# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

## IN COMPUTER & COMMUNICATION ENGINEERING

**Impact Factor: 8.379**

INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

# A Review Paper on Prediction Customer Churn over Telecom Data Using Data Mining Techniques

**Kanchan Choudhary**

M.Tech Research Scholar, Department of Computer Science and Engineering, Oriental Institute of Science and

Technology, Bhopal, India

**ABSTRACT—** Predicting customer churn in telecommunication industries becomes a most important topic for research in recent years. Because its helps in detecting which customer are likely to change or cancel their subscription to a service. Now a days the mobile telecom market has growing market rapidly and all the telecommunication industries focused on building a large customer base into keeping customers in house. So it is very important to find which customers are wants to switch to a other competitor by cancel their subscription in the near future. Analysis of data which is extracted from telecom companies can helps to find the reasons of customer churn and also uses the information to retain the customers. So predicting churn is very important for telecom companies to retain their customers. In this paper we can focuses on various data mining techniques for predicting customer churn.

**KEYWORDS*:*** Churn prediction, data mining, telecom system ,Customer retention, classification system.

## I. INTRODUCTION

Customer churn prediction in Telecom industry is one of the most prominent research topics in recent years. It consists of detecting customers who are likely to cancel a subscription to a service. Recently, the mobile telecommunication market has changed from a rapidly growing market into a state of saturation and fierce competition. The focus of telecommunication companies has therefore shifted from building a large customer base into keeping customers in house. For that reason, it is valuable to know which customers are likely to switch to a competitor in the near future. The data extracted from telecom industry can help analyze the reasons of customer churn and use that information to retain the customers. So churn prediction is very essential in telecom industries to retain their customers. In this thesis we can use classification techniques along with decision tree to better predicting churn in telecom sector.

Predicting customer churn in telecommunication industries becomes a most important topic for research in recent years. Because its helps in detecting which customer are likely to change or cancel their subscription to a service. Now a days the mobile telecom market has growing market rapidly and all the telecommunication industries focused on building a large customer base into keeping customers in house. So it is very important to find which customers are wants to switch to a other competitor by cancel their subscription in the near future. Analysis of data which is extracted from telecom companies can helps to find the reasons of customer churn and also uses the information to retain the customers. So predicting churn is very important for telecom companies to retain their customers. In this we can focuses on various data mining techniques for predicting customer churn.

## II. MOTIVATION

Telecommunication data pose several interesting issues for data mining.

a) Telecommunication databases may contain billions of records and are amongst the largest in the world.

b) The raw data is often not suitable for data mining. The extraction of the raw data from the huge data files and summarization for the data mining is the time consuming tasks. This paper focuses on the customer relation data and call quality data to identify the customers who may or may not leave the company service. The customer dissatisfaction due to the call quality affect customer churn. The analysis of these datasets will help to understand customer satisfaction with the company services. The dissatisfaction indicators such as number of complaints and call drop rate due to poor connectivity quality have significant impact on the customer churn.

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.

## III. DATA MINING

Data Mining [9] 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."

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.

### Data Mining In Telecom

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.

### Churn Analysis in Telecommunication Sector

Data volume has been rising at an incredible step over the last two decades due to progressions in information knowledge [5]. At the same time there has been huge progress in data mining. Many new approaches & methods have been added to process data & collect information. The data collected from any source is a raw data in which valuable info is unknown.

Data mining can be clear as the process of eliminating valuable information from data. Data mining methods have been effectively useful in many different domains. The toughest difficulty faced by the telecom industry is customer churn. The customer churn models goal to distinguish customers with the high probability to jump / leave the service supplier. A database of clients who might churn permits the company to target those clients& start maintenance strategies that decrease the percentage of customer churning.

Retention of old clients is always desirable option to the company. Attracting new clients costs almost 5-6 times more than retentive the old clients. Attracting a new client comprises new recruits of manpower, cost of publicity & discounts. A loyal client, who has been with the business for quite the long time, tends to produce higher profits& is less complex to competitor prices. Such clients also cost less to keep & in addition, offer valuable word-of-mouth marketing to the professional by mentioning their relations, friends, & other associates. In telecom Industry, the scheme is built to offer service to some average number of clients, when the client number falls below the planned number. It is careful as loss to company.

A minor step towards retentive a current customer can lead to an important growth in revenues & profits. The condition of the recollecting customers desires for correct models of customer churn prediction that are both accurate& comprehensible. The Models have to classify customers who are about to churn & their reason for churn to evade losses to industry of telecom, the model should be recognised to classify the reasons to churn & the enhancements required to recollect customers.

## IV. LITERATURE REVIEW

According to the paper [1] Customer Churn (CC) is a major issue and important concerns for large organizations and businesses alike. Telecom industries are attempting to improve methods to predict possible customer churn due to the immediate impact on revenue, particularly in the telecom sector. This paper discusses the various ML algorithms used to construct the churn model that helps telecom operators to predict customers who are likely to churn. The experimental results are compared to predict the best model among various techniques. As a result, the use of the Random Forest combined with SMOTE-ENN outperforms best result than other in terms of Fl-score. According to our analysis, the maximum prediction is 95 percent based on Fl-score.

According to [02] , Telecom industry has become one of the leading companies in the entire world. To grow further in this competitive world, it is important for companies to lessen the probability of churn. Churn is the shifting of customers across various service providers. One of the most important task that Telco industry has to deal with is predicting churn. Customers churn due to several factors including the exciting offers provided by competitors. Churn rate has an significant impact on the lifetime of customer. As churn has a direct impact on the revenue of industry, companies are looking forward to model with best accuracy. In the proposed system model is developed using Machine Learning. Companies will be able to identify the consumers who are most likely to leave and the causes influencing them with the aid of machine learning models. This data can be used by the service providers to improve their services and thereby reduce churn.

According to [03], In today's time, customer churn is one of the major issues in many large-scale industries. Because it has a direct impact on the company's revenue, it is necessary to find a solution to determine which customers are most likely to churn so that large-scale companies can make wise decisions and take steps to deal with customer churn. Customers may leave for a variety of reasons, but the most common is that they are dissatisfied with the services provided by the companies. In this proposed solution, we will be building a machine learning model that has the capability to predict the potential churn so that the Telecom companies can make proper marketing retention strategies as time passes. In this system, we will be using existing datasets and necessary pre-processing techniques like bivariate and univariate analysis, further using data visualisation to understand the dataset properly. After this, we will be building different classification models by applying and comparing different supervised machine learning algorithm such as the Logistic Regression, the Support Vector Machine algorithm, Decision Tree Classifier, and the Random Forest algorithm . The best machine learning algorithm is chosen by using performance metrics like accuracy, F1 recall and so on.

| Authors | Year | Title | Classification Techniques | Summary |
|---|---|---|---|---|
| R. Srinivasan; D. Rajeswari; G. Elangovan | 2023 | Customer Churn Prediction Using Machine Learning Approaches | Random Forest | This paper discusses the various ML algorithms used to construct the churn model that helps telecom operators to predict customers who are likely to churn. The experimental results are compared to predict the best model among various techniques. As a result, the use of the Random Forest combined with SMOTE-ENN outperforms best result than other in terms of Fl-score. |
| Yashraj Bharambe; Pranav Deshmukh; Pranav Karanjawane; Diptesh Chaudhari | 2023 | Churn Prediction in Telecommunication Industry | ML Algorithm | Telecom industry has become one of the leading companies in the entire world. To grow further in this competitive world, it is important for companies to lessen the probability of churn. Churn is the shifting of customers across various service providers. One of the most important task that Telco industry has to deal with is predicting churn. |
| O. Pandithurai; H. Humaid Ahmed; Hrudhai | 2023 | Telecom Customer Churn Prediction Using Supervised Machine Learning | Logistic Regression, the Support Vector Machine algorithm, | In this system, we will be using existing datasets and necessary pre-processing techniques like bivariate and univariate analysis, further using data visualisation to understand the dataset properly. After this, we will be building different classification models by applying and |

| | | Techniques | Decision Tree Classifier , and the Random Forest algorithm | comparing different supervised machine learning algorithm |
|---|---|---|---|---|
| Narayan. S; B. Sriman; Seetha R | | | | |
| Sahar F. Sabbeh | 2018 | Machine-Learning Techniques for Customer Retention: A Comparative Study | logistic regression and Linear Discriminant Analysis (LDA) | This study tries to gift a benchmark for the foremost wide used state of the humanities for churn classification. The accuracy of the chosen models was evaluated on a public dataset of consumers in medium Company. supported the findings of this study, ensemble – based mostly learning techniques ar counseled as each Random forest and Ad boost models gave the simplest accuracy. However, the study may be extended by together with hybrid models and deep learning models. alternative performance metrics may be used for performance analysis. temporal arrangement measures of the models also can be a serious indicator for performance. Models also can value against totally different datasets from different domains |
| Saran Kumar A., Chandrakala D., | 2016 | A Survey on Customer Churn Prediction using Machine Learning Techniques | machine learning algorithms | This paper provides a close study on the strategies used for the method of client churn prediction. every of the on top of churn prediction models has low accuracy and prediction. therefore an honest prediction model is needed so as to avoid the client churn downside. this will be achieved by combining SVM with boosting algorithms for higher accuracy and performance which may be thought of as a future work for Churn prediction. smart prediction models got to be perpetually developed and a mix of the projected strategies must be used. |
| Federico Castanedo , Gabriel Valverde, Jaime Zaratiegui, Alfonso Vazquez | 2015 | Using Deep Learning to Predict Customer Churn in a Mobile Telecommunication Network | deep learning | To apply deep learning models to predict client churn prediction during a mobile telecommunication network, we tend to enforced a multi-layer feedforward design and introduced a unique thanks to encipher computer file to find out ranked illustration on real datasets. |
| Junxiang Lu | 2015 | Predicting Customer Churn in the Telecommunications Industry — An Application of Survival Analysis Modeling Using SAS_ | SAS | Our experiments recommend multi-layer feedforward models ar an efficient formula for predicting churn and capture the advanced dependency within the information. Experiments shown that the model is sort of stable on totally different months, so generalize well with future instances and don't overfit the coaching information. |
| Antonio Canale, Nicola Lunardon | 2014 | CHURN PREDICTION IN TELECOMMUNICATIONS INDUSTRY. A STUDY BASED ON BAGGING CLASSIFIERS | Classification | In this article, we tend to use billions of decision records from associate degree enterprise business intelligence system and gift our current work towards exploitation deep learning for predicting churn during a paid mobile |

Table 1. Related Work

## V. CHALLENGES & FUTURE DIRECTION

In a business setting, the term, client attrition merely refers to the purchasers exploit one business service to a different. client churn or subscriber churn is additionally kind of like attrition, that is that the method of shoppers shift from one service supplier to a different anonymously. From a machine learning perspective, churn prediction could be a supervised (i.e. labeled) downside outlined as follows: Given a predefined forecast horizon, the goal is to predict the longer term churners over that horizon, given the info related to every subscriber within the network. The churn prediction downside diagrammatical here involves three phases, namely,

    i)   the training part,
    ii)  testing part,
    iii) prediction section.

The input for this downside includes the info on past necessitate every mobile subscriber, along with all personal and business data that's maintained by the service supplier. additionally, for the training section, labels are provided within the type of an inventory of churners. when the model is trained with highest accuracy, the model should be able to predict the list of churners from the important dataset that doesn't embody any churn label. within the perspective of information discovery method, this downside is categorized as prognostic mining or prognostic modeling.

In the proposed system advance programming like python or R will be used to build the model for churn prediction. It is widely used among statisticians and data miners for developing statistical software and data analysis. It is freely available and a powerful statistical analysis tool which has not yet been explored for building model for churn prediction[7].
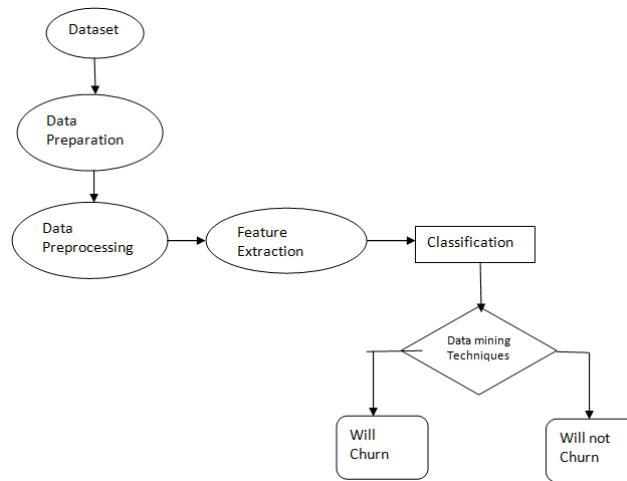


Figure 1. Churn Prediction Framework

This is where the churn prediction model [4] can help the business to identify such high risk customers and thereby helps in maintaining the existing customer base and increase in revenues. Churn prediction is also important because of the fact that acquiring new customers is much costly than retaining the existing one. As the telecom users are billions in number even a small fraction of churn leads to high loss of revenue. Retention has become crucial especially in the present situation because of the increasing number of service providers and the competition between them, where everyone is trying to attract new customers and lure them to switch to their service.

With a large customer base and the information available about them data mining techniques proves to be a viable option for making predictions about the customers that have high probability to churn based on the historical records available. The data mining techniques can help find the pattern among the already churned customers and provide

useful insights which can then be used strategically to retain customers.

## VI. CONCLUSION

Churn prediction modelling systems serve as a means of understanding the customer's exact behaviour and work as an alert to the danger and timing of customer churn. Using data mining techniques helped us to accurately predict customer churn as well as defining the causes that lead to customer retention.

## REFERENCES

[01] R. Srinivasan; D. Rajeswari; G. Elangovan, "Customer Churn Prediction Using Machine Learning Approaches" , in IEEE 2023.

[02] Yashraj Bharambe; Pranav Deshmukh; Pranav Karanjawane; Diptesh Chaudhari, "Churn Prediction in Telecommunication Industry" in IEEE 2023.

[03] O. Pandithurai; H. Humaid Ahmed; Hrudhai Narayan. S; B. Sriman; Seetha R, "Telecom Customer Churn Prediction Using Supervised Machine Learning Techniques" in 2023 IEEE.

[04] Manpreet Singh, Sarbjeet Singh, Nadesh Seen, Sakshi Kaushal and Harish Kumar, "Comparison of learning techniques for prediction ofcustomer churn in telecommunication" in IEEE 2018.

[05] Muhammad Ali, Aziz Ur Rehman, Shamaz Hafeez, "Prediction of Churning Behavior of Customers inTelecom Sector Using Supervised LearningTechniques" in 2018 IEEE.

[06] Ibrahim M.M.Mitkees, Asist. Prof. Sherif M Badr, Dr. Ahmed Ibrahim Bahgat ElSeddawy, "Customer Churn Prediction Model using Data Mining techniques" in IEEE 2017.

[07] AMMAR A AHMED, Dr. D. Maheswari linen, "A Review And Analysis Of Churn Prediction Methods For Customer Retention In Telecom Industries" in 2017 International Conference on Advanced Computing and Communication Systems (ICACCS -2017), Jan. 06 – 07, 2017, Coimbatore, INDIA.

[08] J. Burez and Dirk Van den Poel. "Handling class imbalance in customer churn prediction." Expert Systems with Applications, vol. 36, no. 3, pp. 4626-4636, 2009.

[09] Veronikha Effendy and ZK AbdurahmanBaizal. "Handling imbalanced data in customer churn prediction using combined sampling and weighted random forest." In 2014 2nd International Conference on Information and Communication Technology (ICoICT), pp. 325-330. IEEE, 2014

[10] Ning Lu, Hua Lin, Jie Lu, and Guangquan Zhang. "A customer churn prediction model in telecom industry using boosting." IEEE Transactions on Industrial Informatics, vol. 10, no. 2, pp. 1659-1665, 2014.

[11] Xiaojun Wu and SufangMeng. "E-commerce customer churn prediction based on improved SMOTE and AdaBoost." In 2016 13th International Conference on Service Systems and Service Management (ICSSSM), pp. 1-5. IEEE, 2016.

[12] G. Ganesh Sundarkumar, Vadlamani Ravi, and V. Siddeshwar. "One-class support vector machine based undersampling: Application to churn prediction and insurance fraud detection." In 2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), pp. 1-7. IEEE, 2015.

[13] Qiu Yihui, and Zhang Chiyu. "Research of indicator system in customer churn prediction for telecom industry." In 2016 11th International Conference on Computer Science & Education (ICCSE), pp. 123-130. IEEE, 2016

[14] Qiuhua Shen, Hong Li, Qin Liao, Wei Zhang, and KoneKalilou. "Improving churn prediction in telecommunications using complementary fusion of multilayer features based on factorization and construction." The 26th Chinese Control and Decision Conference (2014 CCDC), pp. 2250-2255. IEEE, 2014.

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

## IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462     6381 907 438     ijircce@gmail.com

Scan to save the contact details