

# Survey on Prediction of Customer Churn Analysis in a Telecommunications Industry

<sup>1</sup>Shreya Monani, <sup>2</sup>Shubham Gupta, <sup>3</sup>Shubhangi Jain, <sup>4</sup>Vidushi Mishra, <sup>5</sup>Prof. Artika Singh

<sup>1,2,3,4,5</sup>NMIMS Mukesh Patel School of Technology Management and Engineering, Mumbai, Maharashtra, India.

**Abstract :** In this paper, we study and discuss the customer retention strategies in a telecommunication industry. The goal of this study is to apply various analysis techniques along with data mining techniques to predict customer tenure and churn patterns by using data from a telecommunications company. These techniques include survival analysis or churn analysis. The study will help telecommunications companies to understand customer churn risk and customer churn hazard in a timing manner by predicting which customer is likely to churn along with what strategies can be adopted to stop a particular customer from churning, which helps optimize their customer retention and resources in their churn reduction efforts.

**Keywords—** Churn analysis Customer churn, Customer retention, Data mining techniques, Survival analysis Telecom industry.

## I. INTRODUCTION

While selecting a service provider in a telecommunications industry, there are multiple services to choose from and customers actively switch from one provider to another. There is a cutthroat telecommunications market wherein the customers demand better services day by day, at lower prices. This makes retention of customers over a maximum tenure the focus of all service providers.

A company experiences an overall of 30% annual churn rate and it is very expensive for the company to recruit new customers. Therefore, the strategy followed by the service providers is to maintain the already existing customers, rather than recruiting new ones. Through this study, if the telecommunications companies know which users are at high risk of churning and when they will churn, the service providers will be able to design customized customer plans and treatment programs for a set of customers in a timely efficient manner.

Many aspects may influence a customer to churn. Prepaid customers are not constrained by service contracts and so they churn for this reason. Hence, it is difficult to predict their churn rate. Another factor is customer reliability and sincerity, which can be determined by service and outcome quality given by the telecom industry. Issues like network coverage and reception quality may influence customers to move to the rival companies. Other factors that increase probability of customers churning include slow or faulty response to complaints and billing errors. Factors such as encasing prices, inadequate features, and primitive

technology may also cause customers to switch to the competitors. Customers often compare their providers with others and churn to whoever they feel provides better overall value for money.

To make the directing process easier, it is important to come up with robust tools of data analytics that can be important in the eradication of hidden patterns and knowledge in the database of the telecom industry. There are many statistical and data mining modules and techniques, which can be used for analyzing and predicting customer, churn. Some of these data mining techniques are logistic regression, decision trees; k means cluster method, and neural networks, etc. Survival analysis and churn analysis diagnose the customers who might be at high risk of churning and how to prevent that from happening. Solutions also include construction of an integrated model comprising of data mining approach and survival analysis to segregate the customers into a group of heterogeneous clusters by prediction of churn trend, which aid the service providers to retain customers by analyzing the churn trend of each customer cluster.

These techniques lead to several actionable Key Performance Indicators that are worth to continuously track the customers. These include the Average Customer Lifetime, Total Customer Value by taking feedbacks and Next Most Probable Exit-Point.

## II. LITERATURE REVIEW

1. Churn: A churn is defined as customer attrition or loss in a telecom industry when customers terminate their contracts or their usage and switch to another service provider.

2. High Valued Customers: These are the type of customers that share important service features like higher bills, long-distance transitions, etc. They spend more than the average rate and their expectations from the service providers is high, which make the retention of such customers more important than the rest as they are beneficial to the company. The company must allocate more resources to them to decrease the churn rate.

3. Customer Relationship Management (CRM): Customer churn prediction has become important to the telecom industry to protect their customers based on their loyalty to the company for their service given. Hence, it is vital for all companies to improve their Customer Relationship Management. It involves knowing the customer's attainment so it can retain the most profitable customers and identify those whose churn no longer makes any difference to the company. It aims at better advisement and customized plans for their high valued customers in order to retain them from churning to a competitor.

## III. DATA MINING TECHNIQUES

Data mining techniques are the techniques that are used for automated analysis of massive data sets. Data mining is defined as finding the hidden information from a database. The other names of data mining are Exploratory Data Analysis, Data Driven Discovery and Deductive learning. There are different algorithms to perform data mining techniques. A data algorithm consists of three parts:

**Model:** The data should fit into a model is the main purpose of the algorithm.

**Predictive Model:** The prediction about values of data is done using known results found from different data.

**Descriptive Model:** The patterns or relationships in data are identified by Descriptive Model. A descriptive model is used to explore the properties of the data that are examined.

There are different types of data mining techniques that are used:

### 1. Decision Tree:

It uses a tree-like graph. It takes into consideration all the possible consequences and makes a graph accordingly [1]. The most difficult part in decision tree algorithm is the creation of the tree.

There are two types of methods that can be used in decision tree algorithm:

### 1. Balanced[1]:

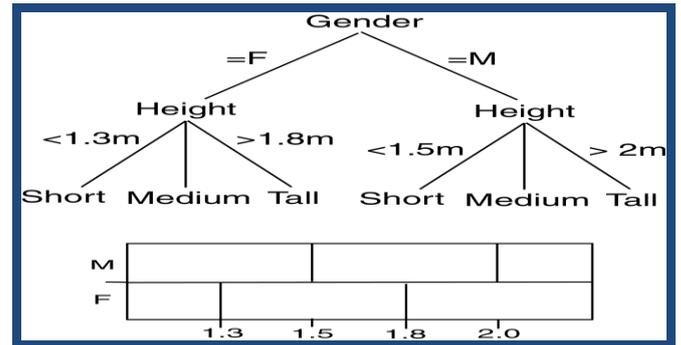


Fig 1.1

### 2. Deep:

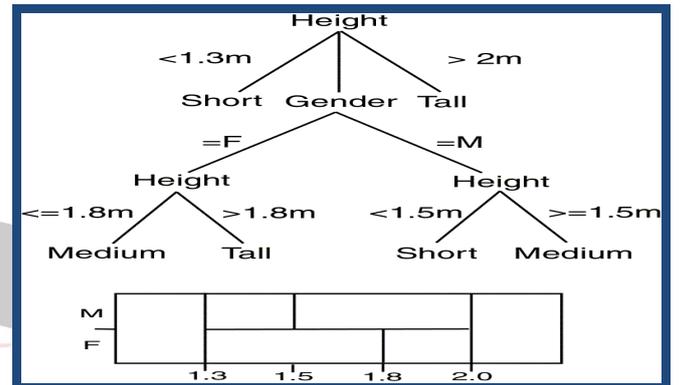


Fig 1.2

Advantages:

1. They are very easy to understand.
2. Rules can be easily generated.

Disadvantages:

1. It may suffer from over-fitting.
2. It involves classification by rectangular partitioning.

### 2. Neural Networks:

It is a technique that is based on the functioning of human brain [1]. Neural network is used in pattern recognition, speech recognition, computer vision etc.

Advantages:

1. Learning can be continued even after training set has been applied.
2. It has easy parallelization

Disadvantages:

1. It is difficult to understand
2. It may suffer from over-fitting

### 3. K-Means:

K-means clustering is a method of vector quantization, originally from signal processing, popular for cluster analysis in data mining. The k-means algorithm is implemented in four steps:

1. Partition objects into k nonempty subsets.

2. Compute seed points as the centroids of the clusters of the current partition (the centroid is the center, i.e., mean point, of the cluster)
3. Assign each object to the cluster with the nearest seed point
4. Go back to Step 2, stop when no more new assignment

#### Advantages:

1. If variables are huge, then K-Means most of the times computationally faster than hierarchical clustering, if we keep k small.
2. K-Means produce tighter clusters (especially if the clusters are globular), than the hierarchical clustering method.

#### Disadvantages:

1. Difficult to predict K-Value.
2. With global cluster, it didn't work well.

#### 4. Time Series Analysis

It is an ordered sequence of values of a variable at equally spaced time intervals.

The usage of time series models is twofold:

- Understanding the underlying forces and structure that was used to produce the observed data.
- Fit a model and after that proceed to monitoring, forecasting or even feedback and feed forward control.

It comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data. Time series forecasting is the use of a model to predict future values based on previously observed values.

#### Advantages:

1. It is useful for identifying seasonal variables which can help planning different times of years.
2. It can be reasonably accurate in short term if firm is in a stable environment.

#### Disadvantages:

1. Long, winding and complex process.
2. Not very useful for long-term forecasting.

#### 5. Linear Regression

In linear regression, the relationships are modeled using linear predictor functions whose unknown model parameters are estimated from the data. Such models are called linear models. Like all forms of regression analysis, linear regression focuses on the conditional probability distribution of  $y$  given  $X$ , rather than on the joint probability distribution of  $y$  and  $X$ , which is the domain of multivariate analysis.

#### Advantages:

1. It is simple, interpretable, scientifically acceptable, and widely available.
2. It is the first method to be used for many problems.

## IV. AN INTEGRATED DATA MINING AND SURVIVAL ANALYSIS MODEL FOR CUSTOMER SEGMENTATION

This area focuses on the integrated model of data mining and survival analysis to help analyze customer churn and provide appropriate actions for enterprises to retain customer. It contains two parts, firstly, clustering of the customers into heterogeneous clusters using data mining techniques, based on their survival characters. Secondly, prediction of cluster's survival and hazard function using survival analysis and also to test its correctness for customer segmentation. This model was applied in a dataset from one of the best [3] telecommunications company from china. Customer segmentation is used to find out; how to maximize the value of customer, which is evaluated by researchers and managers both to generate different strategies.

Normally, customer segmentation process includes experience description, traditional statistical, and non-statistical methods (Per Vagn Freytag, et al, 2001, Lei-da Chen et al., 2000, E. H. Sub et al., 1999). Non-statistical methods mainly are the applications of data mining techniques in segmentation (Agnes Nairn, and Paul Bottomley, 2003, Verhoef P.C, et al., 2003, Jon Kleinberg, et al., 2004). Jaesoo Kim etc (2003) researched the application of ANN in tour customer segmentation. Fraley, C. and Raftery, A.E (2002) researched the application of clustering approaches in customer segmentation.

Data Mining is one of the application of predictive and descriptive analytics for large datasets to support different functions like sales, marketing, supply chain, service, finance etc. It works in number of ways like prediction, estimation, clustering etc. It is used in almost every industry today. It uses two approaches: directed and undirected; using input data to predict a specified output and exploration of data sets to check what can be learned from them, respectively.

Survival analysis is a field of statistics to model any population whose individuals move between a set of well-defined states. It can be applied to study the behavior of customers in a telecommunication company.

#### Model Implementation Steps:

This model consists of clusters and survival analysis, as shown in the below figure.

The customer cluster is extracted according to the industry and also the customer behavior attributes. Take a telecommunication company as an example, customer's calls hours and expense in recent few months are widely used for churn prediction and they include some steps.

Steps are:

1. Mission defining: Approving the customer attributes for the mission.
2. Data preparation and Data extraction: Extracting the important attributes from the data warehouse.
3. Clustering: Using data mining techniques such as K-means, where the cluster customers are created based on similar survival/ hazard possibility.
4. Appending a new attribute: For e.g., cluster Number, after clustering the attributes, each customer will have a new attribute named cluster number, which is necessary for survival analysis.
5. Survival analysis: It requires only three important attributes, namely: Months with Service, Clustering Number and Customer Number.
6. Finding out of the different patterns for clustering.
7. Successful Implementation of the action of customer retaining.

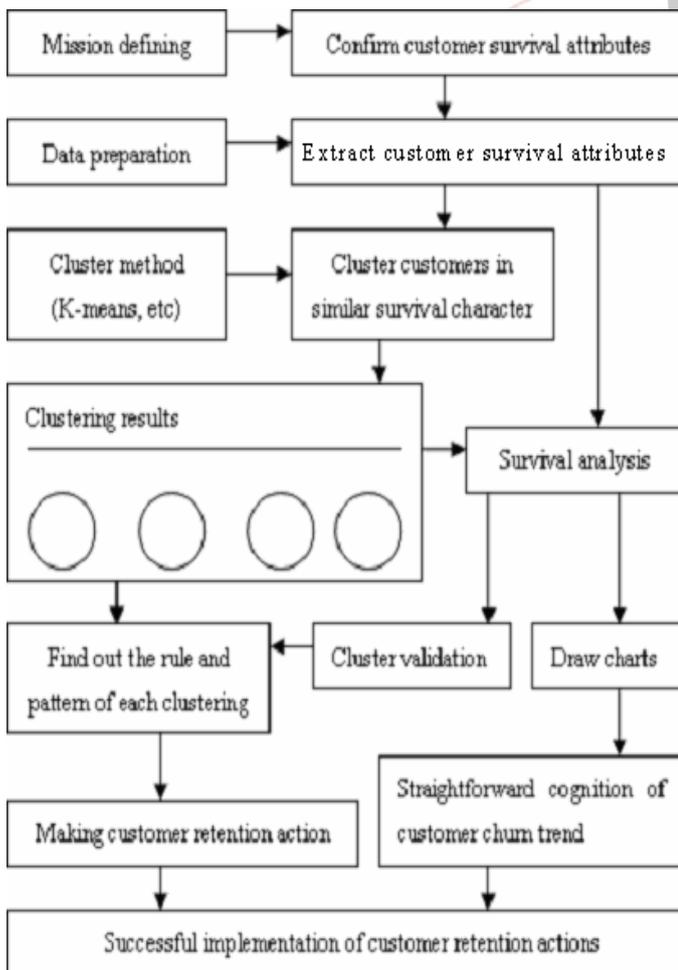


Fig 1.3: Customer Segmentation

## V. SURVIVAL FUNCTION

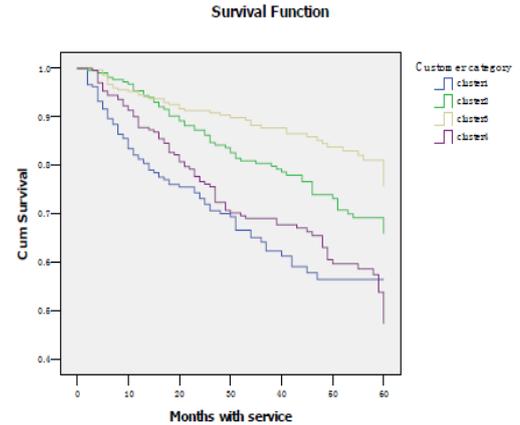


Fig 1.4: Survival function

Survival function maps a set of events, onto time which calculates the probability. The estimation of survival and hazard function helps us to obtain customer churn and survival information, which uses few variables like months, Customer ID etc. [3]

## VI. CLUSTER FEATURES

class	
1	More fees, more long-distance transactions, highest churn
2	Normal fees, second lowest churn
3	Little transaction each month and more note, lowest churn
4	More transactions each month, more call in and less call out, second highest churn

Fig 1.5: Types of Clusters

### Clustering 1:

These customers share important features such as more long distance, more fees etc. Their expectations are high because they expend more, so we must hold on to them to decrease the churn.

### Clustering 2:

These customers have normal fees hence the churn is comparatively low. The company must provide appropriate resources to hold on to them, as they satisfy the company's service.

#### Clustering 3:

These customers' expenditures are lowest and hence the churn is the lowest comparatively. They satisfy the company's service, but the contribution is very low hence the companies needn't provide the resources to them as they hardly effect on churn analysis.

#### Clustering 4:

These customers have higher spending with many calls, hence they are the most valuable customers, and also the customer churn for these clusters is high. Companies must provide some resources to them to retain the customers.

For a company, the way to use data mining, and also how to practice its tactics, and should we use it in determining the segmentation? Hence, to answer these questions, this field proposes an Integrated Data Mining and Survival Analysis Model for Customer Segmentation. This model clusters the customers into heterogeneous clusters, where each clustering has its unique churn rate, so the company can take retention action according to this. After observing the sharp of the survival function curve, company can get straightforward understanding of the customer churn rate. This model can be used in other industry such as finance service etc. [3] Therefore, the future researchers can focus on testing this model in other services.

## VII. CONCLUSION

Customer churn has been identified as a major problem in Telecom industry and aggressive research has been conducted in this by applying various data mining techniques. Many data mining techniques are generally applied in customer churn. On reviewing literature we found that neural network outdo most of methods in terms of accuracy. Survival analysis can be applied to check the probability of a customer churning at a particular time.

There are likely to be tremendous rates of research in data mining and their applications in customer churn, but still it is an active research field and researchers are searching for more accurate solutions. However from the literature survey it is evident that there has been little research work on survival analysis and their applications in customer churn. Our future work will be applying survival analysis on telecom datasets and combine the result with some of the most commonly used predictive analysis techniques in churn prediction as they are very suitable tools for data mining.

## REFERENCES

- [1] Dr. M.H Dunham, *Data Mining Introductory and Advanced Topics*, Prentice,2002
- [2] Kiran Dahiya, Surbhi Bhatia, *Customer Churn Analysis in Telecom Industry*.
- [3] Guozheng Zhang, Yun Chen, *An Intergrated Data Mining and Survival Analysis Model for Customer Segmentation*, College of Business, Houzhou Dianzi University, P.R.China, 310018, School of Public Economy Administration, Shanghai University of finance
- [4] Amal M. Almana, Mehmet Sabih Aksoy, Rasheed Alzahrani, *A Survey On Data Mining Techniques In Customer Churn Analysis For Telecom Industry*, Department of Information Systems, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia, ISSN : 2248-9622, Vol. 4, Issue 5(Version 6), May 2014, pp.165-171
- [5] Simon Hefti, *Improved Customer Understanding with Survival Analysis*.
- [6] Junxiang Lu, *Predicting Customer Churn in the Telecommunications Industry — An Application of Survival Analysis Modeling Using SAS*, Ph.D. Sprint Communications Company Overland Park, Kansas, Paper 114-27
- [7] Rud, Olivia Parr, *Data Mining Cookbook*, New York: John Wiley & Sons, 2001.