International Journal of Electronics Communication and Computer Engineering
Volume 13, Issue 4, ISSN (Online): 2249–071X

An Analysis of Customer Churn Predictions in the Telecommunications Sector

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 **– The Telecommunications (telecom) Industry is saturated 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. So customer churn is an important area of concern. This research work aims at carrying out a literature review for the past decade reviewing around 50 research papers in the area of telecom churn with two perspectives: mining technique applied and publication year. This review looks at the existing models in literature, using 30 selected CCP studies particularly from 2014 to 2020. 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 derived variables improves the accuracy of churn prediction which assists Telecom industries to retain their customers more efficiently.

Comparatively, a smaller study was done on user's perspective, taking 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.

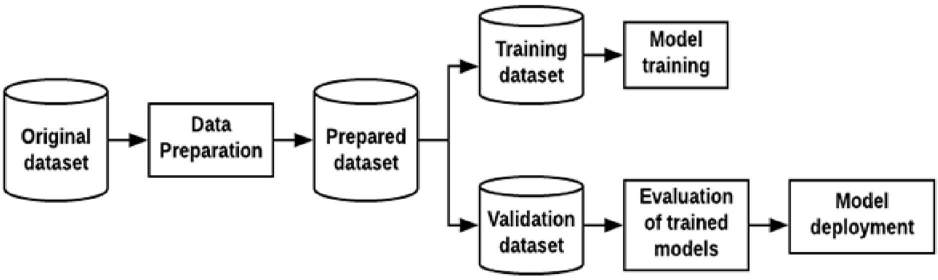

Fig. 1. Churn prediction modeling process.

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.

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 [4], 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.

C. *Motivation*

The Telecom industry is humongous, vibrant and dynamic with extremely large base of customers, making customer acquisition and customer retention imperative concerns for its survival and good profitability. The new entrants focus on customer acquisition, while old and matured one emphasize to focus on customer retention. Globalization enables customers to choose the best available services, which encourages the customers not to stick with a single company, rather opt from a diverse range of products/services. Customer churning is directly related to customer satisfaction [1]. Since the cost of acquiring new customer is much higher than retaining old news, operators lay preeminent significance on various customer related methodologies and analytics to ensure customer retention.

There is no clear common consensus on the prediction technique to be used to identify churn. Significant research in the field of churn prediction is being carried out using various statistical and data mining techniques since a decade. Big Data analytics with Machine Learning were found to be an efficient way for churn prediction. Several previous works [1] focused on various data mining techniques for churn prediction based on call detail records. The work in [13] focused on service failures and disconnections recorded to identify churn. Study [5] focuses to detect early warnings of churn by assigning “Churn Score” for numerous customer transaction logs.

So far, customer churn has been majorly studied on network parameters. Barely, any study could be addressed regarding churn prediction with user’s perspective taking their Quality of Experience into consideration. No study was done taking only data usage volumes into consideration. This thesis aims to predict customer churn using Big Data analytics, decision tree on a python based tool, considering only users’ data usage volumes from three different datasets. There is no standard model which addresses the churning issues of global telecom service providers accurately. This thesis predominantly focuses to identify churn using decision trees, one of the most popular data mining techniques. A decision tree is an eminent categorizer that use a flowchart-like process for categorizing instances. During the process of customer churn prediction, Telecom operators would often need to analyze the steps to figure out the probable cause and rationale instigating customers to churn. This could be only possible with decision trees as they are easy to interpret, visualize and analyze.

II. Problem Formulation

Through literature study, surveys and previous works, various discrepancies, challenges and difficulties are identified. After the data acquisition from the anonymous Telecom provider and an experimental survey by Mounika Reddy Chandiri, statistical and Big Data analytics were carried out to draw different convictions on usage trends. From the individual data usage traffic analysis by Hemanth Kumar Ravuri, a certain trend of variation on the usage pattern is expected. The analysis is also expected to result in certain correlations between the varying data traffic, annoyance, churn risk and the quality of experience with respect to users and the Big Data analytics indicating churn. From the study of the thesis work, a derivation of a general relation between users’ satisfaction and users’ traffic volume is expected to be reached.

In the competitive Telecom industry, public policies and standardization of mobile communication allow customers to easily switch over from one carrier to another, resulting in a strained fluidic market. Churn prediction, or the task of identifying customers who are likely to discontinue use of a service, is an important and lucrative concern of the Telecom industry. The aim of this thesis is to study and analyze customer churn prediction based on mobile data usage volumes with respect to QoE and users’ perspective with the help of Big Data analytics.

III. Literature Review

Adnan Amin (2019) is shows Uncertain samples have been shown to make a difference when applied to CCP in TCI. Through a novel method, the extracted insightful uncertain samples, and build a CCP models. Further, the have empirically evaluated the impact of the uncertain samples on the performance of the predictive model in the context of TCI. Overall the propose study offers two main contributions to the existing literature such as: (i) introduced a novel method to identify efficiently uncertain samples in the TCI dataset, and (ii) an empirical comparison of the lower distance and upper distance samples in the context of their impact on the performance of the prediction model. Moreover, the performance of the resulting models was evaluated with four measures which gave consistent and robust results. A benchmarking of the propose model for CCP in TCI using uncertain samples on this scale, does not have precedence in the current studies to the best of our knowledge. Hashmi, Butt and Iqbal (2013) carried out a systematic review on customer churn prediction in the telecom industry, considering studies carried out from 2002 to 2013. The author's selected 61 articles from the 834 initially availed in their primary search. In their study, it was found out that Decision Tree was the most used churn prediction technique. The study also revealed that the huge volumes, high dimensionality and imbalanced structure of telecom data is a big challenge for industry practitioners as they try to draw actionable insights from it. These issues give this study a foundation to consider literature from 2014 to 2017 and to further review literature as well as seeking solutions to the highlighted issues. KhakAbi et al. provides a brief review of papers from two perspectives: techniques used(model applied) and statistical reports. Through their findings, they have tried to point out the gaps and strengths in the area of churn. Another model that has been built and tested by Shyam and Ravi to predict future churn for wireless customers. This model obtained a 68% predictive accuracy, using the Naive Bayes algorithm. In their model, Shyam and Ravi has suggested a very unique approach to "mine their own data", which means that the organizations should mine their existing customer information to develop models of predicting customer behavior. Junxiang Lu in his research paper, says that methods like decision tree, logistic regression, etc. are predicting customer churn but they hardly tell that when the churn will happen. He has used survival analysis techniques to predict which customer will churn and when the churn will happen, thereafter, helping the telecom companies in customizing their customer treatment programs. The work of Chen et al. is based on the use of SAS Enterprise Miner to design a predictive model for churn management, providing the foundation for predicting the data of customer churn and reduce the number of churns. One more study by Shan Jin et al. emphasizes on a predictive model to find out possible churners and provide personalized services. The researchers have explored a few data mining techniques, namely decision tree c4.5, BPN and compared the effectiveness of each for customer churn.

Several other studies have been carried out in form of surveys about CCP methods. Umayaparvathi and Iyakutti (2016) studied the open datasets, methods and metrics in telecom CCP. The authors found out that Support Vector Machines (SVM) out-performed Decision Tree and Neural Networks in churn classification during CCP. A comparison study carried out by also revealed SVM-POLY as the best classifier. However in Monani et al. (2016)'s survey still on CCP in telecom, Neural networks were found to outperform the rest of the techniques in terms of performance. The authors' research also recommended survival analysis as a solution for determining the possibility of a customer churning at a particular moment in time.

Research in the area of customer churn is always a trending topic. The unbridled growth of databases in recent years brings data mining to the forefront of new business technologies., becoming our only hope for elucidating the patterns that underlie it. Significant research in the field of churn prediction is being carried out using various statistical and data mining techniques since a decade. This chapter presents the recent and prominent publications on churn prediction in the recent years.

Saad Ahmed Qureshi et. al. aims to present commonly used data mining techniques for churn prediction. The dataset used was obtained from Customer DNA website and contains traffic data of 1,06,000 customers and their usage behavior for three months. The class imbalance problem was solved by re-sampling. Regression analysis, Artificial Neural Networks, K-Means Clustering, Decision Trees including CHAID, Exhaustive CHAID, CART and QUEST were taken into consideration to identify churn. The results were compared based on the values of precision, recall and F-measure. Decision trees, especially Exhaustive CHAID were found to be the most accurate algorithm in identifying potential churners.

Muhammad Raza Khan et. al. presented a unified analytic framework for detecting the early warnings of churn, and assigning a “Churn Score” to each customer that indicates the likelihood of a particular customer to churn within a predefined amount of time. The approach uses a brute force approach to feature engineering that generates a large number of overlapping features from customer transaction logs, then uses two related techniques to identify the features and metrics that are most predictive of customer churn. These features are then fed into a series of supervised learning algorithms that can accurately predict subscriber churn. For a dataset of roughly 1,00,000 subscribers from a South Asian mobile operator observed for 6 months, an approximate of 90 percent accuracy was achieved.

A.Idris and A. Khan et. al. a dataset of 40,000 instances provided by cell2cell Telecom Company was pre-processed to a balanced form. In the preprocessing stage, in order to provide discriminating features to the classifiers mRMR, Fisher’s ratio and F-Score feature extraction methods were used. For each of these methods, a linear search is performed to select the features which provide maximum discriminating information to the classifiers and hence produce better performance. When a linear search is performed for all the methods with rotation forest, for mRMR the accuracy for predicting the churners was 76.2%, while it was 69.1% and 65.2% for Fisher’s ratio and F-Score respectively. For Random Forest, the accuracy of churn prediction for mRMR, Fisher’s Ratio and F-Score were 74.2%, 71.6% and 71.3% respectively.

IV. Method

A. *Churn Prediction Techniques*

A variety of churn prediction techniques were used. In certain studies, a combination of two or more techniques were 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. *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.

V. Conclusion

This paper is brief analysis of different author work about the churn production. This research work aims at c-

-arrying out a literature review for the past decade reviewing around 50 research papers in the area of telecom churn with two perspectives: mining technique applied and publication year. 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.

REFERENCES

[1] Jose Saias, Luís Rato and Teresa Goncalves "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 Brândusoiu, 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.

[15] Aimée Backiel, Bart Baesens and Gerda Claeskens "Predicting time-to-churn of prepaid mobile telephone customers using social network analysis" Journal of the Operational Research Society 2010.

[16] Michael J. Perez, William T. Flannery "A study of the relationships between service failures and customer churn in a telecommunicati--ons environment" PICMET '09 - 2009.

[17] Luo Bin; Shao Peiji Liu Juan " Customer Churn Prediction Based on the Decision Tree in Personal Handyphone System Service" 2007 International Conference on Service Systems and Service Management DOI: 10.1109/ICSSSM12473.20079-11 June 2007.

[18] Dahiya, K., Bhatia, S.: Customer churn analysis in telecom industry. In: 2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions), pp. 1-6 (2015).

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.