



## Telecom Customers Churn Modeling

Dr.N.DIVYA

Dr.N.SREENIVAS

SEELAM VARUN REDDY

MADI BHARATH REDDY

B.AKASH KUMAR

**Abstract**—Customer churn is one of the main problems in the telecommunications industry. Several studies have shown that attracting new customers is much more expensive than retaining existing ones. Therefore, companies are focusing on developing accurate and reliable predictive models to identify potential customers that will churn in the near future. The telecommunication industry in the recent years is a subject of major changes and from a fast-growing industry has come to a state of saturation accompanied with strong competitive market. Customers starve for better services and prices, while their requirements are extremely complex and difficult to understand. Customers starve for better services and prices, while their requirements are extremely complex and difficult to understand. The aim of this project is investigating the main reasons for churn in telecommunication sector. The aim of this project is investigating the main reasons for churn in telecommunication sector. The proposed methodology for analysis of churn prediction covers several phases: understanding the business; selection, analysis and data processing; implementing various algorithms for classification; evaluation of the classifiers and choosing the best one for prediction

**Keywords** — Churn Prediction, machine learning.

### I. INTRODUCTION

Accreation of new clients are one of the most key concerns of businesses. While many companies concentrate on obtaining new customers, intelligent ones try to focus on retaining the existing ones in order to provide themselves with the opportunity to grow the company. One of the most appreciable ways of increasing customer base for a company is to concentrate on both acquisition and retention. In the new age emergence of electronic commerce has increased the available information, and the internet has empowered the customers who are no longer stuck with the decisions of a single company and has led to amplifying in the competition, while competitors are only one step away, customer empowerment is likely to aggravate the weakening the rate of a company's customers. Facing this problems companies should be prepared with the most efficient and effective methods of understanding their client's behavior by making predictions on their possible future. The study aims at finding an efficient and accurate predictive model

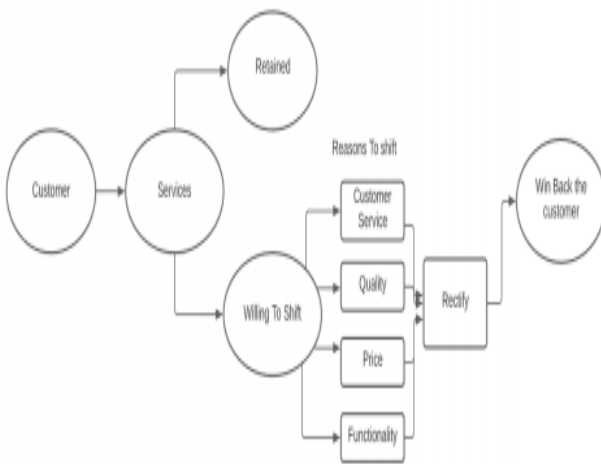
for customer churn in mobile telephony market segment by utilizing machine learning techniques. With the purpose of making you more accustomed with the research's area and its significance. In this report we provide you with statistics regarding the customer churn exaggeration in telecommunications industry and also we discuss our problem definition and the purpose of our research. Customer churn is the main concern of most developed companies which are highly active in industries. The statistics display that the annual churn rate in telecommunication industry is approximately 36%. This means wasting the money and efforts for acquiring new customers when the existing customers are leaving. In order to address this issue we must understand the churners before they churn, so by developing a model which predicts the customers who are future churners seems to be the best solution for this problem. The model which we prepared has ability to recognize the customers who are going to churn in the near future. But, due to the nature of mobile telephony market which is a democratic kind of industry where customer has privilege to leave the network if he/she had issues with it. Customer churn is not easily traceable, thus building a predictive model would be of high complexity. In order to achieve such a goal in telecom market segment the initial step that need to be taken is to be define the difference between churner and a non-churner and then predicting the churner. Following above, Due to the less percentage of churn rate in the company's data set compared t the non-churners it is complex to understand the patterns for a machine. It can also plays a vital role in defining the accuracy of the prediction model. But, when the data is dumped more and more in a daily basis the prediction model get the better understanding about the prediction. And in future it improves the accuracy and makes better predictions.

Customer churn prediction has been performed using various techniques, including data mining, machine learning, and hybrid technologies. These techniques enable and support companies in identifying, predicting, and retaining churn customers. They also help industries in CRM and decision making. Most of them used decision trees in common as it is one of the recognized methods to find out the customer churn, but it is not appropriate for complex problems. In some cases, data mining algorithms are used for customer prediction and historical analysis. The techniques of regression trees were discussed with other commonly used data mining methods like decision

trees. Using the decision trees has their own drawbacks if the data is increasing then the model's accuracy gets decreased because the decision tree algorithms doesn't perform well if the data set size is large. To address this problem that we implemented a classifier machine learning algorithm "Ada Boost" classifier, Better suited for large data set's. Due to its learning nature from the error's it makes it improves its accuracy for every new prediction and new data addition

**II. LITERATURE SURVEY**

Literature review mainly consist of three individual parts. section 1 Finding the Customer Relationship with the company which is called as (CRM) this give a better understanding about the customer from the time he enters as new customer to the end time which means whether the customer is going to retain the same network or whether he had plans to shift the network. Section 2 is about the data mining process of customers data to get patterns that are required for understanding the reasons for customers leaving the network. Section 3 machine learning models used for classification of customers to two classes who are going to churn and who are not going to churn. Although the focus of this research is on machine learning predictive models for customer churn, this chapter has taken a look at churn literature from both explanatory and predictive point of view in order to broaden the visions toward all sides of churn issue.



The CRM main aim is to define the flow of the customer from his onboard to the network and to the stage where he is going to churn or not. It includes providing the services to the customers and finding the reasons for his churn to the other network. Bose (2002) viewed CRM as an integration of technologies and business processes used to satisfy the needs of a customer during any given interaction more specifically from his point of view Customer relationship management (CRM) involves acquisition, analysis and use of knowledge about customers in order to sell goods or services and to do it more efficiently. Richards and Jones (2008) have defined CRM as "a set of business activities supported by both technology and processes that is directed by strategy and is designed to improve business performance in an area of customer management". Data mining is useful for extraction of information from data set which will be useful for making business decisions. It can be done by using various advanced algorithms and advanced tools used for Making data mining processes.

Classification algorithms are used when the desired output is a discrete label. In other words, they're helpful when the answer to your question about your business falls under a

finite set of possible outcomes. Many use cases, such as determining whether an email is spam or not, have only two possible outcomes. Classification can be performed on structured or unstructured data. Classification is a technique where we categorize data into a given number of classes. The main goal of a classification problem is to identify the category/class to which a new data will fall under.

**III. PROPOSED SYSTEM**

**A. Architecture phases**

The architecture design of this modeling starts with collection of the data of telecom customers. Since every company holds their data for making better business decision making. So we need to extract the data and need to use data preprocessing tools which includes removing the abnormalities in the dataset and converting categorical features to numerical features etc. Later the next step is feature selection it may be increasing the dimensions of dataset or maybe decreasing. This process is necessary because we need to select features which are useful to the model and remove the features which doesn't have any impact on the target which saves the lot of computational power. After preprocessing the data we need to split it for making machine train and test to get accuracy. For this, we use machine learning algorithms.

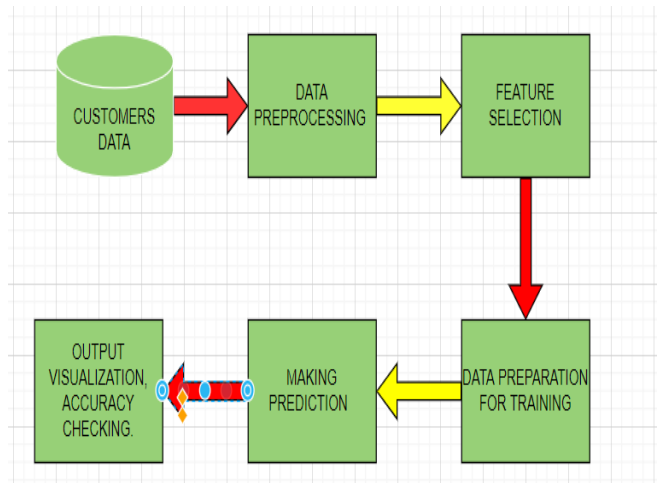


Fig-1

**B. EXPLORATORY DATA ANALYSIS(EDA) :**

After data collection, several steps are carried out to explore the data. Goal of this step is to get an understanding of the data structure, conduct initial preprocessing, clean the data, identify patterns and inconsistencies in the data (i.e. skewness, outliers, missing values) and build and validate hypotheses. By inspecting the columns and their unique values, a general understanding about the features can be build.

**C. PROPOSED ALGORITHM:**

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.

Step 1: Train the AdaBoost classification algorithm with preprocessed data.

Step2: Check in terms of accuracy of the model by using test data.

Step3: Classify the output for validation set and check the accuracy before putting it into real use.

**D. MODEL EVALUATION METRICS:**

For performance assessment of the chosen models, various metrics are used:

Feature weights: Indicates the top features used by the model to generate the predictions

Confusion matrix: Shows a grid of true and false predictions compared to the actual values

Accuracy score: Shows the overall accuracy of the model for training set and test set

**IV. DESCRIPTIVE DATA STATISTICS**

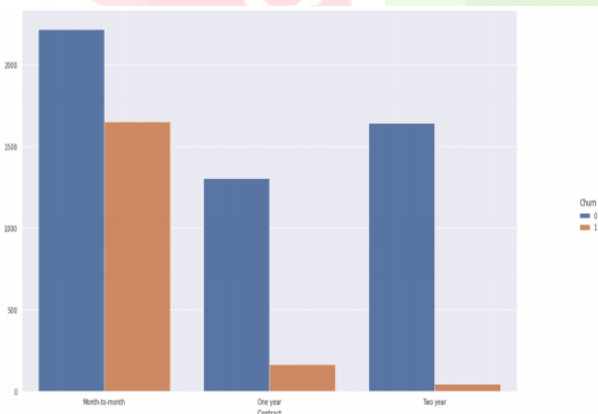
**A. CORRELATION OF FEATURES WITH CHURN:**

“Correlation” is a statistical term describing the degree to which two variables move in coordination with one-another.

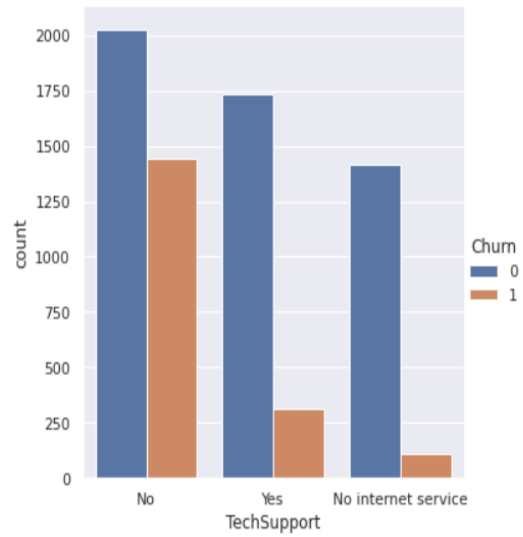
1. Positive Correlation: means that if feature A increases then feature B also increases or if feature A decreases then feature B also decreases. Both features move in tandem and they have a linear relationship.
2. No Correlation: No relationship between those two attributes.
3. Negative Correlation: means that if feature A increases then feature B decreases or if feature A decreases then feature B increases. Both features move in opposite directions and they have a non-linear relationship.

**B. FEATURES WITH HIGH RELATION TO CHURN:**

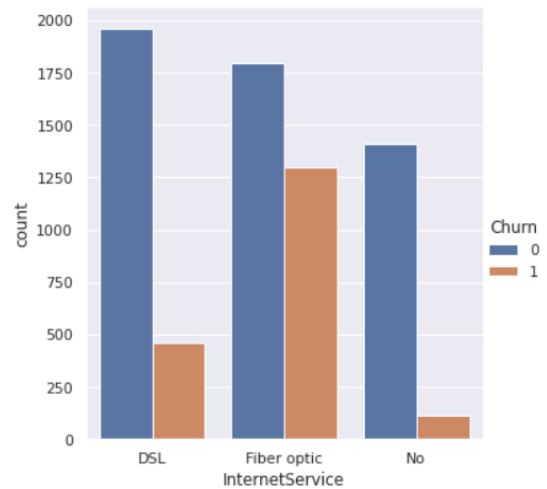
1. Type Of Contract:



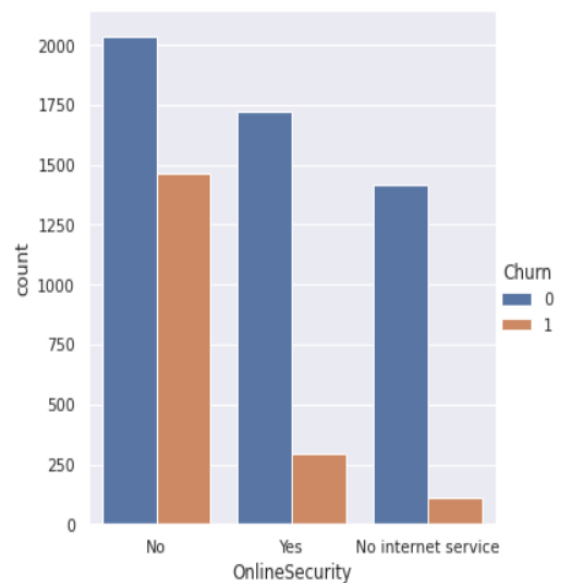
2. Tech Support:-



3. Type of Internet Service:-



4. Online Security:-



IV. RESULTS

V. CONCLUSION

The project aims to predict whether the customer is going to move to another network in the future by using the customer's data. There are two classes of classification that can be done by using the Ada Boost classification algorithm.

Class 1:- It represents the churn value as 1 which implies the customer is going to shift to another network. Our model also provides the reasons of customer shifting the network by comparing his data values with the existing non-churner customer's data features and establishes correlation with those records and Identifies the week regions and produces them as reasons for customers leaving.

Class 2: It represents the churn value as 0 which means the customer is not willing to shift to another network in meantime.

**Telecom Churn Modeling**

Gender :  Male  Female  
 SeniorCitizen  0  1  
 Partner  No  Yes  
 Dependents  No  Yes  
 tenure 23  
 PhoneService  No  Yes  
 MultipleLines  No  Yes  No phone service  
 InternetService  Fiber optic  DSL  No  
 OnlineSecurity  No  Yes  No internet service  
 OnlineBackup  No  Yes  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  Yes  No internet service  
 PaperlessBilling  No  Yes  
 PaymentMethod  Electronic check  Mailed check  Bank transfer (automatic)  Credit card (automatic)  
 MonthlyCharges 130  
 TotalCharges 1800  
 Predict

1. THE CUSTOMER IS GOING TO LEAVE THE COMPANY
2. Low tenure can be a reason--> fault
3. Have problems in Internet Service due to Fiberoptic--> fault
4. online security is good-->support
5. online backup is good-->support
6. contract is good-->support
7. Technical support is good-->supportive

**Telecom Churn Modeling**

Gender :  Male  Female  
 SeniorCitizen  0  1  
 Partner  No  Yes  
 Dependents  No  Yes  
 tenure 60  
 PhoneService  No  Yes  
 MultipleLines  No  Yes  No phone service  
 InternetService  Fiber optic  DSL  No  
 OnlineSecurity  No  Yes  No internet service  
 OnlineBackup  No  Yes  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  Two year  One year  
 PaperlessBilling  No  Yes  
 PaymentMethod  Electronic check  Mailed check  Bank transfer (automatic)  Credit card (automatic)  
 MonthlyCharges 190  
 TotalCharges 2100  
 Predict

THE CUSTOMER IS NOT LEAVING THE COMPANY

This model helps in Predicting customers overall satisfaction as well as their experience with service quality. Identifying potential network issues, competitive threats, and at-risk customers. Identifying the negative customer experience trends and reducing attrition levels. Building a robust predictive model and gathering data.

In the competitive telecom sector standardization and public policies of mobile communication permits customers to switch over from one carrier to another carrier easily resulting in a competitive market. The prediction of churn or the task of recognizing customers who are probable to discontinue service use is a lucrative and essential issue of the telecom sector. Customer churn is often a critical problem for the telecom sector as customers do not delay to leave if they do not predict what they are viewing for. Customers mainly need value for money, competitive cost and greater service quality. Customer churning is associated directly with customer satisfaction. It is a known fact that the customer acquisition cost is larger than customer retention cost that makes the retention a difficult prototype of business.

The factors influencing customers switching behaviour would be the quality of call, satisfaction level, level of tariff, image of brand, handsets, tenure and income. Some customers are much cost sensitive and move to other telecom service provider if they get better cost and also, they chose the service provider chosen by their friends and family. Another factors which impact a customer to select a service provider is cost and communication followed by responsiveness to their service complaints. Quality is also one of the factors which impact customers to move from one service provide to another service provider.

The cost of obtaining new customers can be greater than that of customer retention. One of the best way for customer retention is to reduce customers churn rate where churn refers to migration of customer from one service provider to another service provider or terminating particular services overparticular periods for several reasons that can be predicted previously if the firm examines its records of data and uses machine learning technique which enhances the firms to find customers who are probable to churn. Several algorithms are available to reduce the churn rate in telecom companies. The telecommunication service providers use advanced analytics algorithms to mine through huge number of data of customers. This algorithm is smart enough to recognize hidden characteristics to find which customers are much probable to churn. Data mining plays an essential role in telecom firms and their effort to reduce overall churn develops good marketing strategies, recognize fraudulent activities and consumers and manage their network better. A proper algorithm is chosen relying on the problem nature and that of feasible data. It can be concluded that machine learning algorithms is regarded as one of the best solutions for the telecom sector to reduce the churn rate.

## REFERENCES

- [1] Ahmad, A.M., Jafar, A, Aljoumaa, K. (2019). Customer churn prediction in telecom using machine learning in big data platform. Journal of Big Data.
- [2] Ahna, J.H., Hana, S.P., Lee, Y.S. (2006). Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry".
- [3] Asthana, P. (2018). A comparison of machine learning techniques for customer churn prediction. International Journal of Pure and Applied Mathematics.
- [4] Azeem, M, Usman, M. (2018). A fuzzy based churn prediction and retention model for prepaid customers in telecom industry. International Journal of Computational Intelligence Systems.
- [5] Brandusoiu I, Todorean, G, Ha, B. (2016). Methods for churn prediction in the prepaid mobile telecommunications industry. In: International conference on communications.
- [6] Adebisi, S.O., Oyatoye, E.O., Amole, B.B. (2016). Relevant Drivers for Customers` Churn and Retention Decision in the Nigerian Mobile Telecommunication Industry. Journal of Competitiveness communications.
- [7] Almana, A.M., Aksoy, M.S., Alzahrani, R. (2014). A Survey on Data Mining Techniques in Customer Churn Analysis for Telecom Industry. Int. Journal of engineering Research and Applications.
- [8] Adebisi, S.O., Oyatoye, E.O., Amole, B.B. (2016). Relevant Drivers for Customers` Churn and Retention Decision in the Nigerian Mobile Telecommunication Industry. Journal of Competitiveness.
- [9] Alwin, P.K.D.N.M., Kumara, B.T.G.S., Hapuarachchi, H.A.C.S. (2018). Customer Churn Analysis and Prediction in Telecommunication for Decision Making. 2018 International Conference on Business Innovation (ICOBI).

