting the customer churn in advance will help the telecom industry and the CRM department to identify which customer is going to leave the network.

To classify each subscriber as potential churning or potential non churning, we have used the framework know as KDD (Knowledge Discovery Data) process. It includes five modules. First module includes the Data Collection. In our project we collected the data from Kaggle website given by IBM Community. It is used to analyze the marketing tendency of customers. We represent churn in the data set in the form of bar graph and pie chart.

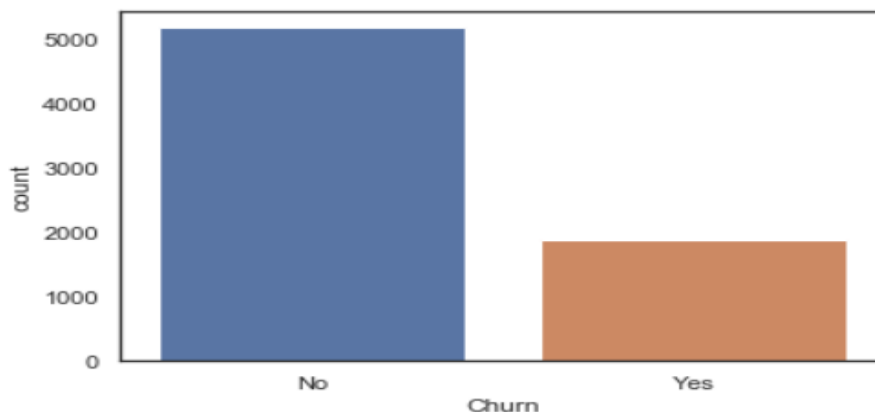


Fig1 Bar graph Representation of Churn Analysis

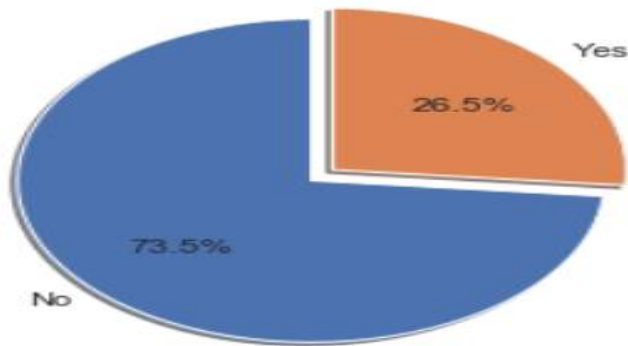


Fig2 Pie chart Representation of Churn Analysis

In order to analyze the churn first we need to know the factors which are affecting the churn analysis. These are the factors or fields that are taken into consideration. They are specified here below with their unique values. Among these we consider one of the factor tenure groups wherein we divide the data into yearly basis to analyze year by year the churn rate. This analysis is also represented in the form of bar graph for us to understand in pictorial manner.

```

SeniorCitizen : [0 1]
Partner : ['Yes' 'No']
Dependents : ['No' 'Yes']
tenure : [ 1 34  2 45  8 22 10 28 62 13 16 58 49 25 69 52 71 21 12 30 47 72 17 27
          5 46 11 70 63 43 15 60 18 66  9  3 31 50 64 56  7 42 35 48 29 65 38 68
          32 55 37 36 41  6  4 33 67 23 57 61 14 20 53 40 59 24 44 19 54 51 26 39]
PhoneService : ['No' 'Yes']
MultipleLines : ['No phone service' 'No' 'Yes']
InternetService : ['DSL' 'Fiber optic' 'No']
OnlineSecurity : ['No' 'Yes' 'No internet service']
OnlineBackup : ['Yes' 'No' 'No internet service']
DeviceProtection : ['No' 'Yes' 'No internet service']
TechSupport : ['No' 'Yes' 'No internet service']
StreamingTV : ['No' 'Yes' 'No internet service']
StreamingMovies : ['No' 'Yes' 'No internet service']
Contract : ['Month-to-month' 'One year' 'Two year']
PaperlessBilling : ['Yes' 'No']
PaymentMethod : ['Electronic check' 'Mailed check' 'Bank transfer (automatic)'
                 'Credit card (automatic)']
    
```

Fig3. Fields with unique values

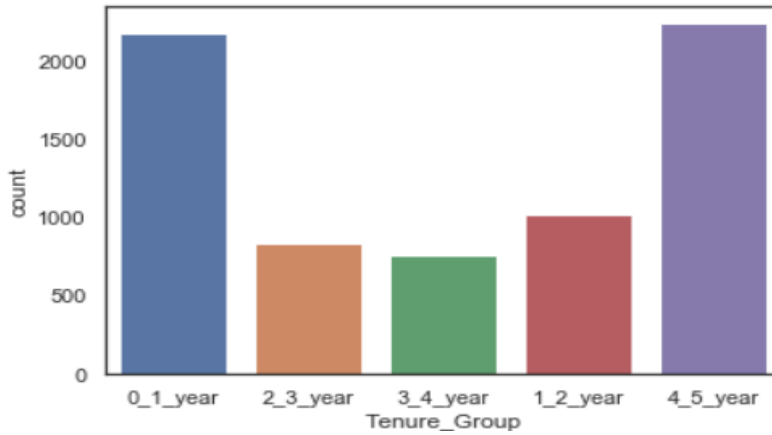


Fig4. Bar Tenure Churn Analysis

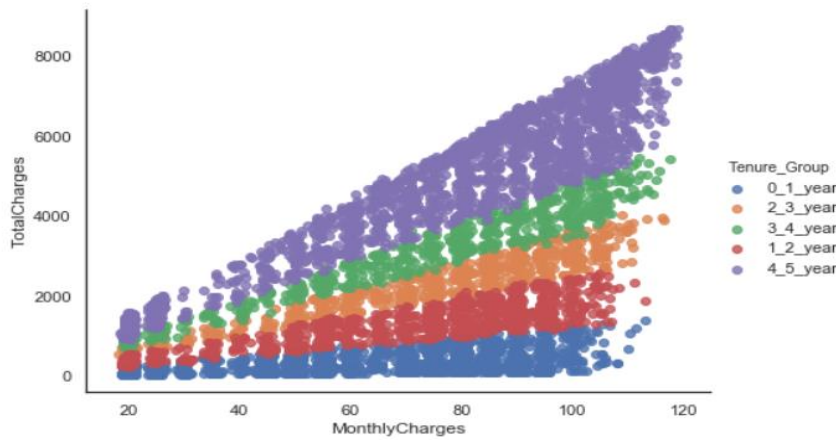


Fig5. Scatter plot Tenure Analysis

The second module is data pre-processing. The collected data consists of ambiguities, errors, redundancy which needs to be cleaned as the complete data collected is not suitable for modeling purposes. Fields with huge null values are discarded. The result we get after data preprocessing is represented with the help of heap map.



Fig6. Analysis of three factors which can help to predict churn

The third module is the Feature Extraction. The attributes are identified for classifying process. In our work, we have worked with numerical and categorical values. The fourth module is the evaluation model. We evaluate the accuracy, precision, recall, f-score. Out of all we have analyzed say that XGBoost gives more accuracy.

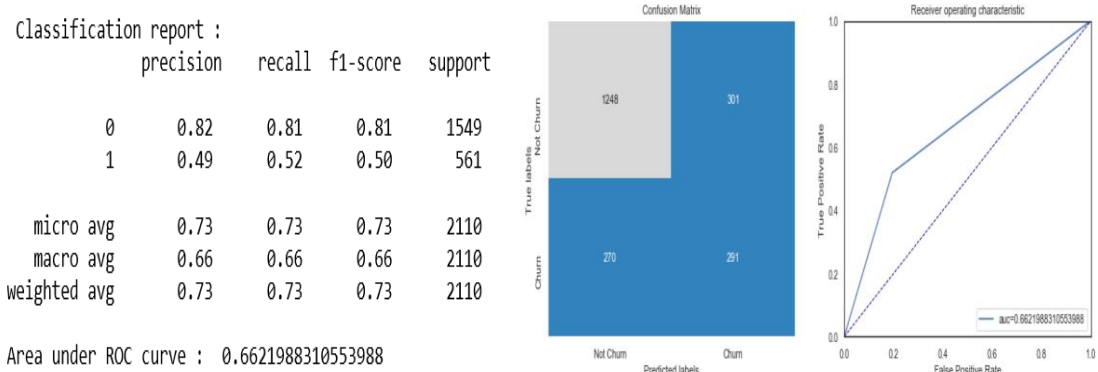


Fig7. Decision Tree Algorithm Analysis

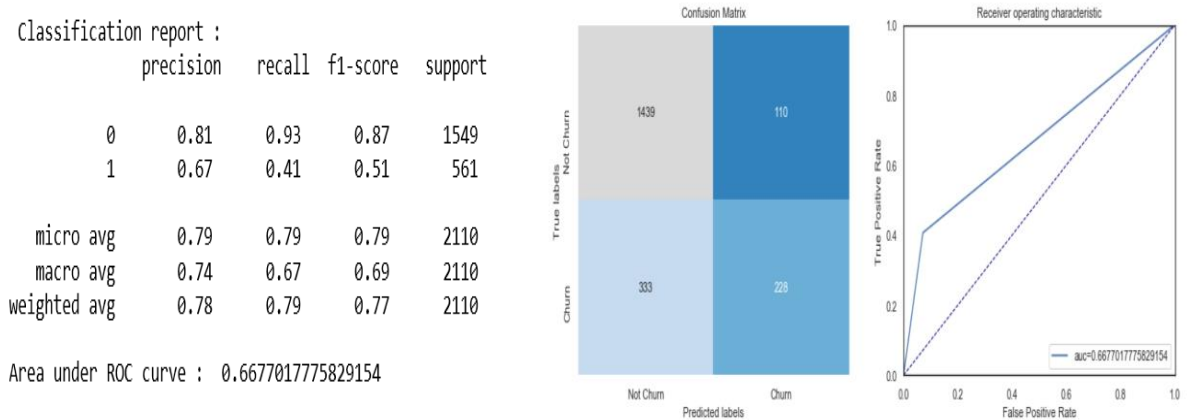


Fig8. SVM Algorithm Analysis

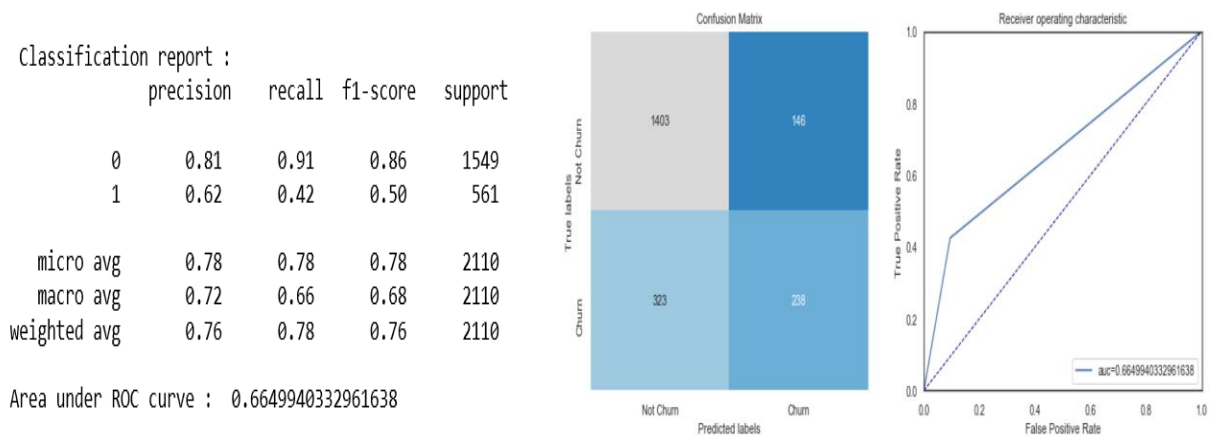


Fig9. Random Forest Algorithm Analysis

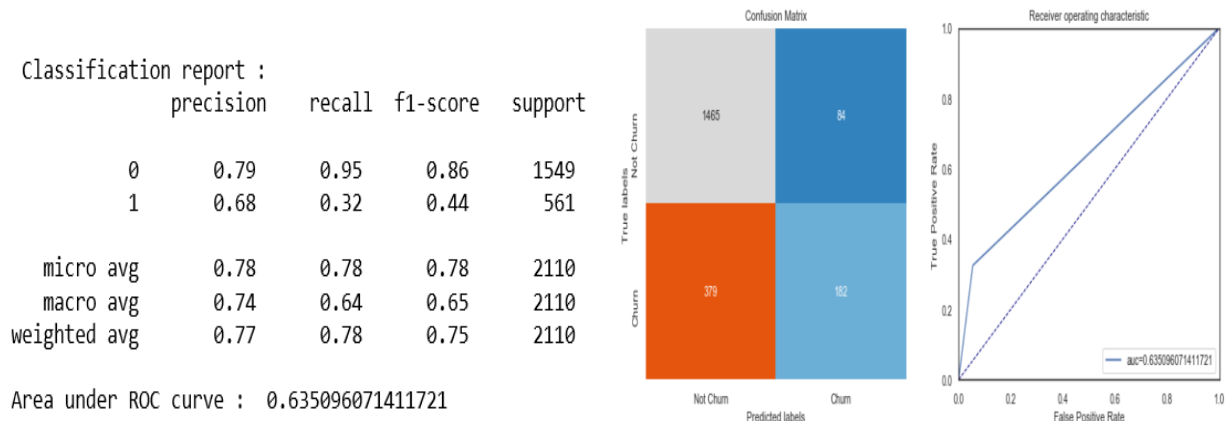


Fig10. Adaboost Algorithm Analysis

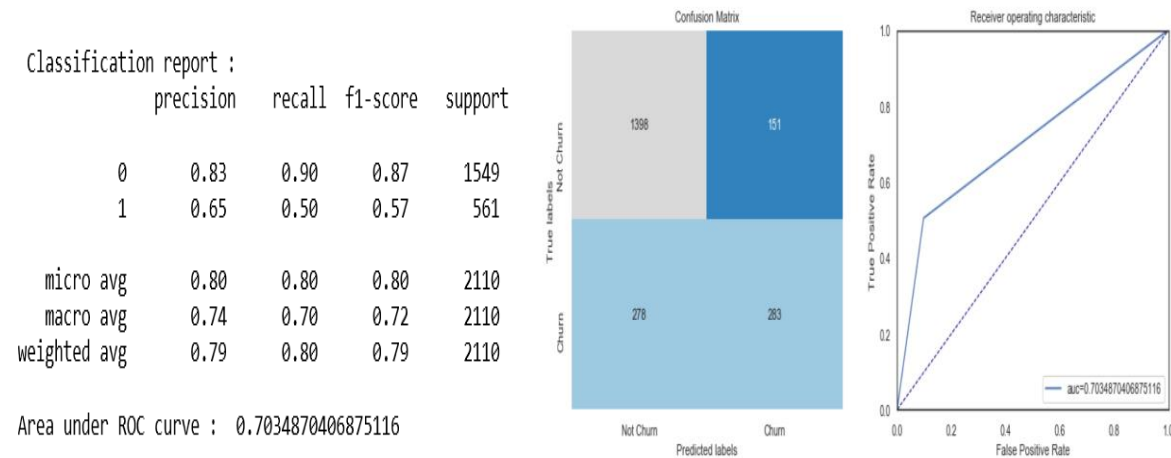


Fig11. XGBoost Algorithm Analysis

The fifth and the final module is to predict the result to be churn or non churn customers. We developed a desktop web application using flask. It takes the input values from the user and XGBoost is the algorithm used in this application. The values should be either in 0 or in 1 in some fields and other values can be given for some other fields they are described below. For the fields like senior citizen, phone service, online service, online backup, Tech Support it takes yes as 1 and no as 0. Gender field takes male as 1 and female as 0. Internet Service takes fiber optics as 1 and DSL as 0. Contract takes yearly subscription as 1 and a monthly subscription as 0. Payment Mode takes values credit card as 0, Bank as 1, electronic mode as 2 and mailed check as 3. The result is displayed whether churn or not churn customers.

III. EXPERIMENT AND RESULT

The desktop web application takes the user details; the user has to enter the details. The output is given either churn or not churn. Churn depicts that customer is thinking to leave the subscription. The company must take certain measures by giving the customer offers to retain in subscription. Not Churn depicts that customer will not leave the company. Both cases examples have been shown below.

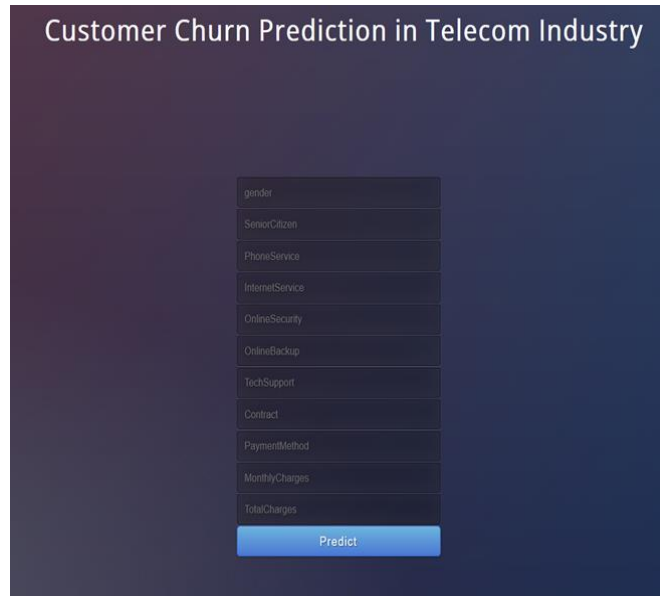
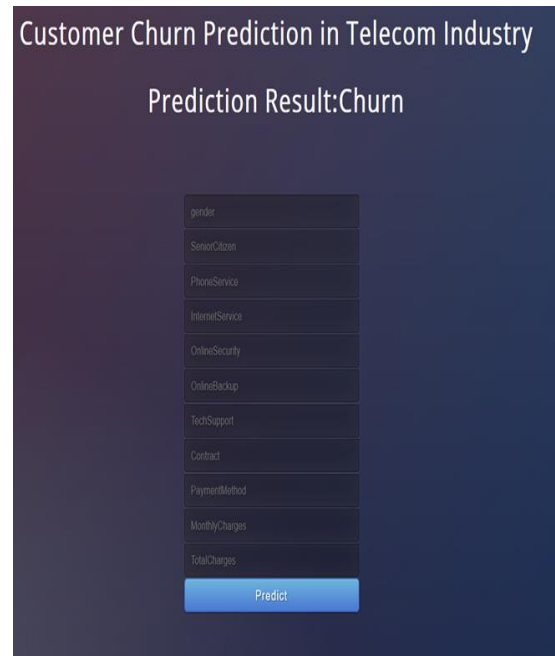
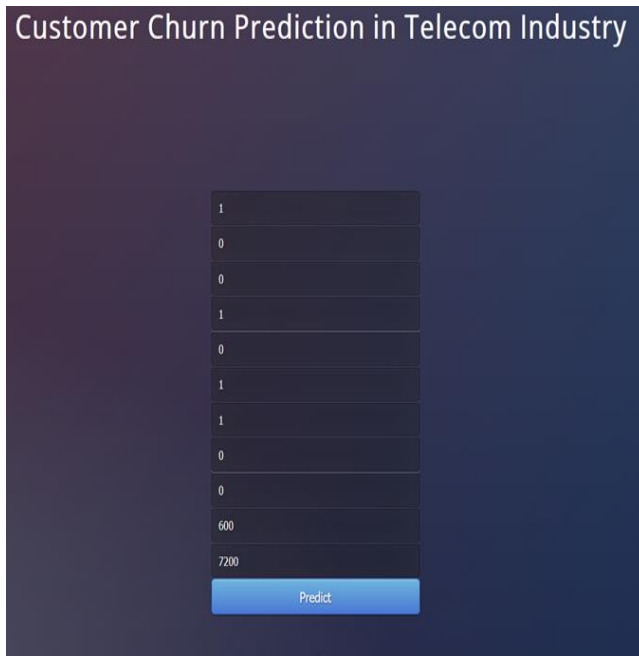


Fig12 Desktop web application for churn prediction

CASE 1: Churn Customer



CASE 2: Non-Churn customer

Customer Churn Prediction in Telecom Industry

1
0
1
0
1
1
0
0
3
199
2380

Predict

Customer Churn Prediction in Telecom Industry

Prediction Result: Not Churn

gender
SeniorCitizen
PhoneService
InternetService
OnlineSecurity
OnlineBackup
TechSupport
Contract
PaymentMethod
MonthlyCharges
TotalCharges

Predict

IV. CONCLUSION

In the present competitive market of telecom domain, churn prediction is a significant issue of the CRM (Customer relation management) to retain valuable customers by identifying similar groups of customers and providing competitive offers/services to the respective groups. Therefore, in this domain, the researchers have been looking at the key factors of churn to retain customers and solve the problems of CRM. In this study, a customer churn model is provided for data analytics and validated through standard evaluation metrics. The obtained results show that our proposed churn model performed better by using several machine learning techniques.

REFERENCES

- [1] K.B.Oseman, S.B.M.Shukor, N.A.Haris, F. Bakar, Data Mining in Churn Analysis Model for Telecommunication Industry, Journal of Statistical Modeling and Analytics, Vol. 1 No. 19-27, 2010.
- [2] T. Mutanen, Customer churn analysis – a case study, Technical Report, retrieved http://www.vtt.fi/inf/julkaisut/muut/2006/customer_churn_case_study.pdf, April 12, 2014
- [3] Customer Churn Prediction and Prevention by the Optimove <https://www.optimove.com/learning-center/customer-churn-prediction-and-prevention>
- [4] S.V.Nath, Customer Churn Analysis in the Wireless Industry: A Data Mining Approach, Technical Report has been taken <http://download.oracle.com/owsf/2003/40332.pdf>, April 14, 2014.
- [5] Understanding the confusion matrix by Sarang Narkede <https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62>
- [6] <https://datascienceplus.com/predict-customer-churn-logistic-regression-decision-tree-and-random-forest>
- [7] A Gentle introduction to XGBoost for applied Machine Learning by machine learning mastery in given website <https://machinelearningmastery.com/gentle-introduction-xgboost-applied-machine-learning>