International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 02 | Feb 2020 www.irjet.net p-ISSN: 2395-0072

Customer Churn analysis in Telecom Industry

Venkatesh Hariharan[1], Irshad Khan[2], Pravin Katkade[3], Prof. Sejal D' mello[4]

[1,2,3]Student, Department of Information and Technology, Atharva College of Engineering, Mumbai, Maharashtra, India.
[4]Professor, Department of Information and Technology, Atharva College of Engineering, Mumbai, Maharashtra, India

---

Abstract - Telecom industry is rapidly growing day by day. Customer is always is looking for better services and scheme. Churn analysis helps to identify customers who are probably to cancel a subscription to a service. Telecommunication companies now focus on to have a more extensive customer base; therefore, retaining existing customers becomes a considerable challenge. It is essential to know which customers are likely to cancel the subscription. In this scenario, data plays a critical role in getting knowledge of customer behavior. With the help of some Algorithm and information, we can predict possibilities of customer churn and improve their service to change the churn decision

Key Words: Telecom, CRM, Retention, Customer Serive, Usage detection, Logistic Regression algorithm, Decision Tree algorithm, Recommender system (TCPRS).

1. INTRODUCTION

Customer churn refers to that customer switch over from one service provider to another service provider. Churn is a challenging problem for a subscription services provider. There are many Telecom Companies are coming with various schemes and services, which will help that company attract numerous customers to switch from competitor's service to their service. The main goal of those companies to retain their existing customers. Those companies need to know in advance, which customers may switch from their service to another service.

Churn prediction model can be used to analyze the historical data available within the business organization to identify the customers who are at high risk of churning; this will assist the telecom industry to get their attention on developing retention strategies on those customers. Individualized customer retention is severe because businesses usually have a big customer database and cannot afford to spend money and time. If we could predict in advance which customers are at risk of leaving, then we can lower the customer retention efforts by directing our attention toward such customers.

Now, this is where the churn prediction model can help the business to identify such high-risk customers and thereby helps in maintaining the existing customer base and increase in revenues.

A churn prediction model is famous because of the getting new customers is much costly than retaining the existing one. Retention has become necessary because of the increasing number of service providers and the competition between them, where everyone is trying to attract new customers and give them exciting to offer them to switch to their service.

Using Data mining techniques on customer database helps to know the customers that have a high probability to churn based on the historical records available. The data mining techniques can be helpful in finding a pattern among the already churned customers. The rest of this paper is represented as follows. Section 2 discusses related research in churn analysis. Section 3, represents the system design and personalized recommendation system for telecom and services, TCPRS, using a hybrid approach that combines item-based and user-based collaborative filtering methods. A complete analysis of performance of the TCPRS is shown in Section 4.

Problem Statement

Customer churn refers to when a customer switches from one service provider to another. Churn is a problem for any provider of a subscription service or recurring purchasable. This is where the churn prediction model 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.

2. RELATED WORK

A lot of research has been done in various industries for the retention of customers to develop and build an efficient model so that a group of customers can be retained. Different types of data mining techniques have been used for churn prediction, of which some great techniques include Bayesian Models, Regression models, Neural Networks, Decision trees, Clustering, SVM, etc. The goal of the research was to show that the model built on data mining techniques can explain the churn behavior with more accuracy and that in some of the models, the reason for churn can be identified. They used Logistic Regression in parallel with Voted Perceptron for classification and then combined with clustering for churn prediction.

Qureshi et al. [2] in their examination took active churners in the Telecom industry into account by applying various data

mining such as K- Means Clustering, Logistic Regression, Neural Network, Linear. In [5], the authors have used neural networks and decision trees for churn prediction. They have described the process of churn prediction right from data collection of churn analysis. They have also focused on pre-processing of data in which data cleaning, data transformation, data reduction feature abstraction is used before giving it as input to the algorithms. In these, they have proposed that selecting the fixing the proper threshold values and the right combination of attributes may produce more accurate results. Their study limits the prediction of churn, and there are no steps analyzed to include retention policies. A recent review based on the techniques used for churn prediction [3] Decision tree-based techniques, Neural Network-based techniques, and regression techniques are applied in customer churn. Hence as seen in the literature survey, the most likely methods used for churn prediction include logistic regression, neural networks, and decision trees, out of which we have proposed to study the performance of logistic regression on telecom data publicly available.

Neural networks also perform very well in case of churn prediction. It has limitations such as it shows well for massive datasets only and takes a lot of time for training even with small datasets. Logistic regression is used, which helps analyze the factors causing churn in an effective and understandable way.

Decision tree provides a graphical overview of the available data from which strategies can be built for customer retention, and logistic regression helps to understand each feature affects the decision of churn. Decision trees and logistic regression has been used for churn prediction for credit card customers. The study, as well as the comparison of these two algorithms, has not yet been.explored for the telecom industry. One is the data mining technique, and the other is a statistical model that can provide an idea to the telecom industry about which algorithm suits their data.

3. PROPOSED SYSTEM

In the proposed system, python programming will be used in our churn prediction. It is used for statisticians as well as data mining. The system will have three main options, namely View, analysis, and performance, which displays the results obtained by applying on the available dataset — testing to make a list of customers which have a high probability to churn, given that the attributes of the input data are same as the available dataset used for training. Training and testing which create a model. In performance analysis, the results after using logistic regression on the available dataset are illustrated using confusion matrix analysis. In operation, the user can provide data for testing the system provided the data are the same as that used for training using the publicly available dataset.

If the user wishes to use the system for their data of different formats, then they can do so by using the training and testing option where the system will use the new training data to build the model first and then use it for testing. Hence the system is not limited to use only for some specific type of data. The data cleaning method will be included in the training and testing phase explicitly to improve system efficiency. The system is made usable through a web interface that will graphically explain the results of the system.

The system-building has phases, i.e., developing the web feature extraction, interface module, module, and prediction module. The web interface will provide a graphical representation of results, which can be created. The feature extraction module will consist of the estimation of parameters in logistic regression. Logistic regression uses maximum likelihood estimation for transforming the dependent variable into a logistic variable. Logistic regression uses the linear regression function to estimate the value of the dependent variable by module and prediction module. The web interface will provide a graphical overview of the results obtained, which can be created. The feature extraction module will consist of the estimation of parameters in logistic regression. Logistic regression uses maximum likelihood estimation for transforming the dependent variable into a logistic variable. Logistic regression uses the linear regression function to estimate the value of the dependent variable by estimating the parameters for the linear equation.

As shown in (1) α, b1, b2...bn are the parameters to be calculated using the training data and the equation will then be used to predict P(X) which represents the dependent variable value if values of features x1,x2,…xn are given in Equation 1

$$Y = B0 + B1X1 + B2X2 + \ldots\ldots + BkXk + \epsilon \quad \ldots\ldots(1)$$

Y stands for the dependent variable that needs to be predicted. β0 is the Y-intercept, which is the point on the line which touches the y-axis.β1 is the slope of the line (the hill can be negative or positive depending on the relationship between the dependent variable and the independent variable.)X here represents the independent variable that is used to predict our resultant dependent value. Logistic regression can be used to model binary dependent variables. The structure of the logistic regression model is designed to produce binary results. Least square regression is not built for binary classification as logical regression performs a better job at classifying data points and has a better logarithmic loss function as opposed to least squares regression.

$$Q(X) = \frac{1}{1 + e^{-(a + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n)}}.$$

……(2)

As the value of Q(X) is a number between 0 and 1, this value can be considered as the probability of the particular outcome. For example, if the output is 0.8, it means there are 80% chances of getting the production as one, and it can, therefore, be safely predicted that for the given set of input attributes, the production would be 1. The prediction for the negative case or 0 can be calculated as one minus the probability of a positive occurrence. For the above example, it will be 0.2, which is far less, and hence the prediction is given to be as 1. Typically, 0.5 is used as a threshold to decide what prediction is to be delivered. Any value above 0.5 is predicted as a positive case and anything below as a negative case value.

Figure 1 provides an overview of the proposed system

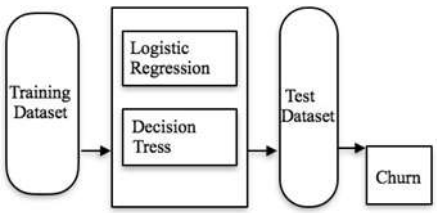

Fig 1: System Architecture

The churn prediction system will build the prediction model based on the logistic regression algorithm, which will then be used on the testing dataset to measure the accuracy of the system. Finally, they will output the list of customers that have high probability and are possible churners in the future, along with the efficiency of the system. The input to the method includes the options, training data, and test data where Option comprises of View Performance, Train and Test, Test. Training Dataset Attributes to contain State, Area Code, Phone, Day Mins, Night Mins, International Mins, Customer Service Calls, International Plan, Voicemail Plan, Day Calls.

3.1 Recommender System

There will be a large number of web services and telecom services provided by third-party providers and or end-users. Then, how to recommend a minor service set that is related and interesting for a particular user is a significant problem that needs to be solved because it will ease the distribution of services to the most suitable users.

Recommender system has become an vital research area in the last decade, and there has been much work done in the industry and academic center. [6]. Ideally, recommender systems are classified into three types [6]:

1. **Content-based recommendations:** This method uses only information of the description and attributes of the item users have previously consumed to model user's preferences. In other words, these algorithms try to recommend things that are similar to those that a user liked in the past. Various items that were previously rated by users are distinguished and recommended to the consumer.

2. **Collaborative recommendations**: This method makes automatic predictions about the interests of a user by collecting preferences or taste information from many users. The underlying assumption of the collaborative filtering approach is that if person A has the same opinion as a person B on a set of items, A is more likely to have B's idea for a given object than that of a randomly chosen person.

3. **Hybrid approaches:** Recent research has demonstrated that a hybrid approach, combining collaborative filtering and content-based filtering, could be more effective than simple approaches in some cases. Cold start and Sparsity problems can be solved using these methods.

3.2 Algorithm Description

1. **User-based collaborative filtering**: In this model, products that are liked or preferred by other consumers are recommended to the user. For example, if Derrick and Dennis like the same movies and a new movie come out that Derick wants, then we can recommend that movie to Dennis because Derrick and Dennis seem to like the same film.

2. **Item-based collaborative filtering**: These systems is used to identify similar items based on previous users' ratings. For example, if users 1, 2, and 3 gave a 5-star rating to books A and B, then when a user 4 buys book B they also get a recommendation to purchase book A because the system identifies book A and B as similar based on the ratings of users 1, 2, and 3.

4. RESULT OF IMPLEMENTATION

Logistic regression is a statistical method for predicting binary classes. The outcome or target variable is binary in nature. It computes the probability of an event occurrence. A logistic regression predicts the probability that an observation falls into one of two categories of a dichotomous dependent variable based on one or more independent variables that can be either continuous or categorical.

Logistic Regression predicts the probability of occurrence of a binary event utilizing a logit function. The dependent variable in logistic regression follows Bernoulli Distribution. Estimation is done through maximum likelihood.

4.1 CHURN PREDICTION MODEL

Precision is about being precise i.e., when a model makes a prediction, how often it is correct. In our prediction case, the logistic regression model predicted the outcome with 80% accuracy.

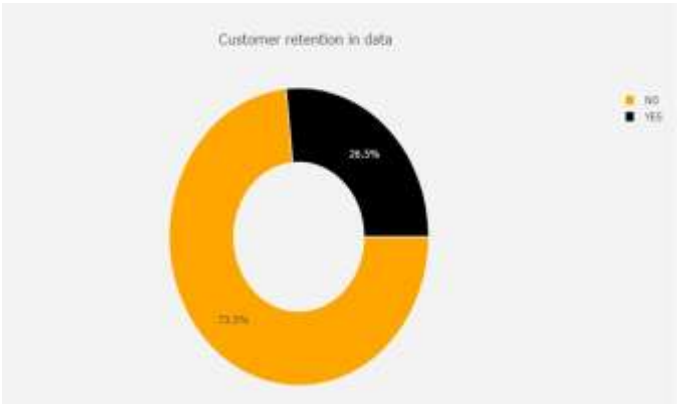

Fig 2: Churn Prediction

Above Fig 2 portrays customers who are at the risk of churning to a new service provider According to Fig 2, 73.5% of customers will not churn and will continue using the current service provider, while the remaining 26.5% are at a high risk of churning and hence proper counter measures must be taken in order to retain them.

4.2 IDENTIFYING CHURNING CUSTOMERS

```
Enter Customer ID 7156-MHUGY
Top 10 similar customers like 7156-MHUGY  are:

The Customer with ID 3887-PBQAO wont churn
The Customer with ID 7596-IIWYC wont churn
The Customer with ID 3849-NDXFL Will churn
The Customer with ID 7156-MHUGY Will churn ---User Input Customer

The Customer with ID 6922-NCEDI wont churn
The Customer with ID 3134-DSHVC wont churn
The Customer with ID 7762-ONLJY wont churn
The Customer with ID 9770-KXGQU wont churn
The Customer with ID 9752-ZNQUT wont churn
The Customer with ID 1577-HKTFG wont churn
The Customer with ID 5868-YWPDW Will churn
```

Fig 3: Identifying churning customers

Above Fig 3 represents the required customer's churning status, along with 10 other customers who are similar to the required customer based on feature selection. According to Fig 3, customer no 7156-MHUGY will churn, but some customers who are similar to the customer no 7156-MHUGY may not churn to a new service provider like customer no. 7596-IIWYC.

4.3 RECOMMENDER SYSTEM

```
Enter Customer ID 7156-MHUGY

Most suited plan for customer with ID: 7156-MHUGY is:['RS 98 ' '28 DAYS' '2  GB' '2 GB/Month']
REASON:Customer Is a Senior citizen,most likely he/she  will need less datapack plan
```

Fig 4: Recommender System

Above Fig 4 represents the churning customer a suitable plan based on his data usage and attributes, above customer is a senior citizen and based on his usage statistics, Rs 98 plan is suggested as the customer doesn't require much internet.

5. CONCLUSION

This system gives a prediction about customer churn alignment and fixing the proper threshold values, and the right combination of attributes may produce more accurate results of predicting churn customers. This system model suggests that data mining techniques gives solution for the customer churn management. Using this model, the telecom companies can predict in advance which customers are at risk of churn, and can target those customers and can recommend them suitable recharge plan. This system saves lots of revenues and time, which is used for replacing the already lost customers and also the ones that are already loyal customer.

ACKNOWLEDGEMENT

We express our gratitude to **Prof. Sejal D'mello**, our project guide who gave us all the guidance and motivation and at any time available to us for doubts solving, grateful for giving us the required help, point by point recommendations and furthermore support to do the venture. We are excited and happy to offer our thanks to the Head of the Information Technology Department **Prof. Deepali Maste**, for her endorsement of this undertaking. We also want to profoundly offer our genuine thanks to our regarded head **Prof. Dr. Shrikant Kallurkar** and the administration of Atharva College of Engineering for giving such a perfect environment to develop this undertaking with well-furnished library with all the most extreme important reference materials and forward-thinking IT Laboratories We are amazingly appreciative to all staff and the administration of the school for giving every one of us the offices and assets required.

REFERENCES

[1] A Hybrid Churn Predictio Model in Mobile Telecommunication Industry, Georges D. Olle and Shuqin Cai, International Journal of e-Education, e-Business, e-Management and e-Learning, Volume 4, Issue 1, February 2014.

[2] Telecommunication Subscribers' Churn Prediction Model Using Machine Learning, Saad Ahmed Qureshi, Ammar Saleem Rehman, Ali Mustafa Qamar, Aatif Kamal, Ahsan Rehman, IEEE International Conference on Digital Information Management (ICDIM), 2013 8th International Conference on, Volume 4, Issue 5, April 2013.

[3] Customer Churn Prediction in Telecommunication Industries using Data Mining Techniques- A Review, Kiran Dahiya and Kanika Talwar, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 4, March 2015.

[4] Churn Prediction in Telecommunication Using Classification Techniques Based on Data Mining: A Survey, Nisha Saini and Monika, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 3, March 2015.

[5] Churn Prediction In Mobile Telecom System Using Data Mining Techniques, Dr. M. Balasubramaniam, M.Selvarani, International Journal of Scientific and Research Publications, Volume 4, Issue 4, April 2014.

[6] A personalised recommender system for Telelcom products and services ,Zui Zhang, Kun Liu, William Wang, Tai Zhang and Jie Lu, Volume 5, Issue 4,March 2016