Customer Churn Prediction

[1] Mittal Patil, [2] Mahima Rawal, [3] Mukta Sawant, [4] Shradhha Rajput, [5] Prof. Abhay Gaidhani

[1,2,3,4] Student, [5] Professor, [1,2,3,4,5] Computer Engineering,

Sandip Institute of Technology and Research Centre, Nashik (422213), India

Abstract:

Customers, additionally taken into consideration as the primary income source, are the maximum important belongings in any industry. The telecom enterprise has been increasingly aggressive in latest years. The fee of attracting new clients is greater than preserving current telecom clients. Telecom organization desires to apprehend client churn via client dating management (CRM). To are expecting which clients will churn CRM analysers are required. Customer churn is a huge problem and one of the maximum critical issues for massive corporations. Because of the direct effect on organization earnings, especially withinside the telecommunications industry, groups are searching to create fashions to research patron churn. Thus, figuring out the elements that make contributions to client churn is crucial if you want to make the right movements to lessen this churn. As existence appears to be steadily depending on using telecom merchandise and services, the worldwide market for telecommunication is slated to develop at a splendid price over the imminent years. The international telecommunication marketplace is constantly reworking because of the continued improvements and tendencies taking area consistently and at a quick pace. So for the boom of the telecommunication marketplace, the loyalty of the client has turned out to be a key factor. Nowadays, there are lots of churn clients withinside the telecom industries, that's why there's a want for a website professional for predicting the churn clients. The strategies utilized in growing this version are gadget studying strategies and constructing a brand-new manner of features' engineering and selection. The version became advanced and demonstrated the usage of the Spark surroundings on a massive dataset acquired via way of means of reworking widespread quantities of uncooked statistics furnished via way of means of the telecom provider. The dataset contained all clients' statistics like Customer Id, Gender, Senior Citizen, Partner, Tenure, Phone Service, Multiple Lines, Internet Service, Online Security, online backup, Device Protection, Tech Support, Streaming TV, Streaming Movies, Contract, Paperless Billing, Payment Method, Monthly Charges, Total Charges, Churn and became used to train, test, and examine the machine at Telecom Company. We enforce churn prediction the usage of categorization.

Keywords: Machine Learning, Random Forest, Decision Tree, XGBoost, Prediction, Churn.

INTRODUCTION

In developed countries, telecommunications have become one of the most important industries. Customer attrition is a significant issue that is frequently linked to the industry's life cycle. When a business is in its growth phase, sales increase at an exponential rate, and the number of new customers vastly outnumbers the number of churners. Companies in the mature stage of their life cycle, on the other hand, focus on lowering customer churn rates. The most common reasons for customer churn are classified into two categories: accidental and intentional. Accidental churn occurs when conditions change and customers are unable to use the services in the future. Customers choose to switch to another company that provides similar services, which is known as intentional churn. This is the type of churn that most industries prefer to control. Retention is one of the major growth pillars for products with a subscription-based business model. A customer may leave after a few poor experiences – or simply one. And if a large number of dissatisfied consumers leave at a rapid pace, both economic losses and reputational harm would be tremendous. Companies in these industries frequently employ customer service divisions that strive to win back defecting consumers, because recovered long-term customers can be far more valuable to a company than newly recruited clients. Companies generally distinguish between voluntary and involuntary churn. Voluntary churn arises when a customer decides to switch to another company or service provider; involuntary churn occurs when a customer is relocated to a long-term care facility, death, or relocates to a distant region. Analysts focus on voluntary churn because it often occurs as a result of aspects in the company-customer relationship that companies control, such as how billing transactions are handled or how after-sales support is offered. Customers (of subscription-based businesses) choose a product or service for a specific period of time, which can be as short as a month. As a result, a customer remains open to new fascinating or lucrative offers. Furthermore, if a customer's current commitment expires, they have the option to reconsider and decide not to continue doing business with the company. Natural churn is unavoidable, and the percentage varies by company. A greater churn figure, on the other hand, is a result of the fact that a company is doing something wrong. These models are successful in focusing customer retention marketing operations on the fraction of the customer base that is most prone to churn because they generate a short-prioritized list of probable defectors. The purpose of this model was to investigate the primary causes of churn among fixed telephony subscribers in a telecom company.

LITERATURE REVIEW:

This section provides a summary of some of the previous work on developing ML models using various algorithms to enhance the Churn prediction process. In telecom industries, there were many methods to predict the churn. Most of them used machine learning and data mining. The majority of related work aimed at using only one machine learning method to extract knowledge, while some focused on comparing various ways to predict churn.

1)Gavril et al. proposed an advanced system in which some features include information regarding the number of incoming and outgoing messages, as well as voicemail for each customer. To minimize data dimensionality, the author used the principal component analysis algorithm "PCA." To identify the churn factor, three machine learning algorithms were used: neural networks, support vector machines, and Bayes networks. The dataset employed in this study was tiny, and there were no missing values. [1]

2) He et al. suggested a prediction model based on the Neural Network algorithm to overcome the problem of customer attrition in a large Chinese telecom company with approximately 5.23 million consumers. The overall accuracy rate reached 91.1 percent. [2]

3) Idris proposed a method for predicting the churn problem in telecoms based on genetic programming with AdaBoost. Two standard data sets were used to test the model. One by Orange Telecom and one by cell2cell, with the cell2cell dataset having 89 percent accuracy and the other having 63 percent. [3]

4) Huang et al. researched the issue of consumer churn in a big data platform. The researchers' goal was to demonstrate that big data significantly improves the process of detecting churn based on the amount, diversity, and velocity of the data. To manufacture the cracks, China's top telecoms corporation needed a big data platform to deal with data from the Operation Support department and the Business Support department. AUC was used to evaluate the Random Forest method. [4]

PROBLEM STATEMENT:

To develop an application to Prediction of Telecommunication Churn. The retention and acquisition of users are the key issues in the telecom trade. The subscriber base is increasing rapidly due to the growth of the marketplace in every business. Accordingly, companies have recognized the significance of retaining the customers is on hand. It has become necessary for service providers to reduce the churn rate of customers since inattention might negatively influence the profitability of the company. Churn prediction contributes to identifying those users who are likely to switch a company over another. Telecom has to put up with the problem of an ever-growing churn rate. Accordingly, the current study employs a machine learning algorithm. Machine learning algorithm techniques enable these telecommunications organizations to be safeguarded with effective churn-reduction strategies. Silent churn is one type that is considered complicated to predict since there might have such kinds of users who might probably churn in the near future. It should be the aim of the decision-maker and advertisers to reduce the churn ratio since it is a recognized fact that comparatively existing customers are the most beneficial resources for companies than acquiring new ones.

OBJECTIVE:
1) To investigate the prediction of customer attrition in telecommunications using machine learning.
2) To investigate the impact of customer churn in the telecom industry as a whole.
3) To discuss the significance of customer churn models in the telecom industry.
4) To compare the algorithms that are effective in reducing churn rate in telecom companies.

RESEARCH METHODOLOGY

We've used previous data for predicting future customer churn. We examine data from consumers who have already churned (respond) as well as their attributes/behavior (predictors) prior to the churn. Customers' demographic information, total charges, and the sort of service they receive from the company are all included in the dataset. It is made up of churn data from over a thousand consumers divided across 21 parameters gathered from Kaggle. We will try to predict the reaction for existing customers by fitting statistical models that relate the predictors to the response.

Algorithm:

1) Decision Tree:

Classification bushes are tree fashions wherein the centered variable can take a discrete set of values; in those tree structures, leaves suggest magnificence labels and branches constitute function conjunctions that cause the one's magnificence labels. Regression bushes are choice bushes wherein the goal variable can take non-stop values (typically actual numbers). To create a prediction, this set of rules divides an information pattern into or extra homogeneous units primarily based totally on the maximum sizeable differentiator in entering variables. The tree is created by using department of every split. As a result, a tree containing choice nodes and leaf nodes (that are connected) is formed.

2) Random Forest:

The random woodland is a category set of rules product of several choice bushes. We use Random Forest to forecast whether or not or now no longer the client will terminate his membership. Random Forest makes use of Decision bushes to are expecting whether or not a client might cancel his subscription. A choice tree specializes in one precise elegance. A elegance with the maximum votes may be the classifier for a selected client. Decision bushes are particularly touchy to the records on which they're trained. We use Bagging to keep away from this. Bagging is a way wherein we take a random pattern from a dataset to educate choice bushes.

3) XGBoost:

XGBoost has become widely used model amongst Kaggle competition. It is effortlessly on hand as open-supply software, and it is able to be used on lots of structures and interfaces. XGBoost stands for excessive Gradient Boosting. The key purpose for the use of XGBoost is its execution pace and version performance. XGBoost employs ensemble studying methods, this means that it employs a group of more than one algorithm to supply output. XGBoost gives parallel and disbursed computing at the same time as offering reasonably-priced reminiscence use.

Relevant mathematics Associated with project:

S= I, O,F, DD, NDD, Failure, Success

Where,

S=System

I= Input

O=Output

F=Failure

S=Success

I is Input of system

Input I = set of Inputs

Where,

I= {Telecom User data}

F is Function of system

F = set of Function

Where,

F1= {Input Dataset}

F2={Feature Extraction}

F3= {Train Dataset}

F4= {Feature Matching}

F5= {Test Dataset}

F6= {Random Forest}

F7= {Decision Tree}

F8={XGBoost}

O is Output of system

Output O = {Churn Prediction}

· Success Conditions: Product working Soothly.Churn Prediction successfully.

· Failure Conditions: if internet connection Unavailable.
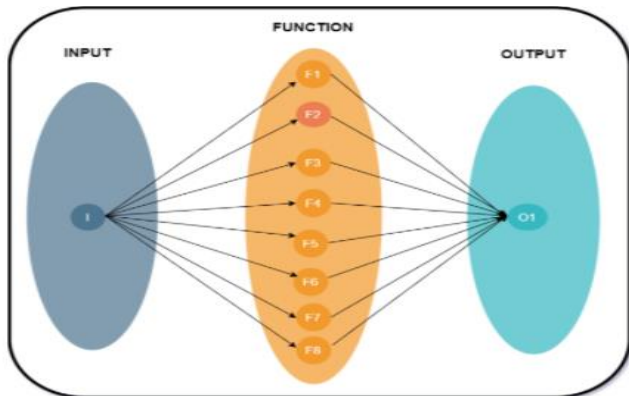
Venn Diagram :

Fig. 1: Venn diagram

Where,

I= {Input Data}

F1={Browse Image}

F2={Feature Extraction}

F3={Train Dataset}

F4={Feature Matching}

F5={Test Dataset}

F6={Random Forest}

F7={Decision Tree}

F8={XGBoost}

O = {Churn Prediction}

SYSTEM ARCHITECTURE:

We examine data from consumers who have already churned (respond) as well as their attributes/behavior (predictors) prior to the churn. Customers' demographic information, total charges, and the sort of service they receive from the company are all included in the dataset. It is made up of churn data from over a thousand consumers divided across 21 parameters gathered from Kaggle. We will try to predict the reaction for existing customers by fitting statistical models that relate the predictors to the response. This strategy is classified as supervised learning.
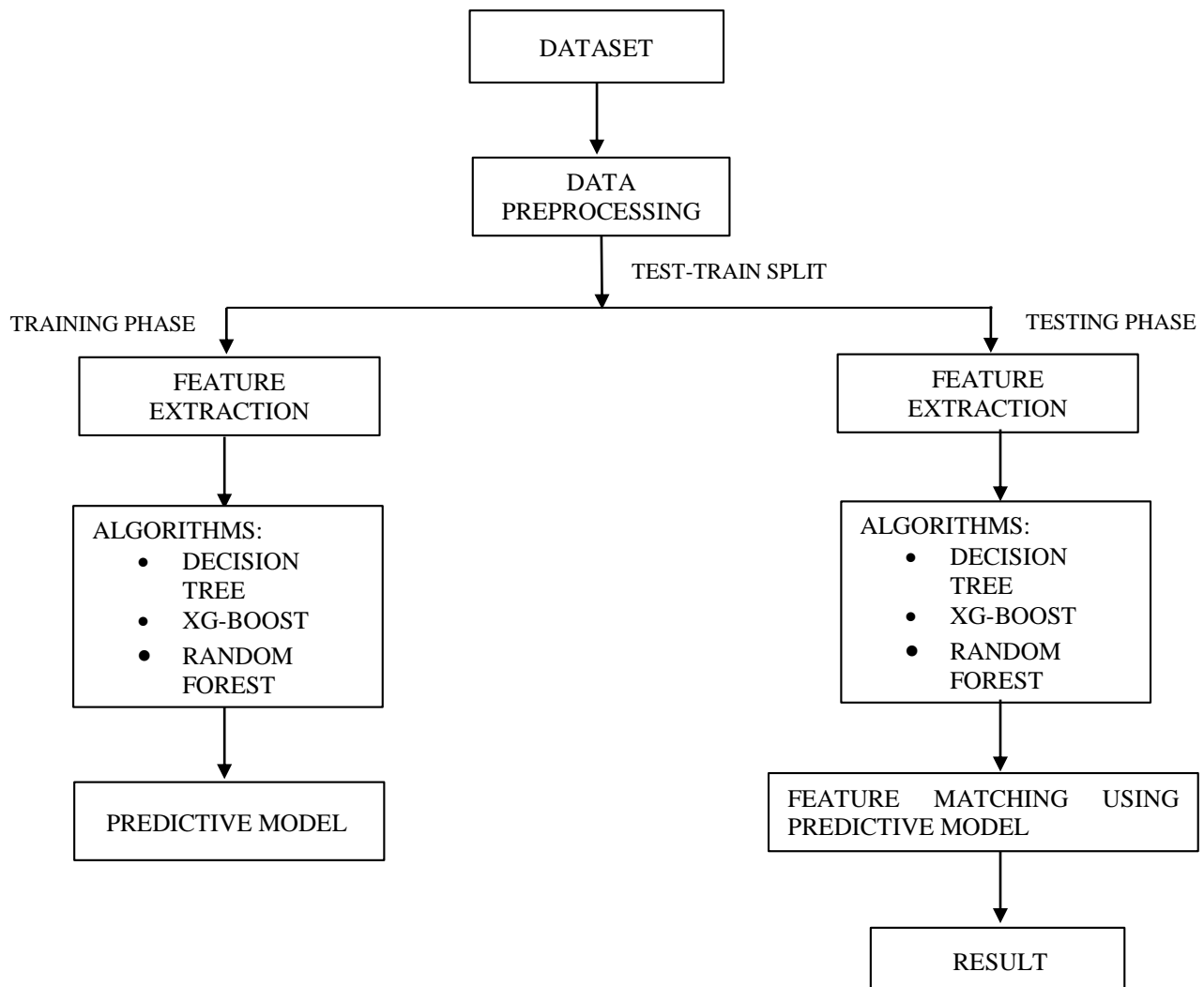
Fig 2: System Architecture

CONCLUSION:

Churn prediction is one of the most efficient retention methods employed in the telecom industry. It immediately leads to better cost allocation in customer relationship management operations, resulting in future revenue and profit retention. In order to maintain a loyal client base, telecom service providers try to keep customers with them. As a result, it is possible to conclude that machine learning approaches have shown to be accurate and useful in predicting client attrition.

REFERENCES:

[1] Brandusoiu I, Toderean G, Ha B. Methods for churn prediction in the prepaid mobile telecommunications industry. In: International conference on communications. 2016. p. 97–100.

[2] He Y, He Z, Zhang D. A study on prediction of customer churn in fixed communication network based on data mining. In: Sixth international conference on fuzzy systems and knowledge discovery, vol. 1. 2009. p. 92–4.

[3] Idris A, Khan A, Lee YS. Genetic programming and adaboosting based churn prediction for telecom. In: IEEE international conference on systems, man, and cybernetics (SMC). 2012. p. 1328–32.

[4] Huang F, Zhu M, Yuan K, Deng EO. Telco churn prediction with big data. In: ACM SIGMOD international conference on management of data. 2015. p .607– 18.

[5] Gupta, M, Gandhi, A.B., Gupta, S.C. (2018). Machine Learning as Intelligent tool for Churn Prediction in Telecommunication Industry. International Journal of Computer Applications, 181(10), pp. 16-22.

[6] Madan, M, Dave, M, Nijhawan, V.K. (2015). A Review on: Data Mining for Telecom Customer Churn Management. International Journal of Advanced Research in Computer Science and Software Engineering, 5(9), pp. 813-817.

[7] Yabas, U, Chankya, H.C. (2013). Churn prediction in subscriber management for mobile and wireless communications services. IEEE Publications.

[8] Zhang, Y.; Qi, J.; Shu, H.; Cao, J. A hybrid KNN-LR classifier and its application in customer churn prediction. In Proceedings of the 2007 IEEE International Conference on Systems, Man and Cybernetics, Montréal, QC, Canada, 7–10 October 2007; pp. 3265–3269.