iJRASET
International Journal For Research in
Applied Science and Engineering Technology

# INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 Issue: XI Month of publication: November 2017

DOI: http://doi.org/10.22214/ijraset.2017.11115

International Journal For Research In Applied Science & Engineering Technology (Ijraset)
Issn: 2321-9653; Ic Value: 45.98; Sj Impact Factor:6.887
Volume 5 Issue X1, November 2017- Available At Www.Ijraset.Com

# Research on Prediction Model of the Detection and Impact of New Telecom Offer's for Retention of Vulnerable Customers

Nikhil Ganage[1], Nishant Rajput[2], Jeet Thakur[3],
[1]Department of Information Technology, MIT Pune, Pune University.
[2]Department of Electronics and Telecommunication, Manav Rachna International University.
[3]Third year Information Technology student at Fr CRCE, Mumbai University,

***Abstract: Due to the quick improvement of telecommunication industry, the service suppliers are slanted more towards development of the subscriber base. To address the issue of making due in the focused condition, the maintenance of existing customers has turned into an enormous test. In the review done in the Telecom business, it is expressed that the cost of obtaining another customer is significantly more that holding the current one. This paper studies about the effect of new telecom services and offers for retention of existing (customers who have ever picked the first telecom service packages) and the income variety from the point of view of utility and client decision conduct. Critical indicators, for example, utility of service packages, exchange likelihood of the customers inside and expected difference in income are obtained, which are helpful for market orientation, income forecast and improvement administration of the new telecom services tariff.***
***Keywords: Telecom, Telecom Offers, Customer retention, Prediction analysis.***

## I. INTRODUCTION

Today, telecommunication market everywhere throughout the world is confronting an extreme loss of income because of wild rivalry and loss of potential customers. To keep the upper hands and procure however many customers as could reasonably be expected, most operators contribute a colossal measure of income to extend their business in the precise beginning [1]. Accordingly, it has turned out to be crucial for the operators to secure the sum contributed and to pick up no less than a base benefit inside a brief timeframe. Since it is in particular testing and repetitive issue to keep the customers in place for a long span because of the opposition engaged with this business field. To make due in the market, telecom operators typically offer an assortment of retention approaches to pull in new customers. This is the significant reason for the subscribers abandoning one network and moving to another which suits their necessities. As indicated by telecom market, the procedure of subscribers (either prepaid or postpaid) changing from keeps on occurring for any telecom industry, it would prompt the colossal loss of income to the organization. In this circumstance, the main solution to overcome such peril situation is to hold in the market, operators are compelled to search for elective methods for utilizing data mining systems and factual instruments to distinguish the reason ahead of time and to take prompt endeavors accordingly. This is conceivable if the previous history of the customers is broke down efficiently. Luckily, telecom ventures produce and keep up an expansive volume of data. They incorporate Billing information, Call detail Data and Network data. These measurable devices are to distinguish the reason ahead of time and to take quick endeavors accordingly. This is conceivable if the past history of the customers is broke down efficiently. Luckily, telecom businesses produce and keep up a huge volume of data. This voluminous sum data guarantees the extension for the use of data mining systems in telecommunication database. As a lot of information is hidden in the data created by the telecom enterprises, there is a ton of degree for the scientists to investigate the data in alternate points of view and to help the operators to enhance their business in different ways. Just the important data things which truly add to the particular analysis must be considered for any study. This study concentrates on churn forecast, the significance of highlight extraction and the utilization of data mining strategies in churn expectation in telecomm data.

## II. PROBLEM DESCRIPTION

In a business domain, the term, customer steady loss essentially alludes to the customers abandoning one business service to another. Customer churn or subscriber churn is additionally like wearing down, which is the procedure of customers changing from one service supplier to another namelessly. From a machine learning point of view, churn expectation is a managed (i.e. marked) issue characterized as takes after: Given a predefined estimate skyline, the objective is to foresee the future churners over that skyline, given the data related with every subscriber in the network [7]. The churn expectation issue discussed here includes 3 stages, to be specific, i) preparing stage, ii) test stage, iii) forecast stage. The contribution for this issue incorporates the data on past requires each mobile subscriber, together with all individual and business information that is kept up by the service supplier. In expansion, for the preparation stage, names are given as a rundown of churners. After the model is prepared with most elevated precision, the model must have the capacity to foresee the rundown of churners from the genuine dataset which does exclude any churn name. In the viewpoint of information disclosure process, this issue is sorted as prescient mining or prescient displaying. Churn Forecast is a wonder which is utilized to recognize the conceivable churners ahead of time before they leave the network. This aides the CRM office to anticipate subscribers who are probably going to churn in future by taking the required retention approaches to draw in the likely churners and to hold them. In this way, the potential loss of the organization could be mollified. This study uses data mining methods to recognize the churners.

## III.PROPOSED METHODOLOGY

The framework discussed below is based on the Knowledge Discovery Data (KDD) process .The issue of our discourse manages the discrete esteemed target variable and our definitive point is to pronounce every subscriber as "possible churner‟ or "possibly non churner", so the KDD work for our concern is characterized to be the grouping issue. The initial phase in prescient displaying is the procurement and arrangement of data. Having the right data is as essential as having the right strategy. The accompanying methodologies are given below

### *A. Data Acquisition*

Data Acquisition is a troublesome issue for the researchers to procure the real dataset from the telecom industries. This is on the grounds that the customer's private details might be abused. The data set utilized as a part of this study was obtained from an on the web source [1]. Since churn forecast models requires the previous history or the use conduct of customers amid a particular timeframe to anticipate their conduct soon, they can't be connected specifically to the genuine dataset. Subsequently, it is the typical practice to play out some sort of collection on the dataset. Amid the procedure of accumulation, notwithstanding the real factors, new factors will be created which show the occasional expending conduct of the customers. These factors have key information to be utilized by the expectation models in determining the conduct of customers ahead of time. The dataset utilized here was collected for 6 months term.

### *B. Data Preparation*

In data mining issues, data readiness expends extensive measure of time. In the data readiness stage, data is gathered, coordinated and cleaned. Coordination of data may require extraction of data from various sources. Once the data has been orchestrated in forbidden frame, it should be completely described. Data should be cleaned by settling any ambiguities, mistakes. Additionally repetitive and risky data things are to be expelled at this stage. Not all fields of the database are constantly reasonable for demonstrating purposes. Fields with extraordinary esteems, similar to addresses or individual open codes are require not be utilized. These don't have prescient incentive as they remarkably distinguish each column. Additionally fields with just a single esteem are forgotten, as these speak to an irrelevant piece of the data. At last, fields with too much "null‟ esteems are likewise prohibited. Moreover, individual data that is accessible is thought to be solid. Likewise in preparing dataset, the vast majorities of the qualities for the characteristic CONTRACT_CLOSURE_PERIOD are indistinct and are given the value 0.

### *C. Derived Value*

Derived variables are new variables in light of unique variables. The best derived variables are those that speak to something in this present reality, for example, a portrayal of some hidden customer conduct [3]. Since the first variables themselves are accumulated, they can likewise be called as derived variables. In our dataset, the variable CONTRACT_CLOSURE_PERIOD is a derived variable which is ascertained by subtracting the beginning date of the association from the present date on which this expectation is connected. There are some broad classes of derived variables, similar to add up to esteems, normal esteems, and proportions. Our study considers the normal incentive in the course of the most recent a half year as a derived variable sort. Likewise, the proportion between the normal in the course of the most recent three months and the normal over all prior months is utilized as a derived variable. What's more various particular derived variables are utilized. A few cases are:No. Of recharges

*a)* Call Average duration
*b)* In/Out call ratio
*c)* Billing summary
*d)* Defaulters

The derived variables clarify customer behavior efficiently than the first variables. For example, knowing how long it was since a customer utilized VAS service is significantly more instructive than knowing whether a customer utilized VAS service this month. Notwithstanding the real qualities, various derived variables are proposed which portray customer behavior properly. The data is as of now amassed for a half year's day and age. Once in like clockwork total is performed. For demonstrating purposes this is the coveted level of collection. Every day or week after week accumulated data won't offer any favorable circumstances over month to month collected data. For preparing reason, a dataset of 25,000 customers are considered and this tally is more than adequate to prepare the model. Furthermore, to test reason, dataset with 10,000 records are utilized. Each dataset comprised of 300 properties and very nearly half of them are derived traits. Not all the 300 properties are utilized for displaying. Just applicable characteristics are removed from (counting both genuine and derived gathering) the dataset.

### *D. Variable Extraction*

By alluding to the past research papers in regards to this study, in light of manual surmising and the information assembled from the telecom organization's personals, we have chosen conceivable variables for demonstrating the decision tree. Among them, the most critical variables that have higher commitment to foresee the churn are chosen. The chosen variables are gathered under 7 classes and are described below.

## IV. IMPLEMENTATION

### *A. Lead Indicators*

*1)* Subscriber active period – Lead Indicator
Subscriber active period is the period during which a customer is availing the services of that particular network. We need to select the customers who are associated with the operator for more than 90 days.

*2)* Subscription expiry date – lead indicator
This indicator points to the customers whose subscription is expiring within 30 Days. The tendency of these customers to churn would be the highest because they might have to avail the pack again and the customers might start to lookout for better options.

*3)* Customer type – Another important factor to distinguish between the customers is by analyzing the customer’s profession. For example, if a customer is a student, he/she will be definitely looking out for cheaper plans. Similarly, we can differentiate between the customers whether the customer is from service background or business etc.

*4)* Customers now a days are very specific to the usage and the amount that they are paying for that service. Now let’s say a cutomer has a plan where he has 10 GB data per month and usage of that particular person is just 1 GB than why would he keep up with that plan, hence leading to high possibility of churning. So, we need to select subscribers who has more ratio of incurred amount to total usage.

*5)* Customers has different nature, different priorities, so we need to smartly judge the customers by the number of complaints and the nature of complaints. We need to distinguish the customers and select subscribers who has more flags like; frequent plan changer, with more complaints

*6)* Churning of customers can also be dependent on a very specific thing which is their movement. There are customers who might have shifted to different state/circle where they have to incur roaming charges. So we need to identify subscribers who are in roaming for more than 30 days. Make a decision tree from above factors. And find the suspected the churners. And then send promotional offers to these subscribers.

## V. MODEL CONSTRUCTION

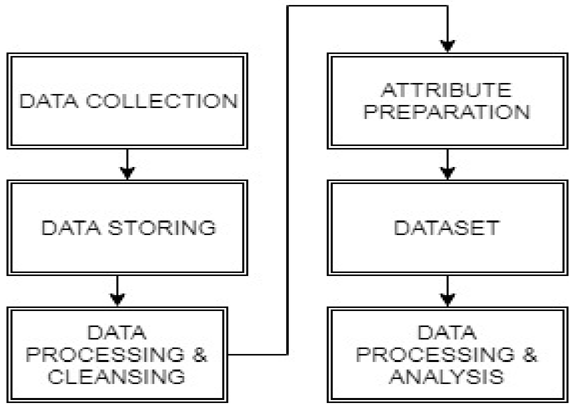



Fig 1: Strategy for data processing and analysis

The model made for this study is appeared in Figure 1. The control set portrayed above for the component variables are utilized for preparing the decision tree display and the neural network model. As there are no measurable techniques connected to the decision of list of capabilities, the Information pick up and Entropy of the attributes are computed to demonstrate the viability in finding the churn. As of now said, data is amassed for a half year. It implies the customer's conduct amid the previous a half year was utilized as a part of foreseeing the churners amid the seventh month. Following are the means to be taken after for the churn expectation

*1)* For each Lead Indicator calculated and threshold value is assigned
*2)* The Lead Indicator values of the training dataset are compared with the Lead Indicator's threshold then all Lead Indictors values are accumulated to declare that a customer will churn or not.
*3)* A model is then developed for the training dataset.
*4)* The model is then connected on the test dataset and the outcomes are recorded.
*5)* Out of total suspected churners, 50% are selected for promotional offers for the retention.
*6)* The above steps can be rehashed by shifting the threshold estimations of the attributes selected.
*7)* After one month check the number of outgoing customers from the churner set and validates the model.

## VI. PERFORMANCE MEASURE

Performance of a classification model depends on the tallies of test records accurately and erroneously predicted by it. These consider are arranged in a table no. 1

| MODEL | Churner Set | |
|---|---|---|
| | Promotion offered(50%) | NO Promotion offered(50%) |
| | 1100 | 1100 |
| Actual Outgoing Customers | 50 | 1000 |

Table 1: Matrix for Predicted and actual churners

THE ACCURACY OF THE MODEL

$$\text{ACCURACY} = \frac{\text{NUMBER OF RETAINED CUSTOMERS}}{\textit{Total Number of Predictions}} = 95.45\%$$

From the above computations, we watched the prescient exactness of 95.45% for our decision model. Thus, the predictive accuracy and different measures are calculated for the neural network model.

## VII. CONCLUSIONS

Telecommunication industry has suffered from high churn rates and immense churning loss. Although the business loss is unavoidable, but still churn can be managed and retained using promotional offers and service enhancement. Good methods need to be developed and existing methods have to be enhanced to prevent the telecommunication industry to face challenges. In this paper we discussed the selecting the right combination of attributes and fixing the proper threshold values may produce more accurate results after applying retention policies.

## VIII. REFERENCES


[1] Yu-Teng Chang, “Applying Data Mining To Telecom Churn Management”, IJRIC, 2009 67 – 77.

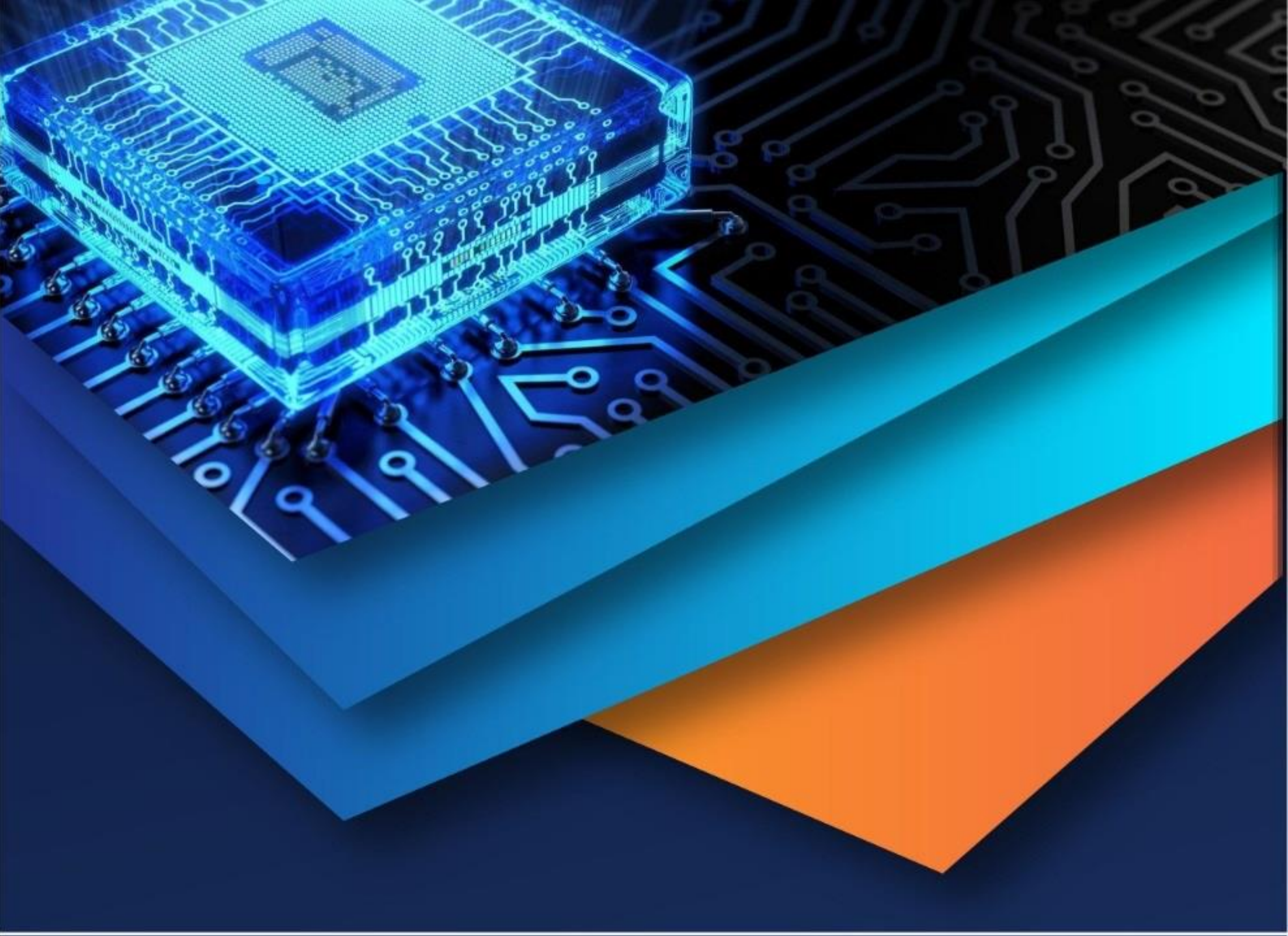
[2] N. kamalraj, A. malathi “A Survey on Churn Prediction Techniques in Communication Sector”., International Journal of Computer Application, feb 2013.

[3] Keyvan Vahidy Rodpysh, “Model to Predict the Behavior of Customers Churn at the Industry” International Journal of Computer Applications (0975 – 8887) Volume 49– No.15, July 2012

[4] Emilía Huong Xuan Nguyen ,Customer Churn Prediction for the Icelandic Mobile Telephony Market , in proc. The Faculty of Industrial Engineering, Mechanical Engineering and Computer Science University of Iceland in September 2011.

[5] Khalida, Sunarti, Norazrina, Faizin, ―Data Mining in Churn Analysis Model for Telecommunication Industry‖ Journal of Statistical Modeling and Analytics, Vol. 1, No.19- 27,ISSN: 2180-3102, 2010, Pp. 19-27.

[6] Amal M. Almana, Mehmet Sabih Aksoy, Rasheed Alzaharni, ―A Survey on Data MiningTechniques in Customer Churn Analysis for Telecom Industry‖, International Journal of Engineering Research and Applications, Vol. 4, Issue 5, ISSN: 2248-9622, May 2014, Pp. 161



10.22214/IJRASET

INDEX COPERNICUS 45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY