IMPACT: International Journal of Research in Humanities, Arts and Literature (IMPACT: IJRHAL)
ISSN (P): 2347-4564; ISSN (E): 2321-8878
Vol. 7, Issue 5, Apr 2019, 573-580

A TWO STEP HYPOTHETICAL CHURN MODELLING AND PREDICTION MODEL

Namrata S. Gupta[1] & Bijendra S. Agrawal[2]

[1]Assistant Professor, Smt. BK Mehta IT Centre (BCA College), Palanpur, Gujarat, India

[2]Research Scholar, Rollwala Computer Centre, Gujarat University, Gujarat, India

Received: 17 Apr 2019 **Accepted: 24 Apr 2019** **Published: 30 Apr 2019**

ABSTRACT

In the age of wireless communication, the term churn is arising due to facility race in mobile phone companies. Churn means the movement of the customer from the existing company for better services which are the migration of customer from one service provider to another. At present the Telecommunication Company or market, the struggle is on their extreme and the products and offerings are more and more analogous. This activity gives a direct loss to the company. In that context, necessary action and step can be taken if the reason behind it or churner may be predicted before leaving the services. So there is a need to understand and simplify the model to deal with churn problem. This paper gives two-step churn prediction model which tries to design a simple methodology to overcome such problem via data mining tools and process.

KEYWORDS: *Churn, Telecommunication, Algorithm, Predicted, Model*

INTRODUCTION

The churn prediction in the telecommunication companies is a typical task which cannot be sent percent predictable. By means of clutching and remain of possibly churning customers' has arisen to be as essential for service supplier as the attainment of new customers. Far above the ground churn rates and substantial revenue loss due to churning have turned correct churn prediction and prevention to a vital business process. Even though churn is inescapable, but it may be managed and kept at an acceptable level.

In general, there are many diverse conducts of churn prediction and novel techniques continue to emerge with the conventional statistical methods. High-quality prediction models have to be continually urbanized for the betterment. Valuable customers have to be identified, thus leading to a combination of churn prediction methods with customer lifetime value techniques. Here in the paper, a two-step method is proposed to contribute in the direction of solving the churn prediction problem.

REVIEW OF LITERATURE

We can be wrapping up on the basis of various models which elaborate on the importance of the work and suggest model as the extensions. In all Predictive model customer churn has been identified which is a major problem in the Telecom industry and hard-line research has been conducted with the support of the various data mining techniques. The core techniques of data mining Decision tree & its extensions, Neural Network based techniques and regression techniques are usually functional in customer churn. As of review and comparisons of the model and literature, it is observed that decision tree based techniques, particularly C5.0 and CART, have performed some of the existing data mining techniques such as regression in terms of accuracy. Despite this neural networks outdo the previous techniques due to the size of datasets used and diverse feature selection methods applied. According to this comparisons, data mining methods and their applications based predictive model for customer churn prediction will be the final outcome. So the proposed predictive model will be based on CART algorithms mainly.

Proposed Hypothetical Concept

In the proposed model there are two steps are proposed to generate telecom churn model including data pre-processing step:

- Defining Churn Algorithm and
- Constructing a Predictive Model.

To construct more accuracy churn model, we divide the huge data set into training data set and testing data set in data pre-processing step for constructing and refining churn models.

First, the data scoping includes problem and data understanding are need to define by experts. For example, customer churn problem including contract-end or number-portable customers may happen in some particular business, product, or customer segmentation. Meanwhile, the corresponding data need to meet each specific request via feature analysis methodology. Through a serious of discussion and analysis, experts may decide that historical billing, contract status, or call detailed data will be useful to construct a model.

Second is to set a time window for pre-processing raw data in different churn management problem. We accumulate raw data for 15 days to help training and testing, churn models. We utilize the training data to define time windows and measures and construct the churn predictive models. The second 15 days data set are then used to predict and verify the effectiveness of those models using the effectiveness measure and the result will be used to refine the churn models.

Constructing a churn prediction is the base for the study. Hence, we can use several suitable data mining methodologies and algorithm for measuring the performance, such as decision tree, SVM, Neural network, regression, clustering and so on, to construct churn models according to the appropriate data.

The Probable Model

The model developed in this research is based on using K-means clustering in two stages. At the first stage, data reduction is performed by applying K-means algorithm on the selected training dataset. This process will split the training dataset into a number of smaller sets (clusters). Clusters with churners and non-churners are classified in the first stage. At the second stage, Decision Tree Algorithms and various performance measure data mining algorithms are applied to the selected clusters in order to assess the performance and develop -predictive conclusion for the churn customers.

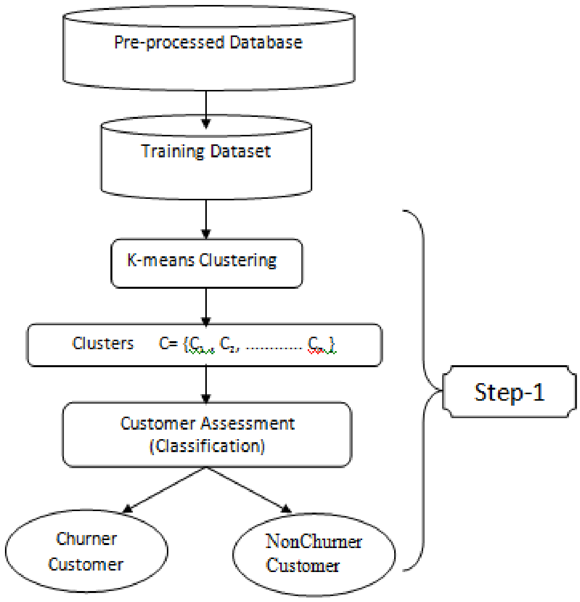

Figure 1: Model for Churn Prediction (Step-1)

Implementation Algorithm for Step-1

Read $C=\{x_1, x_2, x_3... x_n\}$

// C is cluster of customer at one location of the whole data warehouse

// x is customer with churn or non churn

Where $C=C_{nc} + C_{ch}$

nc= Non churner

ch= Churner

Read $C_{nc}=\{ x_{1,} x_{2,} x_{3,}............... x_k \}$

// set of the customer with non churning

$C_{ch}=\{ x_{k+1,} x_{k+2,}............... x_n \}$

// customer with churning

Here each record (x_1) of the cluster has set of 12 variables. Therefore one customer record can be show as:

$x_i=\{ v_{1,} v_{2,} v_{3,}............... v_{12} \}$

where

v_1 = Call Ratio

v_2 = Average Call Distance

v_3 = Last Call Date

v_4 = First Call of Customer

v_5 = Life Span

v_6 = Time Distance between two calls (Call Frequency)

v_7 = Number of Days for Specific Call

v_8 = Total Incoming Call

v_9 = Total Out going Call

v_{10}= Total Cost

v_{11} = Incoming Call Duration

v_{12} = Out going Call Duration

Read (x_i)

If (CN= "POC" OR "COC")

Switch (CF)

{

Case 1:

Call-Frequency >= 1day

CF= "NC"

Break;

Case 2:

Call-Frequency >= 2day

CF= "NC"

Break;

Case 3:

Call-Frequency >= 3days

CF= "POC"

Break;

Case 4:

Call-Frequency >= 7days

CF= "POC"

Break;

Case 5:

Call-Frequency >= 15days

CF= "COC"

Break;

Default:

CF= "POC"

$x_i = x_{k+i}$; i= {1, 2,................................n}

}

else

$x_i = x_{i+1}$;

i=i+1;

}

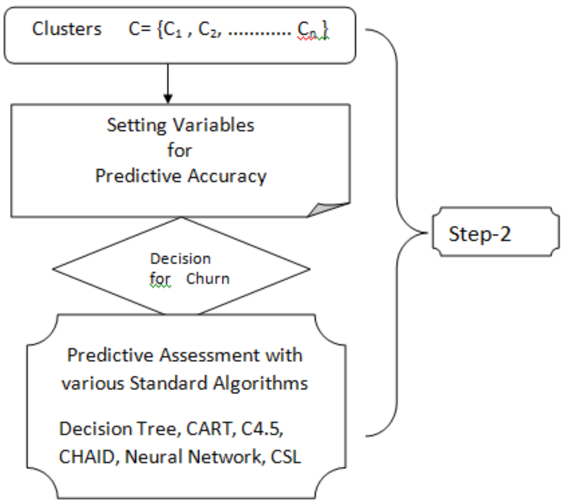

Figure 2: Model for Churn Prediction (Step-2)

Implementation Algorithm for Step-2

The value of Nonchurn customer and Churner customer are classified by considering subperiods of 15 days / day >7 in the two regular sets of observation.

Read $C_T=\{C_1, C_2, C_3... C_n\}$

// C_i = is cluster, where C_T is set of all the clusters

Read $C=\{x_1, x_2, x_3... x_n\}$

Read values from x_i, where ($x_i \in C$)

InMIN : Incoming Minute

InFRQ : Incoming Frequency

OtMIN: Outgoing Minute

OtFRQ : Outgoing Frequency

$$\nabla InMIN = \frac{InMIN\ (2) - InMIN(1)}{InMIN(1)} \ldots\ldots\ldots\ldots\ldots\ldots\ldots 1$$

$$\nabla InFRQ = \frac{InFRQ\ (2) - InFRQ(1)}{InFRQ(1)} \ldots\ldots\ldots\ldots\ldots\ldots\ldots 2$$

$$\nabla OtMIN = \frac{OtMIN\ (2) - OtMIN(1)}{OtMIN(1)} \ldots\ldots\ldots\ldots\ldots\ldots\ldots 3$$

$$\nabla OtFRQ = \frac{OtFRQ\ (2) - OtFRQ(1)}{OtFRQ(1)} \ldots\ldots\ldots\ldots\ldots\ldots\ldots 4$$

If ($\nabla OtFRQ < OtFRQ(1) \mid\mid \nabla InFRQ < InFRQ(1))$

&& ($\nabla OtMIN < OtMIN(1) \mid\mid (\nabla InMIN < InMIN(1))$

{

Churn = “COC”

Else

Churn = “NC”

}

By using above-mentioned features, as the basic input data for the decision tree are suitablefor the predictive model which tries to find out the “churn”.

Now different predictive methods/algorithm/ techniques are used for the different clusters. The training dataset is used and applying Decision Tree (CART Algorithm).

Read $C_T=\{C_1, C_2, C_3... C_n\}$

DT = DT(C_i)

// C_i stands for various clusters

// DT is Decision Tree

Applying different decision tree algorithm on various clusters for the predictive purpose

- CART algorithm
- C5.0 algorithm
- CHAID Algorithm
- Cost-Sensitive learning method
- Neural Network Technique (Performance)

DISCUSSIONS

Stage one of the algorithms is based on findings of symptoms of the churner customer or the possibility of such a case. The prime task which is accomplished here is to clustering of the customer according to various locations. The robustness of the algorithm is to filtering customer through the data of various variables which are almost 10-12 for each one. At last of the stage customer will be categorized in the /concerned category from there we can predict the future possibilities with him/her.

The second stage of the model gives a vast scope to apply traditional data mining on the filtered data simultaneously. This simultaneous operation gives us the facility to compare the result of various algorithms. This situation provides us to analyze the result with different-different angles. Also, we can choose the best result for taking in the action.

CONCLUSIONS

This paper gives us a hypothetical view and algorithm which provide a roadmap to solve the churn prediction model. The model gives a combination of various types of methods which are earlier used for solving such problems separately. This model is contributed combine approach to find out the optimum solution. The two- step strategy of the model gives an ample amount of opportunity to solve the problem from all the direction technically as well as statistically. So we can say that the proposed two step hypothetical model may provide a solution as per the nature of requirements of the researchers. As the base various data mining rules and algorithms are used as the part of the model, therefore, there is less chance to generate the bogus result.

REFERENCES

1. *Bingquan Huang, B.Buckley, T.-M.Kechadi, ―Multiobjective feature selection by using NSGA-II for customer churn prediction in telecommunications‖, Expert Systems with Applications 37 (2010) 3638–3646.*
2. *W. Verbeke, D. Martens, C. Mues, and B. Baesens, ―Building comprehensible customer churn prediction models with advanced rule induction techniques,‖ Expert Syst. Appl., vol. 38, no. 3, pp. 2354–2364, 2011.*
3. *A. M. AlMana and M. S. Aksoy, ―An Overview of Inductive Learning Algorithms,‖ Int. J. Comput. Appl., vol. 88, no. 4, pp. 20–28, 2014.*

4. M. S. Aksoy, H. Mathkour, and B. A. Alasoos, ―Performance evaluation of rules-3 induction system on data mining,‖ Int. J. Innov. Comput. Inf. Control, vol. 6, no. 8, pp. 1–8, 2010.
5. Ng, A. Y., & Jordan, M. (2002). On Discriminative vs. Generative Classifiers: A comparison of logistic regression and Naive Bayes. Neural Information Processing Systems: NIPS 14.
6. Mitchell, T. M. (2005). Generative and Discriminative Classifiers: Naive Bayes and Logistic Regression. In Machine Learning. Unpublished manuscript.
7. RUST, R. T. & ZAHORIK, A. J. (1993) Customer Satisfaction, Customer Retention, and Market Share. Journal of retailing, 69, 193-215.
8. KIM, H. & YOON, C. (2004) Determinants of Subscriber Churn and Customer Loyalty in the Korean Mobile Telephony Market. Telecommunications Policy, 28, 751-765.
9. Souvik Singha & G. K. Mahanti, Encoding Algorithm for Crosstalk Effect Minimization Using Error Correcting Codes, International Journal of Electronics and Communication Engineering(IJECE),Volume 2, Issue 4, September - October 2013, Pp 119-130.
10. AU, W., CHAN, C. C. & YAO, X. (2003) A novel evolutionary data mining algorithm with applications to churn prediction. IEEE transactions on evolutionary computation, 7, 532-545.
11. HWANG, H., JUNG, T. & SUH, E. (2004) An LTV Model and Customer Segmentation Based on Customer Value: A Case Study on the Wireless Telecommunications Industry. Expert systems with applications, 26, 181-188.
12. DATTA, P., MASAND, B., MANI, D. R. & LI, B. (2001) Automated cellular modelling and prediction on a large scale. Issues on the application of data mining, 485-502.
13. S. A. Qureshi, A. S. Rehman, A. M. Qamar, A. Kamal, and A. Rehman, ―Telecommunication subscribers' churn prediction model using machine learning,‖ in Digital Information Management (ICDIM), 2013 Eighth International Conference on, 2013, pp. 131–136.
14. V. Lazarov and M. Capota, ―Churn Prediction,‖ TUM Comput. Sci., 2007.
15. W. H. Au, K. C. C. Chan, and Y. Xin, ―A novel evolutionary data mining algorithm with applications to churn prediction,‖ Evol. Comput. IEEE Trans., vol. 7, no. 6, pp. 532–545, 2003.
16. M. C. Mozer, R. Wolniewicz, D. B. Grimes, E. Johnson, and H. Kaushansky, ―Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry,‖ Neural Networks, IEEE Trans., vol. 11, no. 3, pp. 690–696, 2000.
17. A. Sharma and P. K. Panigrahi, ―A Neural Network based Approach for Predicting Customer Churn in Cellular Network Services,‖ Int. J. Comput. Appl., vol. 27, no. 11, 2011.
18. J. Hadden, A. Tiwari, R. Roy, and D. Ruta. Churn Prediction: Does Technology Matter. International Journal of Intelligent Technology, 1(1):104-110, 2006.