

# CUSTOMER CHURN PREDICTION MODEL FOR TELECOMMUNICATION INDUSTRY: A REVIEW

<sup>1</sup>Shweta P. Pund, <sup>2</sup>Dayanand B. Gore

<sup>1</sup> Research Student, <sup>2</sup> Research Student

<sup>1</sup>st Department of C.S. and I.T., Dr.Babasaheb Ambedkar Marathwada University,  
Aurangabad, Maharashtra 431004, India),

<sup>2</sup>nd Department of C.S. and I.T., Dr.Babasaheb Ambedkar Marathwada University,  
Aurangabad, Maharashtra 431004, India)

**Abstract:** Data mining plays an important role in churn prediction for telecommunication industry. Telecom sector is a rapidly changing sector so it needs to hold the loyal customers for long time. Customer churn prediction is the term which indicates who is in the stage to leave the company. Telecommunication industries need to retain their customers to avoid revenue loss, because retaining old customers is easy than gaining new customers. The need of churn prediction is for the profit maximization in most of the organizations. Churn models is a good method of analyzing “Churn Risk” for any organizations or firms. Prediction models are used in churn management to forecast the likelihood of customer churn. Churn models can run through the basis of profitability and thus in a way consistent with business goals. This paper reviewed 12 research papers in the aim of identifying data mining techniques used to predict the customer churn.

**KEYWORDS - :** *Customer churn, data mining, prediction models, profit maximization, ROC*

## I Introduction

Many organizations have large volume of data to handle. Customer churn provides a way of predicting customers for archiving profit maximization for the organization. In this review paper various aspects of churning is studied by many research papers and verified their output by applying classification algorithms. Customer churn prediction has emerged as an indispensable part of telecom sectors strategic decision making and planning process. Customer retention is one of the main objectives of CRM (Customer Relationship Management). Large companies have millions of customers in telecommunication industries. Basically data depends on three types, customer data, call data and network data. This data base contains customer related information like name, address, gender, marital status, income, service plan, payment history etc. Call data contains call information or call details for every call like call duration, date and time of call etc. Whereas network data are very complex including thousands of inter-related components. It comprises handling network problems, errors, timestamps etc.

## II Review of Literature

1. [ Dr. Mamta Madan, Dr. Meenu Dave, Vani Kapoor Nijhawan]The authors have proposed strategy, issues related with telecom industry. The method of prediction requires for churn prediction, Insolvency Prediction, Fraud Detection. a) Churn Prediction: This is a big issue today because customers are going to change their service providers soon, to gain new plans offered by other companies. To retain these customers, customer churn prediction helps to analyze customer behavior.b) Insolvency Prediction: Insolvency predicts the customers who refuse to pay the bills. These customers can be predicted in advance to cover the loss of company. c) Fraud Detection: Fraud is very expensive activity in the telecom sector. So the telecom companies should try to find the fraudulent users and their patterns.In this paper they focused on various classification method and comparative study of these methods involving decision trees, logistic regression, support vectors machine, neural networks. 2. [Sam Drazin and Matt Montag] They have also tested decision tree algorithms to measure classification performances. Decision tree algorithms provides optimization computational efficiency as well as provides classification accuracy. The data sets were tested using the J48 decision tree-inducing algorithm.3. [Georges D. Olle Olle and Shuqin Cai] In this paper hybrid model is applied logistic regression, Receiver Operating Characteristic Curve. The results show that the hybrid mode (HM) built using LR and VP performs better than each of the individual methods. Here major classifier has predicted the instance class “zero” (or One) then the result true negative and true positive are compared. The technique measures how the classifier rules churners and non-churners.4. [V. Umayaparvathi, K. Iyakutti 2016] this paper focused on existing works on churn prediction in three different perspectives: datasets, methods, and metrics. These metric can be used for analyzing the performance of any model by examining confusion metrics which contains two rows and two columns (FP), false negatives (FN), true positives (TP), and true negatives (TN) provides churn prediction accuracy. It also compares the efficiency of Decision tree and Neural Network classifiers and lists their performances.5. [Thomas Verbraken, Wouter Verbeke, and Bart Baesens] This paper proposed the theory of churn analysis in the context of cost benefit

analysis and identifying misclassification cost. MP (Maximum profit), EMP (Expected Maximum Profit) and H measure (Misclassification cost) are identified by various classifiers involving decision tree approaches, SVM Based methods, statistical classifiers etc. Performance is tested on 8 different telecom data sets.

### Comparative Analysis

Sr. No.	Title, Author, Year	Methodology	Remarks
1.	“ Optimizing Coverage of Churn Prediction in Telecommunication Industry” Adnan Anjum, Adnan Zeb, Imran Uddin Afridi, Pir Masoom Shah, Saeeda Usman, 2017	Data preprocessing, feature elimination, modeling data by N algorithms, Ensemble	Best algorithm is declared for prepaid base
2.	“Churn Analysis in Telecommunication using Logistic Regression” Helen Treasa Sebastian and Rupali Wagh, March 2017	Data visualization tools, Based on Tenure, Based on tech support, SVM (Support vector regression) with R, (backward logistic regression	Minimal loss of customers through predictions with 80.02% accuracy
3.	“Research Model of Churn Prediction Based on Customer Segmentation and Misclassification Cost in the Context of Big Data” Yong Liu, Yongrui Zhuang June 2015	Customer Segmentation by using classifiers and clustering, Decision Tree Algorithm, identifying misclassification cost using Logistic regression and Artificial neural network	Accuracy is higher than 86% and the model coverage is higher than 74%.
4.	“Customer Churn Analysis in Telecom Industry” Kiran Dahiya, Surbhi Bhatia 2015 IEEE	uses datasets of 3 different sizes, compared 3 algorithms J48, Random Tree and SimpleCART	Accuracy comparison between ID3 and C 4.5 algorithm
5.	“Predicting Customer Churn in Mobile Telephony Industry Using Probabilistic Classifiers in Data Mining” Clement Kirui, Li Hong, Wilson Cheruiyot and Hillary Kirui March 2013	C4.5 Decision Tree, Naive Bayes, examining ROC curve	Identification of churners and non-churners through graphically
6.	“Churn prediction on huge telecom data using hybrid firefly based classification” Ammar A.Q. Ahmed, Maheswari March 2017	Firefly algorithm, Hybrid Firefly algorithm	Comparison of both algorithms in context of accuracy and time, Firefly acc.=86.36 % and hybrid firefly acc.=86.38 %

### Development Methodology

#### I. Data Acquisition

Churn analysis is based on the collection of data. This data set may be accessible from the data ware house of respective company [7]. Online data sets are also available as train data and test data. This raw data may contain missing values, redundant values, errors, incorrect values etc. To get correct results, data must be cleaned and errorless. The data is gathered from different sources and combined into a common pool in data warehouse.

#### II. Data Preprocessing

Preprocessing is the method of data mining where raw data is converted into an understandable format. Raw data contained in data warehouse is generally inconsistent, incomplete where some attributes are important and some are not

important. To focus on important attributes and to clean the data by removing missing values and filtering the data preprocessing helps so that it can be in a format accepted by the churning process. Preprocessing performs data cleaning, Data integration, Data transformation, Data reduction and Data discretization.

### III. Feature Extraction

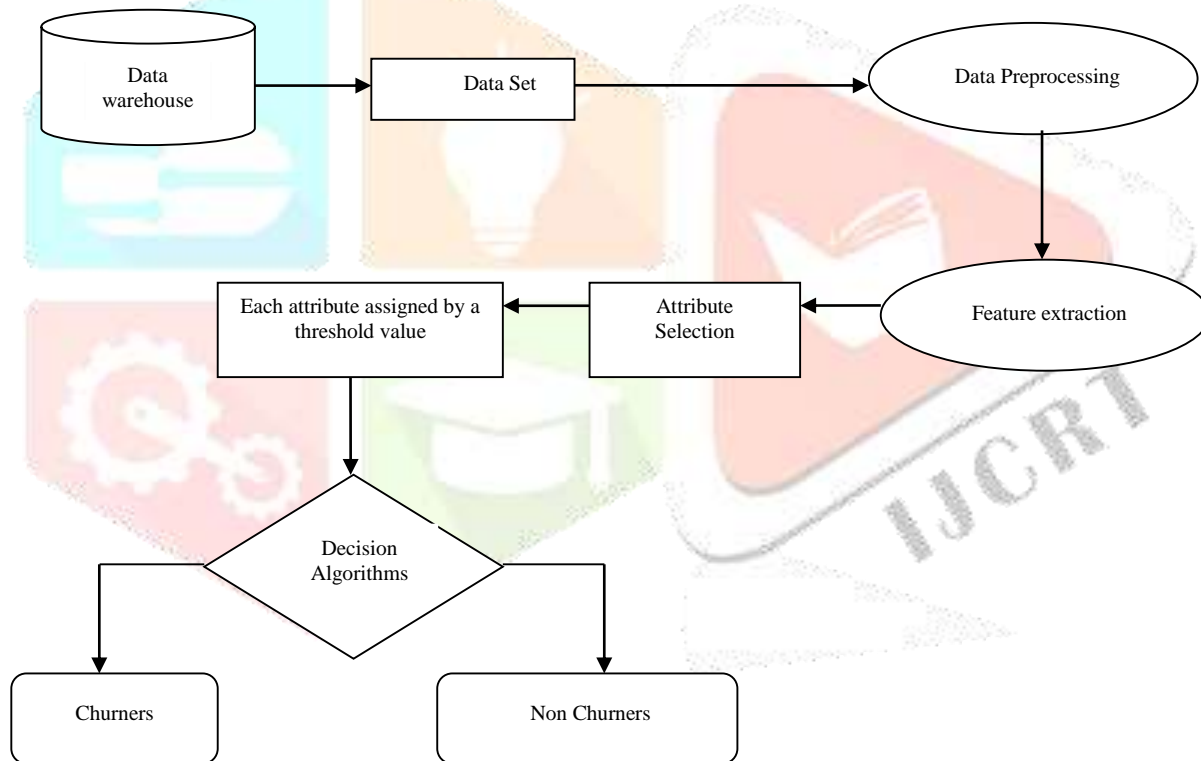
It is used to examine the performance of prediction models. Customer churn prediction involves prediction rates i.e. true positive, true negative, false positive and false negative. Decision Tree, Logistic Regression methods and other classification algorithms are helpful for building the churn prediction model. The test records counts predicted correctly and incorrectly are evaluated on the performance of a classification model. [7]

### IV. Implementation of Classification algorithms

Classification is a task of analyzing data correctly. Classifiers used to classify the raw data with the help of classification rules for various aspects. Classification for prediction can be depending on some measures such as accuracy, speed, robustness, scalability, interpretability etc. Traditional classification algorithms and soft computing algorithms comparison provides more accuracy in churning process.

### V. Result Verification

- I. Classification algorithms generate confusion matrix which can focus on true churners and false churners.
- II. ROC curves help to understand the gap between sensitivity and specificity.
- III. The closer the curve follows the left-hand border and then the top border of the ROC space, the more Accurate the test



**Fig1. Churn Prediction Framework**

### Conclusion

A conclusion section must be included and should indicate clearly the advantages, limitations, and possible applications of the paper. Although a conclusion may review the main points of the paper, do not replicate the abstract as the conclusion. A conclusion might elaborate on the importance of the work or suggest applications and extensions. As telecommunication industry always suffers from high churn rates because other companies offers better plan than current company. So customers churn to other service providers. This may lead to revenue loss. Customer churn prediction is the need of telecommunication industry to gain new customers as well as hold loyal customers. This can be archive by examining behavioral patterns of customer. This paper reviews various methods and algorithms for customer churn prediction which will give more accuracy. Comparison of traditional classifiers verses hybrid classifiers provides more efficiency for churning.



## References

- [1] Thomas Verbraken, Student Member, IEEE, Wouter Verbeke, and Bart Baesens "A Novel Profit Maximizing Metric for Measuring Classification Performance of Customer Churn Prediction Models" **MAY 2013,pp 961**
- [2] Ms Nisha Saini " Churn Prediction in Telecommunication Industry Using Decision Tree" **June 2016**
- [3] Dr. Mamta Madan, Dr. Meenu Dave, Vani Kapoor Nijhawan " Data Mining for Telecom Customer Churn Management" **September 2015, pp. 813-817**
- [4] Georges D. Olle Olle and Shuqin Cai "A Hybrid Churn Prediction Model in Mobile Telecommunication Industry" February 2014, PP 55-62
- [5] V. Umayaparvathi ,K. Iyakutti " Applications of Data Mining Techniques in Telecom Churn Prediction" **March 2012, PP Page 1065-1070**
- [6] Vani Kapoor Nijhawan ,Mamta Madan, Meenu Dave " The Analytical Comparison of ID3 and C4.5 using WEKA" **June 2017**
- [7] Kiran Dahiya, Surbhi Bhatia "Customer Churn Analysis in Telecom Industry" **2015 IEEE** [8] Yong Liu, Yongrui Zhuang "Research Model of Churn Prediction Based on Customer Segmentation and Misclassification Cost in the Context of Big Data" **June 2015**[9] Clement Kirui, Li Hong, Wilson Cheruiyot and Hillary Kirui "Predicting Customer Churn in Mobile Telephony Industry Using Probabilistic Classifiers in Data Mining" **March 2013**
- [10] Dayanand B.Gore, Seema S. Kawthekar, Vipin Y. Borole " Ear Recognition Techniques for Biometrics Using Digital Image Processing: A Review" **February 2017,PP 22-26**
- [11] AMMAR A.Q. AHMED, MAHESWARI "Churn prediction on huge telecom data using hybrid firefly based classification" **March 2017**
- [12] Adnan Anjum, Adnan Zeb, Imran Uddin Afridi, Pir Masoom Shah, Saeeda Usman, " Optimizing Coverage of Churn Prediction in Telecommunication Industry" **2017**
- [13] Sam Drazin and Matt Montag "Decision Tree Analysis using Weka"
- [14] HelenTreasaSebastian\* and RupaliWagh " Churn Analysis in Telecommunication using Logistic Regression" **March 16, 2017, Pgs. 207-212.**