International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

Churn Prediction in Cloud with Fuzzy Boosted Trees

Navneet kaur[1], Naseeb Singh[2], Pawansupreet Kaur[3]

[1]M.Tech, Department of Information Technology, Adesh Institute of Engineering and Technology, Faridkot, PTU
[2]Asst. Professor, Department of Computer Science & Engineering, Adesh Institute of Engineering and Technology, Faridkot, PTU
[3]Asst.Professor, Department of Computer Science & Engineering, Shaheed Bhagat Singh College of Engineering & Technology, Ferozepur, PTU

Abstract: ***Churn prediction is a biggest concern for an organization due to its associated costs. The aim of customer churn prediction is detecting customers with high tendency to leave a company. Although, many modeling techniques have been used in the field of churn prediction, performance of ensemble methods has not been thoroughly investigated yet. With the rapid growth of digital systems and associated information technologies, there is an emerging trend in the global economy to build digital customer relationship management (CRM) systems due to which consideration in churn prediction has proven promising. As the number of suitable classification methods increases, it has become more difficult to assess which one is the most effective for our application and which parameters to use for its validation. Earlier, most researchers that uses boosting as a method to boost the accuracy of a given basis learner, this dissertation tries to separate customers into two clusters based on the weight assigned by the boosting algorithm. For the purpose of improving the predictive accuracy and interpret ability of churn prediction model, fuzzy Boosted Tree algorithm, which enhances the boosted tree, is proposed in this dissertation to predict customer's churn propensities. The proposed algorithm is implemented in Net beans IDE of version 3. Data sets are used, on which Mining is applied with WEKA tool. Experimental evaluation reveals that Fuzzy boosting algorithm also provides a good separation of churn data; thus, Fuzzy boosting algorithm is suggested for churn prediction analysis. The experimental results give satisfactory results and are more pleasing and comfortable.***

Keywords: Data Mining, Churn Prediction, cloud computing, boosted trees, work flow, fuzzy boosted trees algorithm

1. Introduction

In most industries where switching costs are prevalent, the landscape of activities is painted of Customers attrition or churn[1]. Churn prediction aims to detect customers intended to leave a service provider. Retaining one customer costs an organization from 5 to 10 times than gaining a new one. Predictive models can provide correct identification of possible churners in the near future in order to provide a retention solution [2]**.**Churn prediction is a useful tool to predict customer at churn risk. Conventional churn prediction techniques have the advantage of being simple and robust with respect to defects in the input data, they possess serious limitations to the interpretation of reasons for churn. Therefore, measuring the effectiveness of a prediction model depends also on how well the results can be interpreted for inferring the possible reasons of churn [3].

There are two main categories of churners which are voluntary and involuntary. Involuntary churners are the easiest to identify. These are the customers that Telco decides to remove from subscribers list. Therefore' this category includes people that are churned for fraud, non-payment and customers who don't use the phone. Voluntary churner is more difficult to determine; it occurs when a customer makes a decision to terminate his/her service with the provider. When people think about Telco churn it is usually the voluntary kind that comes to mind [4].

Voluntary churn can be sub-divided into two main categories, incidental churn and deliberate churn. Deliberate churn is the problem that most churn management solutions tries to solve [5].

Churn prediction is implemented in Net beans IDE, which is cloud computing environment and uses WEKA tool for data mining techniques.

The four constituents of cloud that are Cloud user, Broker, Physical Machines and VM"s. The cloud users can make their service requests from any location in the world to cloud to fulfill their service requirements. A cloud data centers include physical machines. VM"s are created on the top of physical machines by using virtualizing of technology. The broker works as a mediator between cloud data centers and cloud users. It assigns cloud resources to client's workflow implementations [6].

There are different existing algorithms for churn prediction. But many of these existing algorithms takes more parameters and number of classifiers

There are a large number of chances to enhance the reliability of algorithms to increase their performance. To refine the performance of previous algorithms, there should need to work latest fuzzy boosted algorithm.

2. Related Work

Data Mining is used in areas such as machine learning, database and data warehouse technology, statistics, information retrieval, neural networks, pattern recognition, knowledge-based systems, high-performance computing, artificial intelligence, image and signal processing, spatial or temporal data analysis, and data visualization.

Licensed Under Creative Commons Attribution CC BY

There are many data mining techniques which are used to detect customer behavior patterns to formulate marketing, sales, and customer support strategies. Organizations need to understand the customer behavior to improve their marketing strategies. There are large numbers of existing data mining techniques to design and execute the churn prediction in cloud environment. But these techniques consider a limited number of factors while taking decisions. Hence, many of these can be enhanced further to increase productivity.

Some of the works done in this area are discussed below

According to Agrawal et al.(1993), Association rules are one of the major data mining techniques. They are used to discover multiple independent elements that co-occur frequently and to discover rules that relate to the co-occurred elements in a given dataset

According to Berry and Linoff (1997) ,data mining techniques can be used to retain the loyal customers, look out the right prospects, identify new markets for products and services, and recognize cross-selling opportunities on and off the web
.
According to Rygielski (2002), neural networks provide a more powerful and accurate means of prediction, neural networks is used to uncover patterns in easily understandable form. Largest limitation of neural networks is that they hardly uncover patterns in an easily understandable manner. Neural network neuron with learning algorithms is able to learn from training data; this makes them be referred to as Artificial Neural Networks (ANN).

According to Ahmed (2004), Association aims to establishing relationships between items which exist together in a given record Market basket analysis and cross selling programs are typical examples for which association modeling is usually adopted. Common tools for association modeling are statistics and apriority algorithms.
Hwang et al. (2004) discovered that logistic regression performed best for predicting customer churn when compared with neural networks and decision tree and suggest that logistic regression was the best model for their purpose.

Sharma and Panigrahi (2011) propose a neural network-based approach in the prediction of customer churn in line with cellular wireless services. Accuracy that is achieved by neural networks fully outweighs the limitation that they need large volumes of data sets and a lot of time to calculate a considerable load for the predictor attributes

3. Research Gap

The primary purpose of research is to identify the objectives or need to be addressed throughout the research.

Exploratory research is most useful when a decision maker wishes to better understand a situation or identify the appropriate solutions Exploration is useful when researchers have to suffer from problems during their research work.

The object of descriptive study is to describe market. Explanatory studies establish causal relationship between variables. The purpose of this research is to develop and design enhanced model for churn prediction from existed one. It will help the company to know what really cause it customers to churn to their competitors.

4. Problem Formulation

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. The existing problem is resolved by latest Fuzzy boosted algorithm.

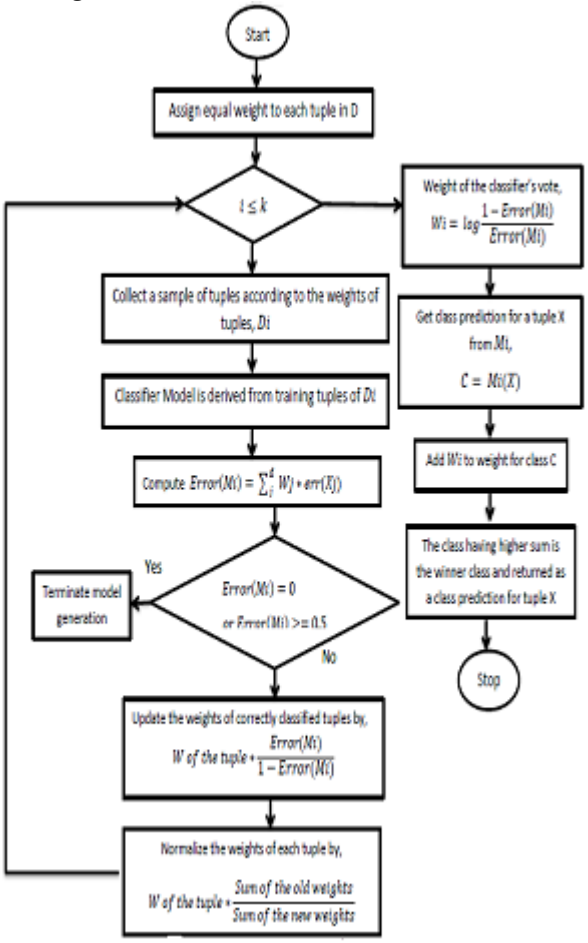

Figure 1: Flow Diagram of Existing Boosted Technique

5. Methodology

This section explains each step conducted out, in the process of achieving research gap mentioned in the scope of work. The steps explained below also lead to the process of evaluation of the implementation.

Licensed Under Creative Commons Attribution CC BY

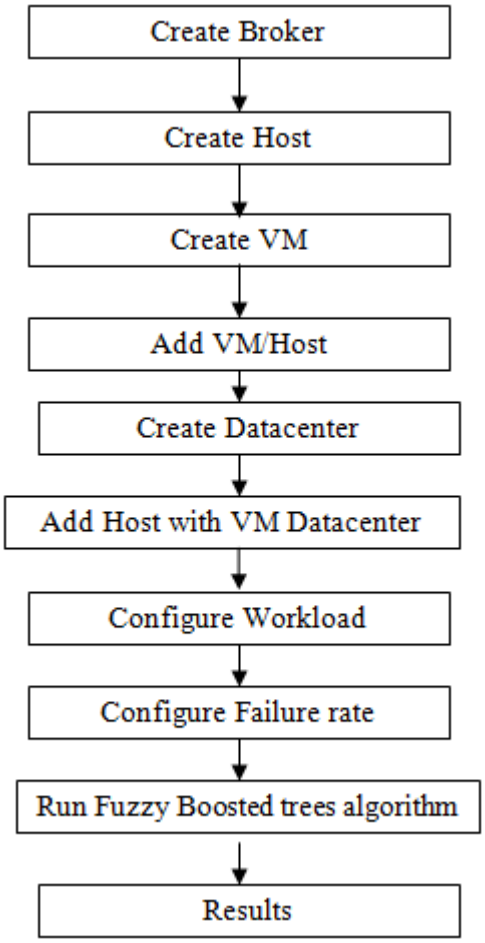

Figure 2: Block Diagram

5.1 Create Broker

We build a datacenter broker in cloud first. The Net Datacenter Broker is a third party individual or business that acts as an intermediary between the purchase of cloud computing services and sellers of that services . In general broker is someone who acts as an intermediary between two parties during negotiation. It hides virtual machine management, as virtual machine creation, submission of cloudlets to this virtual machine and destruction of virtual machines. It helps to the cloud users in the selection of services of cloud. Broker submits the request to the data center, which cloud user wants to use.

5.2 Create Host

Host refers to physical machine. A single host contains number of virtual machine, on which various cloudlets are to be distributed. Cloudlets are number of tasks performed by the user .Cloudlets are distributed to the virtual machines according to the given cloudlet priority. A host has a defined policy for providing memory and bandwidth, as well as an allocation policy for processing elements to virtual machines.

5.3 Create VM

A single host contains number of virtual machine, on which various cloudlets are to be distributed. Cloudlets are number of tasks performed by the user .Cloudlets are distributed to the virtual machines according to the given cloudlet priority. It works as a helper for the small and medium organizations to that requires reliable and a number of services at lower cost rates

5.4 Add VM/Host

Host and virtual machines are added that are already created.

5.5 Create Data Center

Datacenter consists of host and virtual machine, whose host lists are virtualized and network. It is the centralized repository that connects applications, servers and storage services. Enterprises depend upon their data centers to derive the business operations with greater efficiency. Data centers want to be planned and managed carefully to fulfill the user objectives.

5.6 Add Host with VM Data Center

In that step, host added with datacenter of virtual machine

5.7 Configure workload/Submit Workload

In this step path of files is configured and submits the workload. It is a set of processes that can be componentized individually performed upon and evolve a determine result with the abstraction being above the network, hardware and evaluate the performance of the planner to get the optimum results.

5.8 Configure Failure Rate

It collects the failure occurs during the running of algorithms.

5.9 Run Fuzzy Boosted trees algorithm

Algorithm of fuzzy Boosted trees is given as:

Steps:

Step 1: Start with data set having weight of root node equals to 1.

Step 2: At any node p, which is still to be expanded, compute the number of examples of each class. The examples are distributed in part or in whole by branches.

Step3: Now apply iterations to find the strong classifier.

Step 4: At each step of the iteration, a new weak classifier is added and weighted according to learner's accuracy and/or step in the iterative process.

Step 5: A characteristic inherent in Fuzzy trees is that the classification is an example can derive in two or more leaves due to overlapping of fuzzy sets.

Step 6: As a result of increased accuracy and data partitions, more data will be consumed in lesser amount of classifiers.

Licensed Under Creative Commons Attribution CC BY

The lesser number of classifiers will result into lesser decision time and increased performance of the algorithm.

Step 7: 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.

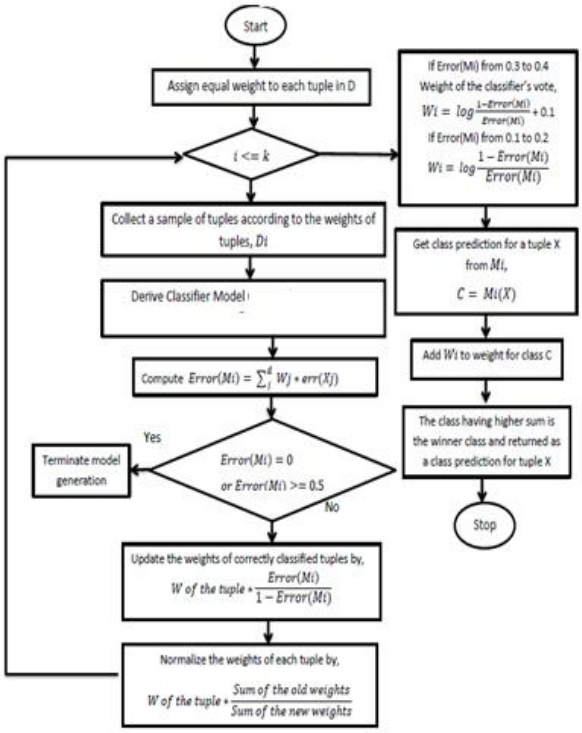

Figure 3: Flow Diagram of Fuzzy Boosted tree Technique

```
========== OUTPUT ==========
Cloudlet ID    STATUS    Data center ID    VM ID    Time    Start Time    Finish Time
    0        SUCCESS        2              0        160        0.1        160.1
Login finished!
```

Welcome admin-When client's registry with cloud server becomes successful, it shows a message "welcome Admin".

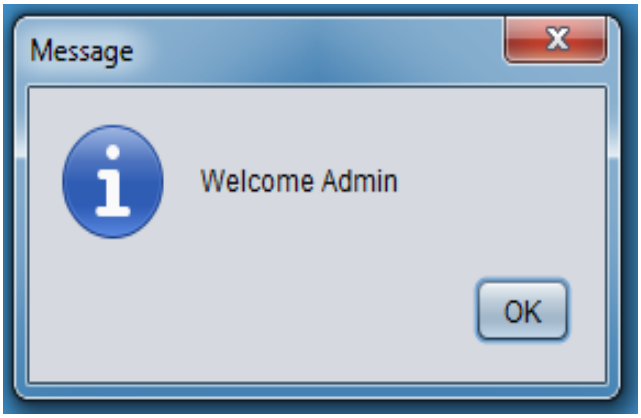

Figure 5: Account Registration

Registry with Cloud Server-Client sends Dataset to cloud server for the prediction technique. Different types of Data Sets are used for prediction.

6. Results

The proposed algorithm is implemented in CloudSim. The observations are carried out by Cloud Sim3.0 simulator. The speed of PE"s (processing elements) indicated in MIPS (Million Instructions per Second) and the length of cloudlets indicated as the number of instructions to be executed. The algorithms are tested by varying the number of cloudless and also randomly varying the length of cloudlets. Also, the number of VMs used to execute the cloudlets, are varied accordingly.

Login –login page is used for invoking in to account, by a client. Client registers connection with cloud server. If client is already registered with cloud server then login directly, otherwise login with new user and register the entry.

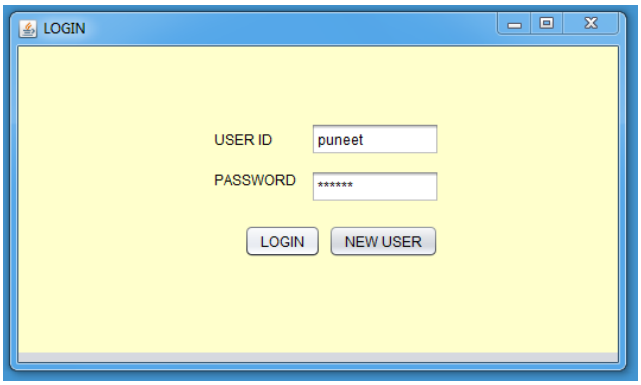

Figure 4: Login Page

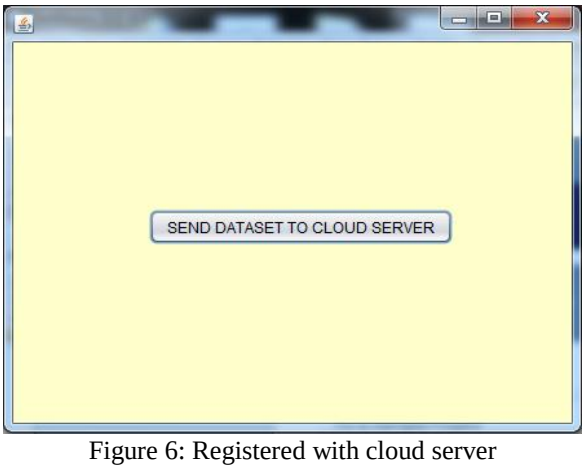

Figure 6: Registered with cloud server

Dataset Collection-Cloud server selects appropriate Dataset from given location and mining is applied on it

Licensed Under Creative Commons Attribution CC BY

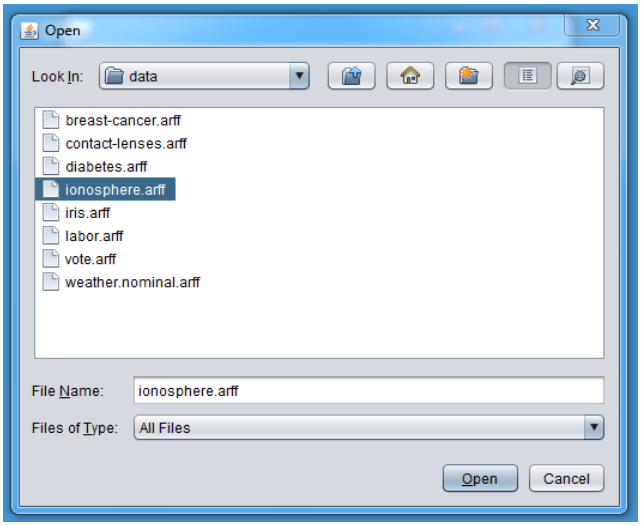

Figure7: Data set for prediction

Submission of Dataset to Cloud Server-Data set is selected by client for prediction and after selection; Data/File is successfully uploaded at the cloud.

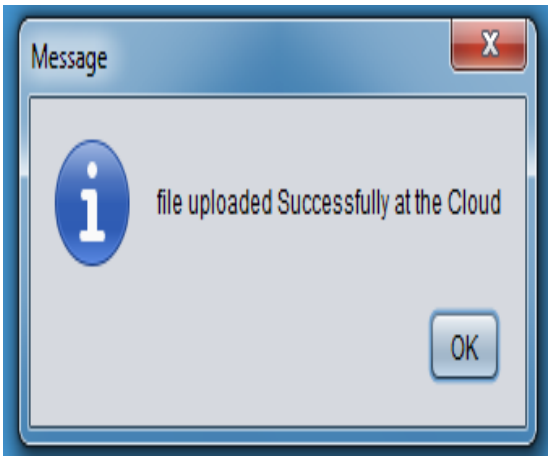

Figure 8: Connection of dataset with cloud server

Representation of Data Set selection by cloud- Various files are shown by cloud server and clients have to select one of them for the churn prediction.

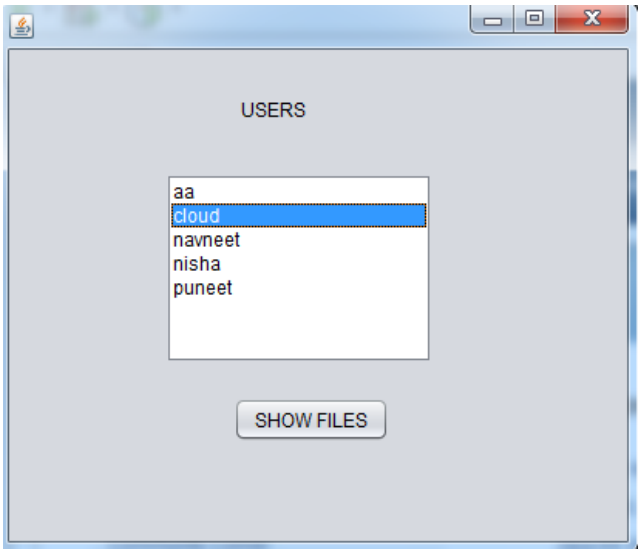

Figure 9: Dataset Selection for mining

Mining Process-Mining process is used to apply mining technique on appropriate data set. Mining is applied with WEKA.

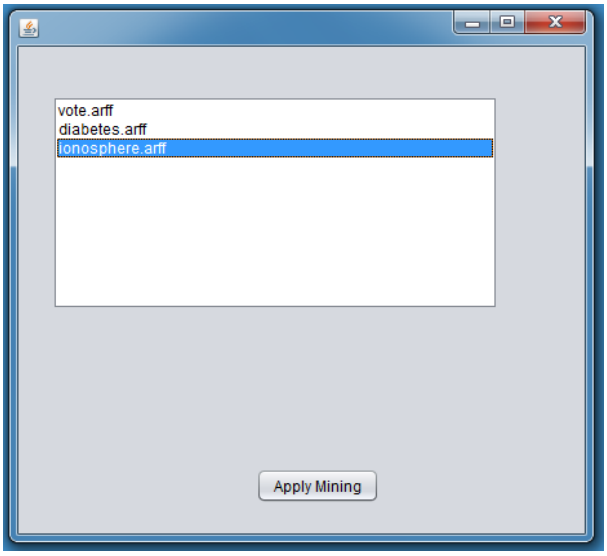

Figure 10: Mining on dataset

Fuzzy Tree Representation-Cloud will develop boosted tree on the given dataset.

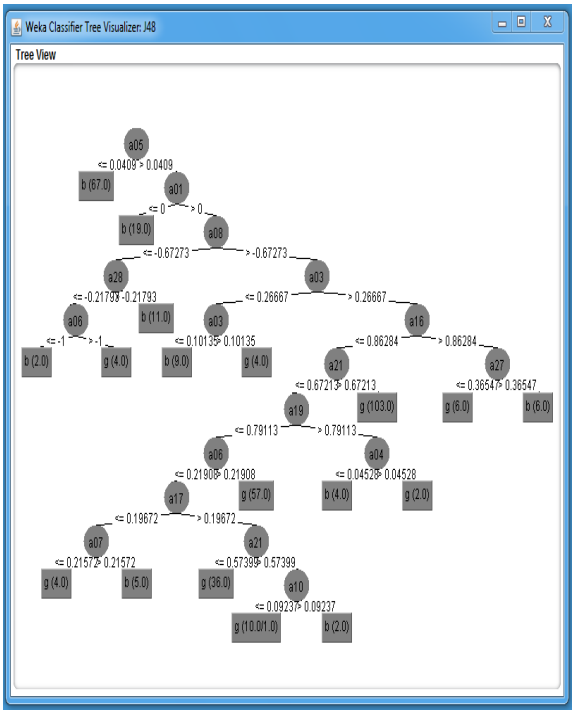

Figure 11: FuzzyBoosted Tree

7. Conclusions and Future Scope

In today's world of competition, organizations need to be more customers-oriented instead of product-oriented and always work for Customer Relationship Management. Customers are the most valuable asset for an organization. So organization must need to retain their profitable customers. For this purpose data mining techniques are used to identify the deliberate churn customers. Fuzzy Boosted tree algorithm is quite successful in its accuracy of churn prediction. Fuzzy boosted tree is implemented in cloud computing and this

Licensed Under Creative Commons Attribution CC BY

information can be used for further analysis. These results are achieved when predictions are made on data sets. Variety of data sets can be introduced for further analysis. Varieties of techniques are available which can be used to incorporate new research in churn prediction. Present method is based on cloud computing but in future it can be incorporated to multiple software platforms to achieve different results. Since the research in churn prediction increasing rapidly. So our requirement to achieve optimal result in less time.

References

[1] Georges D. Olle Olle and Shuqin Cai "A Hybrid Churn Prediction Model in Mobile Telecommunication Industry" International Journal of e-Education, e-Business, e-Management and e-Learning, Vol. 4, No. 1, February 2014
[2] Essam Shaaban, Yehia Helmy, Ayman Khedr, Mona Nasr / International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com Vol. 2, Issue 4, June-July 2012, pp.693-697
[3] V. Lazarov and M. Capota. Churn Prediction. Business Analytics Course. TUM Computer Science, December 2007.http://home.in.tum.de/~lazarov/files/research/papers/churn-prediction.pdf
[4] S. Gotovac. "Modeling Data Mining Applications for Prediction of Prepaid Churn in Telecommunication Services," vol. 51, no. 3, pp. 275-283, 2010
[5] H. Kim, and C. Yoon, "Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market." Telecommunications Policy. Vol. 28 No.: PP. 751-765, 2004
[6] Long, W., Yuqing, L., and Qingxin, X. (2013), "Using CloudSim to Model and Simulate Cloud Computing Environment", 9th International Conference on Computational Intelligence and Security (CIS), IEEE Publication, 14-15 Dec. 2013, Leshan, pp. 323-328
[7] Agrawal, R., Imielinski, T., & Swami, A. (1993), "Mining association rules between sets of items in large databases," in Proceedings of the ACM SIGMOD conference on management of data, pp. 207–216
[8] Anuj Sharma and Dr. Prabin Kumar Panigrah(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
[9] Berry, M. J. A., & Linoff, G. (2004), "Data mining techniques: For marketing, sales, and customer support," John Wiley & Sons publications
[10] Geppert,C. (2002), "Customer Churn Management: Retaining High-Margin Customers with Customer Relationship Management Techniques, KPMG & Associates

Author Profile

Navneet kaur received the B.tech degree in information technology from Bhai Maha Singh College of Engineering & Technology, Muktsar, Punjab in 2012.In 2014, she is currently working on research work. Her area of interest is in cloud environment.

Pawansupreet Kaur is an associate professor in Shaheed Bhagat Singh College of Engineering & Technology, Ferozepur, Punjab. Her area of interest is in cloud environment and in networking.

Licensed Under Creative Commons Attribution CC BY