International Journal of Emerging Engineering Research and Technology
Volume 2, Issue 3, June 2014, PP 101-105
ISSN 2349-4395 (Print) & ISSN 2349-4409 (Online)

Customer Churn Prediction in Cloud Computing by using Fuzzy Boosted Trees

Navneet kaur [1], Rajneesh Narula[2], Naseeb Singh[3]

[1](M.Tech Department of Information Technology, Adesh Institute Of Engineering and Technology, Faridkot, India)

[2](Asst. Professor Department of computer science and engineering, Adesh Institute of Engineering and Technology, Faridkot, India)

[3](Asst. professor Department of computer science and engineering, Adesh Institute of Engineering and Technology, Faridkot, India)

Abstract: *Organizations always take part in competition and as the competition grows; they are more concern about their customers rather than products. Organization always focuses on customer's behaviour to retain in market competition. Churn prediction models are developed to manage and control customer churn in order to retain existing customers. Churn prediction aims to predict profitable customers. To predict the customer's behaviour, data mining techniques are used. Various algorithms are used for the prediction of customer's behaviours. This paper proposes the enhancement in the performance of existing boosted tree algorithm by removing repeated scanning for scoring of the classifiers.This paper presents a new prediction model based on Data Mining (DM) technique that is fuzzy boosted trees. The prediction of churn customers is implemented in cloud computing environment. Net Beans is used with WEKA for this. This approach will improve the performance of existing boosted trees technique.*

Keywords: *Data Mining, Churn Customers, Churn Customer Management, Boosting, Boosted Trees, Customer Churn Prediction*

1. INTRODUCTION

Data mining is 'the extraction of hidden predictive information from large databases'. Data mining techniques allow the transformation of raw data into business knowledge. Data mining as 'the process of selecting, exploring and modeling large amount of data to uncover previously unknown data patterns for business advantage'. The rapid growth of the market in every sector is leading to a bigger subscriber base for service providers. More competitors, new and innovative business models and better services are increasing the cost of customer acquisition In a world of ever growing competition on the market, companies have become aware that they should put much effort not only trying to convince customers to sign contracts, but also to retain existing clients.In the cloud computing environment service providers have realized the importance of the retention of existing customers. Therefore, providers are forced to put more efforts for prediction and prevention of churn. Today, customer relationship management (CRM) systems are replacing traditional mass marketing strategies by selective or personalized marketing practices .These selective marketing practices involve identifying a sub-set of existing customers that are likely to stop using products or services of the company (churn). As existing customer's churning will likely to result in the loss of businesses and thus decline in profit. A small change in the retention rate can result in significant impact on business. Customer relationship management (CRM) always concentrates on loyal customers for managerial decision making. In order to effectively control customer churn, it is important to build a more effective and accurate customer churn prediction model. Statistical and data mining techniques have

been utilized to construct the churn prediction models.

There are several data mining techniques that are proposed to predict potential customers that are most likely to churn. Data mining techniques most commonly used include clustering, associations, rule induction, genetic algorithm, decision tree, and neural network.This research supports the former task, means it intends to illustrate how to apply data mining techniques to predict churn customers. This paper suggests Fuzzy boosted trees technique to predict churn customer's in cloud computing environment.

The rest of the paper is organized as follows. Section 2 reviews the current literature, related to customer churn and boosted trees, and Section 3 presents an approach to enhance the performance of existing boosted trees technique. Section 4 describes the tools and methodology for predicting churn customers. In the last section, conclusion is presented.

2. LITERATURE REVIEW

2.1 Customer Churn

'Chur'' is a word derived from change and turn. It means the discontinuation of a contract.

There are three types of churn:

- Active / deliberate - the customer decides to quit his contract and to switch to another provider. Reasons for this may include: unsatisfaction with the quality of service
- Rotational / incidental - the customer quits contract without the aim of switching to a competitor. Reasons for this are changes in the circumstances that prevent the customer from further requiring the service
- Passive / non-voluntary - the company discontinues the contract itself.

2.2 Customer Churn Management

Customer churn is a term used to indicate the customer movement from one provider to another, and 'churn management' is a term that describes an operator's process to retain profitable customers. The term churn management is used to describe the practices of securing the most important customers for a company. In essence, effective customer management presumes an ability to forecast the customer decision to shift from one service provider to another.

2.3 Boosting

Boosting is a machine learning algorithm that attempts to create a strong classifier starting from a set of weak classifiers. Weak classifiers are those whose error rate is only slightly better than random assignation. Strong classifiers are highly accurate and close to true classification. Boosting algorithms are iterative learning process that combines weak classifiers in order to create a strong classifiers .At each step of iteration new weak classifiers is added and weighted according to learner's accuracy and /or the step in iterative process.

Meanwhile, the data is reweighted by assigning more importance to the still misclassified observations; as a consequence newest classifiers focus more on the population that was not correctly classified in previous steps.

2.4 Boosted Trees

Boosted tree is used for segmentation of churn customers. It assigned a weight for each leaning object. After training the previous classifier, weight of the learning objects is updated so that next

classifier pay more attention to the object if it is not accurately classified by previous classifier. The assigned weight is used to vote for each classifier.

If there is less error rate of classifier then more weight assigned to its vote, this training process is repeated. The weight of classifiers which voted for an object of a class is added. The class which gains higher total weight is the final class and it will introduced as the predictive class for that object. Building classifiers and then weighted vote to the decision of the classifier takes a long time.

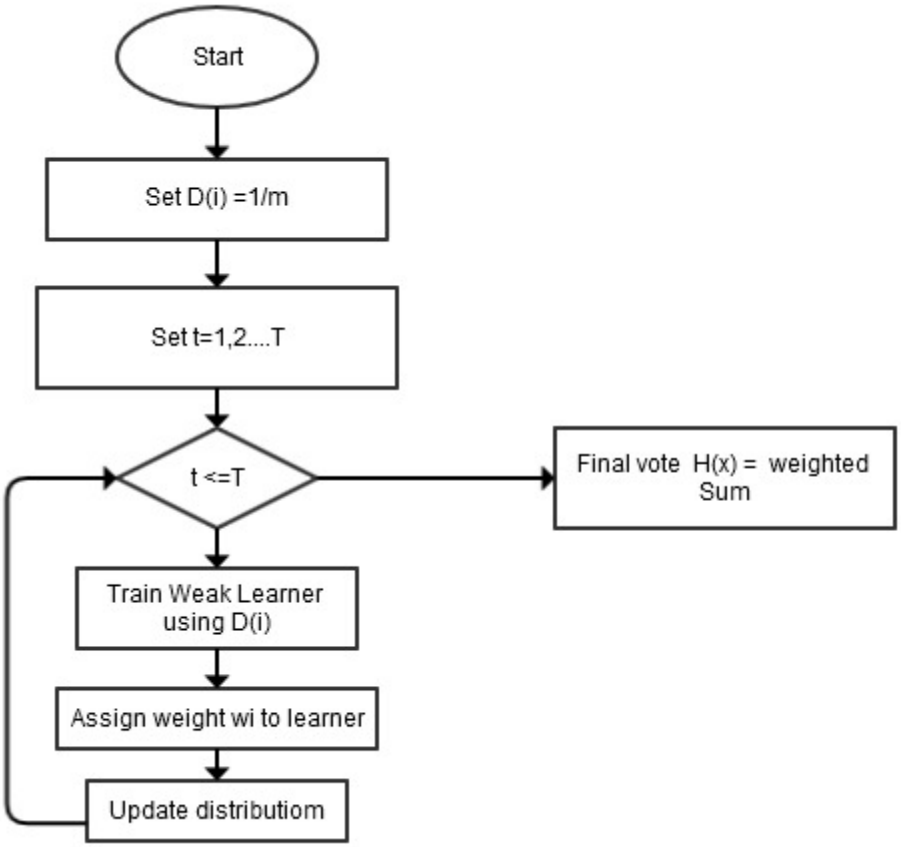

Fig1. *Boosted Trees*

Here "D" is number of data sets used in training. "t" is a variable used for iterating the dataset. It is used to keep track of items. "m" is the a number used to divide data sets so that classifiers can be defined. "T" is transaction set which contains all the transactions. "H" is final decision of the tree.

3. PROPOSED WORK

In existing Boosted tree algorithm, building classifiers and then weighted vote to the decision of the classifier takes a long time. We are reducing this time by increasing the accuracy of the classification using fuzzy logic and removing noise from the test data before starting the processes. As smaller differences in data items can lead to drastic decision changes. The Fuzzy logic will help in decreasing decision change if data contains small changes. The iterations are used to find the strong classifier. At each step of the iteration a new weak classifier is added and weighted according the learner's accuracy and/or the step in the iterative process. A characteristic inherent in fuzzy trees is that the classification of an example can derive in two or more leaves due to the overlapping of the fuzzy sets. As a result of

increased accuracy and data partitions more data will be consumed in lesser number of classifiers. The lesser number of classifiers will result into lesser decision time and increased performance of the algorithm.

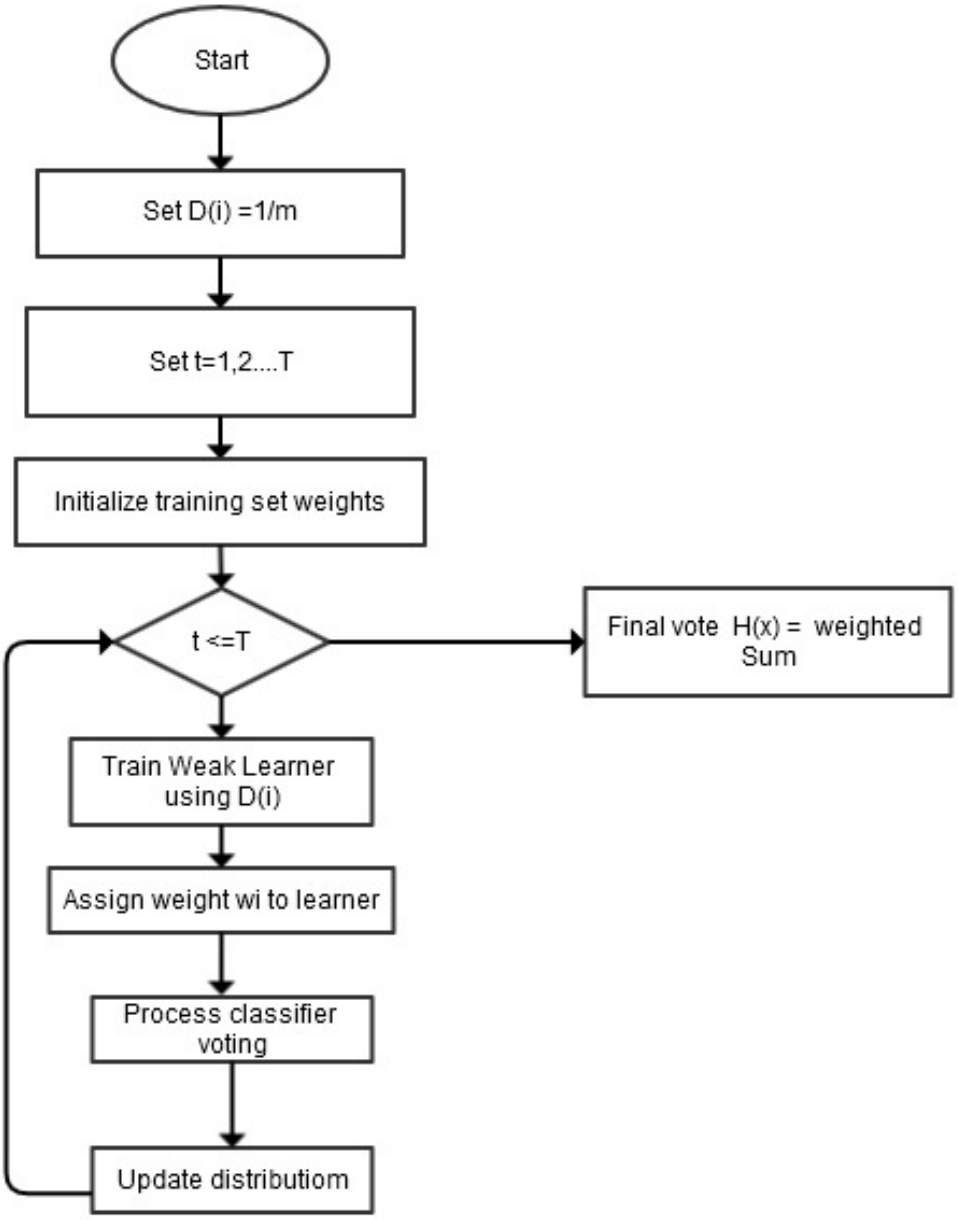

Fig 2. *Fuzzy Boosted Trees*

4. TOOLS AND METHODOLOGY

Tool used

Net beans IDM with WEKA

Methodology

Implement fuzzy boosted tree algorithm after checking the performance of existed boosted tree algorithm. Use the enhanced algorithm for analyzing the customer behavior after comparing the performance of enhance algorithm with existing boosted algorithm.

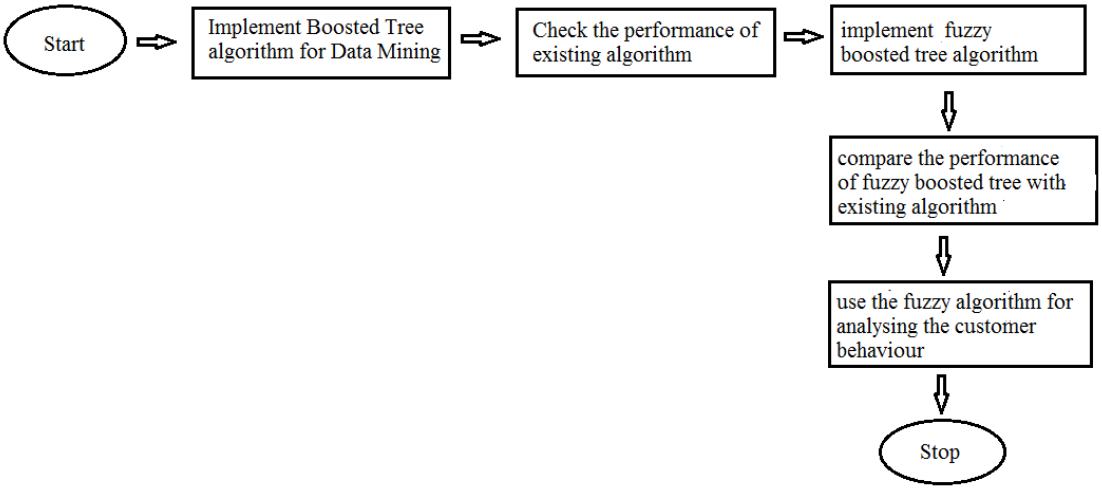

Fig 3 *Research Methodology*

5. CONCLUSION

Organizations need to be more concern about customers and always work for Customer Relationship Management. Losing present customers causes additional costs due to the loss in sale and it generated the need of attracting more new customers. Retention is the most valuable asset for organization. For this purpose data mining techniques are used. So this paper proposed a new approach to enhance the performance of existing boosted trees technique. Then this enhanced approach is used for predicting customer churn. So that proactive measures could be taken by company for churn prevention.

REFERENCES

[1] J. Han and M. Kamber, Data Mining: Concepts and Techniques (Morgan Kaufmann, India, 2006)

[2] S.Y. Hung, D. Yen and H.Y. Wang, Applying Data Mining to Telecom Churn Management, Expert System with Applications, Elsevier, 31(3), 2006, 515-524.

[3] M. M. Rosario and V. Pedro, Tracking of consumer behavior in e-commerce, Proc. 16th International Conference on Information Fusion, Istanbul, Turkey, 2013, 1214-1221

[4] J. Hadden, A. Tiwari, R. Roy, and D. Ruta, Computer assisted customer churn management: State-of-the-art and future trends, Computers and Operations Research, 34(10), 2007, 2902-2917.

[5] E.W.T. Ngai, L. Xiu and D.C.K. Chau, Application of data mining techniques in customer relationship management: A literature review and classification, Expert System with Applications, Elsevier, 36(4), 2009, 2592-2602.

[6] M. J. Berry and G. Linoff, Data Mining Techniques: For Marketing, Sales, and Customer Support (John Wiley & Sons, New York, NY, USA, 1997)

[7] S. Nabavi and S. Jafari, Providing a Customer Churn Prediction Model using Random Forest Technique, Proc. 5th IEEE-Conference on Information and Knowledge Technology (IKT), shiraz, Iran, 2013, 202-207

[8] Kentrias, S. (2001). Customer relationship management: The SAS perspective, www.cm2day.com