



International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Vol. 4, Issue 1, January 2016

Churn Prediction – A Review

Navneet Kaur, Naseeb Singh

M. Tech, Dept. of I.T., AIET, Faridkot, Punjab, India

Assistant Professor, Dept. of CSE, AIET, Faridkot, Punjab, India

ABSTRACT: Churn prediction has become one of the greatest fears in the growth and development of an organization. Company has to pay more attention towards the customers rather than products as customers are more valuable persons to the company. Various Data Mining Techniques have been used in order to separate the customers into loyal and churn customers so that proactive measures would be taken in order to retain them in the company. This paper presents the new Hybrid approach of Boosted tree for churn prediction. The hybrid approach of boosted tree is an enhancement of Boosted tree. This paper mainly focuses on the prediction of churn customers to the company with more accurate and well-defined model.

KEYWORDS: Churn customers, Data Mining, churn prediction, CRM, Boosted trees, Loyal customers.

I. INTRODUCTION

Data Mining is a useful learning technique in order to extract the useful information from large database. There is need of DM technique in order to separate the customers into loyal and churn customers from the given dataset. With the help of various DM techniques, various models have been implemented. Neural networks, Support Vector Machine and Logistic Regression are most popular techniques among them. NN performs the traditional statistical approach and linear regression follows the Restrictive approach [1].

In the competitive environment of technology, market strategy focus on the customers rather than products in order to maximize their profitability. In recent years, market strategies have been changed from “Product-Oriented” to “Customer-Oriented” and hence, focus on customer relationship management. Attract customers to the companies, they provide offers like reducing cost on product, income growth, reducing customer sensitivity to prices and changes. Rate of churn customers increases as the market grows. A small reduction in churn customer can increase the profit of company by 5 percent. Various variables are to be used in order to check the churn prediction. These variables are also known as “Powerful and efficient predictive variables” [2].

CRM offers the company all facilities and services in order to retain the churn customers to the company. Churn prediction model have been implemented with the use of CRISP-DM based on RFM and Random Forest. To separate the customers into loyal and Churn customers, Hybrid approach of Boosted tree and Random Forest has been used. DM technique has been made up of six different phases- Business understanding, Data understanding, Data Preparation, Modelling, Evaluation and Deployment [3]. CRM helps to manage the churn customers by growing them according to the behaviour. It helps to identify those customers by making perceptions and customer relationships.

The rest of paper is represented into different sections- Section 2 describe the concept of churn customers, Section 3 is related with the research problem, Section 4 represents the new Hybrid approach of churn prediction. Section 5 discusses the model representation. At last, conclusion is represented

II. CHURN CUSTOMERS

Every organization depends on the customers for the growth and development of their company. Customers’ can be categorized into Loyal and Churn customers. Customers are god in the rise and fall of the growth of company.

Churn is basically related to the switching off from the contract. Churn customers can be handled by managing them in the company is called churn management.

Loyal customers- Loyal customers are those customers who are satisfied with the company service and product. They are the regular visitors to the company.



International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Vol. 4, Issue 1, January 2016

Churn Customers- These are customers who are dissatisfied with the company due to which they would like to move their interest to another company and hence become churn customers.

There may exist several reasons behind the customers to be churn such as Dissatisfaction, Higher cost, Quality, Lack of Features, New Technology etc [4]. Churn customers can be categorized in two different approaches- Proactive and Reactive approach[5]. Churn customers again subdivided into Voluntary and Involuntary churn customers.

Voluntary churn – This type of churn occurs when customer automatically cancel their relationship with the company, therefore it becomes difficult for the organization to identify such type of customers.

Involuntary churn – It happens when company itself declares the cancellation of relationship between customers and themselves, due to some reasons like Fraud or non-payment etc [6].

III. RESEARCH PROBLEM

Boosted tree Technique- Boosted Tree is one of the most powerful DM techniques in the prediction of churn customers. Boosted tree is used to build strong classifier with the use of weak classifiers by using training sets or iterations. Classifiers are defined as set of classes on where new predictions are to be made on the basis of training sets. Training sets are basically iterations or initial discovery of relationship on the given dataset. Strong classifiers refer to build those classifiers, whose classification is much better than existing.

In boosted tree, weights and training sets are assigned to each object. The process is repeated until the strong classifiers are to be developed. To obtain the strong classifier, more weights are to be assigned so that next classifier pays more attention towards the object as it is not accurately classified. Numbers of classifiers are used in updating the strong classifier. The class with higher weight will be predicted as winner class [7].

Obtaining strong classifiers with iterations and number of classifiers always takes long time in decision making. Lesser the classifiers used, Higher the performance in making decisions in less time period.

IV. PURPOSED APPROACH

This paper is an enhancement of Boosted Tree for churn prediction with more accuracy. We are purposing an algorithm based on boosted tree, where less number of classifiers is to be used, so that appropriate decisions can be taken in less time.

In existing boosted tree, Number of classifiers is to be used in making strong classifiers, which leads to make decisions in long time. The new hybrid approach is developed on the base of boosted tree to develop more effective model for the prediction of churn customers. This new hybrid approach will be based on performance metrics.

TP Rate

TPR is defined as measurement of positive cases that are correctly identified. TPR is also called Sensitivity in some fields.

$$\text{TP Rate} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

FP Rate

FPR is measurement of those instances, which are incorrectly identified as positive. It is also known as fall-out.

$$\text{FP Rate} = \frac{\text{FP}}{\text{FP} + \text{TN}}$$

Precision

It is measurement of all positive cases in making calculations. Higher the precision value, more accurate results are prepared.

$$\text{Precision} = \frac{\text{TP}}{\text{FP} + \text{TP}}$$

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Vol. 4, Issue 1, January 2016

Kappa Statistics

It is measurement between classification and true classes. It is defined as rate of observed agreement to maximum possible agreement.

Accuracy

It is defined as correctly classified instances. It relates to correct or positive instances, while making predictions. Higher the accuracy rate, more effective and well defined algorithm is developed.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}$$

Confusion matrix

It displays the values in the tabular form. It is an overview of performance of an algorithm.

CLASS	Actual Churners	Actual Non-Churners
Predicted Churners	TP	FP
Predicted Non-Churners	FN	TN

V. METHODOLOGY

The new hybrid approach is implemented on the base of Boosted tree. This Hybrid approach helps in predicting the churn customers from the given dataset with more accuracy and less error rate. The Hybrid approach is based on boosted tree that helps to separate the customers into loyal and churn customers from the given dataset.

Figure 1 explains the purposed approach as below:

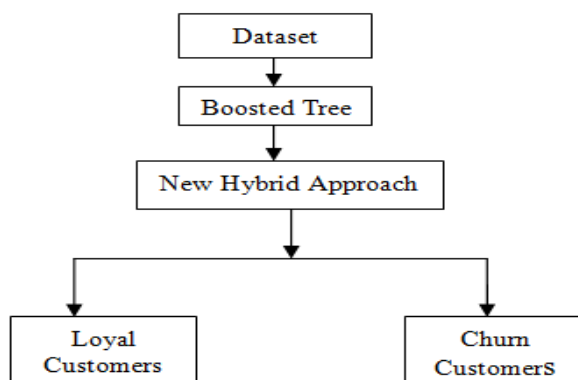


Fig.1 Purposed approach

Boosted tree is implemented and its performance to be checked. New Hybrid approach for the churn prediction is to be implemented and comparisons are to be made in order to predict the accurate model from existing and purposed Hybrid

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Vol. 4, Issue 1, January 2016

Approach. The main idea behind the Purposed Hybrid approach is to develop a more and well defined model on the base of boosted tree in prediction of churn customers.

Figure 2 depicts the research design for the implementation of hybrid approach as shown below:

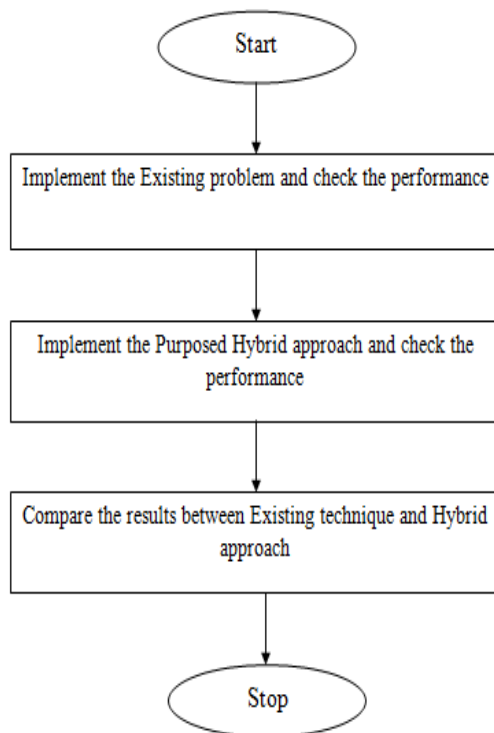


Fig.2 Research Design

VI. CONCLUSION AND FUTURE WORK

Organizations need to more aware about the customers rather than product as customers are more valuable to them. Losing customers becomes the major problem for every organization has to attract the new customers, company have to suffer from problems. Data Mining Techniques have been used so that proactive measures can be taken in securing the churn customers for the profit and survival of a company In future this approach is implemented to conclude the results in making prediction on churn and loyal clients.

REFERENCES

1. Anuj Sharma and Dr. Prabin Kumar Panigraha(2011), "A Neural Network based Approach for Predicting Customer Churn in Cellular Network Services", International Journal of Computer Applications, Vol. 27, No.11,pp. 26-31
2. Sadaf Nabavi, Shahram Jafari(2013) ," Providing a Customer Churn Prediction Model Using Random Forest and Boosted TreesTechniques (Case Study: Solico Food Industries Group)", Journal of Basic and Applied Scientific Research, Vol.3, No.P 6, Pp. 1018-1026
3. Sadaf Nabavi, Shahram Jafari(2013) ," Providing a Customer Churn Prediction Model Using Random Forest Technique (Case Study: Solico Food Industries Group)", IEEE conference on the information and knowledge Technology (IKT).
4. Ahn, P. Hana, and S. Lee (2006), "Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry, Telecommunications Policy", International Journal of Computer Applications, Vol. 30(10-11), pp. 552-568
5. Burez J. and Van D. (2008), "Separating Financial from Commercial Customer Churn: A Modeling Step towards Resolving the Conflict between the Sales and Credit Department," Expert Systems with Applications, Vol. 35, Issue 1, pp. 497-514



ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Vol. 4, Issue 1, January 2016

6. E. Shaaban, Y. Helmy, A. Khedr and M. Nasr(2012) , " *A Proposed Churn Prediction Model*, " International Journal of Engineering Research and Applications, Vol. 2, No. 4, Pp.693-697.
7. Sadaf Nabavi, Shahram Jafari(2013) ," *Providing a Customer Churn Prediction Model Using Random Forest and Boosted TreesTechniques (Case Study: Solico Food Industries Group)*", Journal of Basic and Applied Scientific Research, Vol.3, No. 6, Pp. 1018-1026

BIOGRAPHY

Navneet Kaur is a M. Tech Student in the Information Technology Department, Adesh Insitute of Engineering and Technology, Faridkot, Punjab, India. She received Bachelor degree in IT in 2012 from BMSCE, Muktsar, Punjab, India. Her research interests are Data Mining, algorithms, Networking etc.

Naseeb Singh is an Assistant Professor in CSE in Adesh Institute of Engineering and Technology, Faridkot, Punjab, India. His research area is Data mining, networking, simulation, algorithms etc.