ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

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

Visit: www.ijirset.com

Vol. 8, Issue 6, June 2019

A Massive Statistical Clustering for Mitigating the Danger of Customer Churn

Raghavendra Rao B.G[1], Vasantha S[2], Aravind P N[3],Arunkumar[4], Bala Kiran[5]

Assistant Professor, Department of Master of Computer Science, Sir MVIT, Bengaluru, Karnataka, India[1]

Assistant Professor, Department of Master of Computer Science, Sir MVIT, Bengaluru, Karnataka, India[2]

Assistant Professor, Department of Mathematics, Sir MVIT, Bengaluru, Karnataka, India[3]

PG Student, Department of Master of Computer Science, Sir MVIT, Bengaluru, Karnataka, India[4]

PG Student, Department of Master of Computer Science, Sir MVIT, Bengaluru, Karnataka, India[5]

ABSTRACT*:* The primary objective is onthe churn in telecom industries to accurately estimate the customer survival and customer hazardfunctions to gain the complete knowledge of churn over the customer tenure. This focuseson analyzing the churn prediction techniques to identify the churn behavior and validate thereasons for customer churn. Summarizes the churn prediction techniques in order tohave a deeper understanding of the customer churn and it shows that most accurate churn prediction is given by Apriori algorithm so that telecom industries become aware of the needs ofhigh risk customers and enhance their services to overturn the churn decision. According to this algorithm attribute selection problem can bereplaced by the pinning question of classifier combination and we structure an indicator system of customer churn. Then a parallel Association rule algorithm is implemented through a HadoopMapReduce framework. The proposed parallel Association rule algorithm enjoysa fast running speed when compared with the other methods. Association rule algorithm is implemented through a Hadoop MapReduce framework for real-time and efficient data analysis. The distributed clusteringmethods are implemented to address the problem of customer churn and we used k-meansalgorithm for checking the accuracy for churn rate. We comparing both algorithm in that oneApriori algorithm is giving the accurate result.

KEYWORDS: Hadoop, MapReduce, Apriori algorithm, Churn management, k-means Algorithm.

I. INTRODUCTION

Telecommunication companies are facing severe loss of revenue due to increasing competition among them which leads to loss of customers. Customer churn means customer choose multiple service providers and they have rights to switch another company from one service to another services. Churn prediction and Analysis are very important in telecomm companies. Predictions can improve the company revenue growth and also it provides the details of the customer who are going to leave the company subscriptions. Managing churns is important where churn rate measures the number of customers moving out in certain period of time. Once reasons are found companies can improve their services to fulfill customer needs.

In existing choice tree, neural framework and K-suggests square measure picked by as essential strategies to make observing models for media transmission purchaser beat estimate. Customer related factors cannot be generated effectively in existing system, entire attributes are selected once at a time it delays the result and it takes time. Their experimental assessment shows

that information mining methodology will sufficiently encourage media transmission pro associations to make extra genuine customer mix gauge. In the developing system unstructured data can be effectively processed, accuracy is more where finding churn customers. The proposed attribute selection method on Apriori algorithm that based on the

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

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

Visit: www.ijirset.com

Vol. 8, Issue 6, June 2019

frame level Apriori. Parallel association rule can be implemented in the Hadoop map-reduce concepts which effectively fast this disseminated pack ways that zone unit authorized to manage the matter of customer beat.

II. RELATED WORK

Problem Statement: The current investigative systems don't work okay in overseeing gigantic information. Right off the bat, as specified prior, to absolutely layout "client stir", we'd get a kick out of the chance to with effectiveness mine the unstructured long range interpersonal communication information. Because the relational database which stores traditional data effectively cannot process the unstructured data, so do the analytical methods as the traditional analytical methods are more likely to encounter performance bottlenecks when conducting customer churn analysis, the new analytical method required to accurate as possible.

Objectives: The main objective of the project is to find the churn discovery of the each and every attribute and also based upon the priorities the by applying the Apriori algorithm the churn rate will be discovered and graph of the very attributes with overall attributes percentage of that number can be displayed effectively. After that we implementing the Hadoop Map-Reduce concepts to find the entire details of the customer by entering the phone number the related information will be shown.

Scope:

- The system effectively analyze the telecom industries dataset.
- By using the Apriori algorithm the churn rate will be discovered.
- By using the Hadoop MapReduce system we can search for the mobile number get the related information.

III. METHODOLOGY

Hadoop: Hadoop is an open source structure from Apache and is utilized to store process and break down information which are extremely enormous in volume. Hadoop is composed in Java and isn't OLAP (online systematic handling). It is utilized for group/disconnected processing. It is being utilized by Face book, Yahoo, Google, Twitter, LinkedIn and some more. In addition it can be scaled up just by including hubs in the group.

Modules of Hadoop

- HDFS: Hadoop Distributed File System. Google distributed its paper GFS and based on that HDFS was created. It expresses that the documents will be broken into pieces and put away in hubs over the conveyed design.
- Yarn: Yet another Resource Negotiator is utilized for work booking and deal with the bunch.
- Map Reduce: This is a system which encourages Java projects to do the parallel calculation on information utilizing key esteem match. The Map undertaking takes input information and believers it into an informational collection which can be processed in Key esteem match. The yield of Map errand is devoured by lessen undertaking and after that the out of reducer gives the coveted outcome.
- Hadoop Common: These Java libraries are utilized to begin Hadoop and are utilized by other Hadoop modules.

MapReduce: MapReduce is an overseeing structure and a program appear for spread preparing in a setting of java. The MapReduce check contains two crucial errands, to be particular Map and Reduce. A guide takes a diversion set up of information and followers it into another system of learning, wherever unequivocal square measures are disengaged into tuples. Future, lessen and, that takes the yield from a guide as knowledge and joins those information tuples into a considerable measure of minor technique of tuples. Since the philosophy of the name MapReduce shuts, the diminishing trek is frequently performed when the guide work.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

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

Visit: www.ijirset.com

Vol. 8, Issue 6, June 2019

Apriori Algorithm: The Apriori formula was projected by Agrawal and Srikant in 1994. Apriori is intended to work on databases containing transactions. Alternative algorithms square measure designed for locating association rules in knowledge having no transactions, or having no timestamps. Every transaction is seen as a collection of things. Given a threshold C, the Apriori formula identifies the item sets that square measure subsets of a minimum of C transactions within the information.

Apriori uses a "bottom up" approach, wherever frequent subsets square measure extended one item at a time (a step called candidate generation), and teams of candidates square measure tested against the information. The formula terminates once no more successful extensions square measure found.

Apriori uses breadth-first search and a Hash tree structure to count candidate item sets with efficiency. It generates candidate item sets of length k from item sets of length k − one. Then it prunes the candidates that have associate degree occasional sub pattern. in keeping with the downward closure lemma, the candidate set contains all frequent k-length item sets. After that, it scans the group action information to see frequent item sets among the candidates.

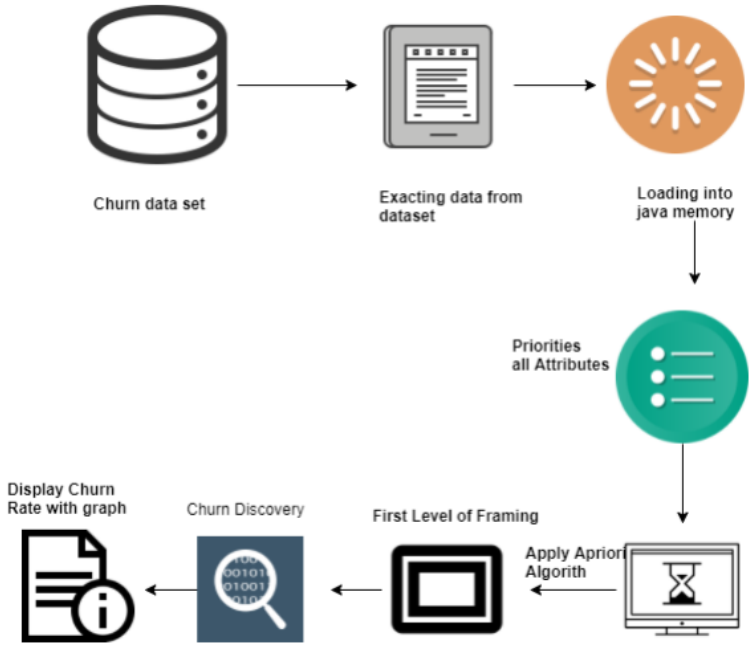

Figure 1. The system architecture for churn rate

V. RESULTS AND DISCUSSION

The essential goal is on the beat in telecom businesses to precisely appraise the client survival and client peril capacities to pick up the entire information of stir over the client tenure. This project centers around examining the agitate expectation procedures to distinguish the beat conduct and approve the explanations behind client agitate. This project compresses the beat expectation

strategies with a specific end goal to have a more profound comprehension of the client stir and it demonstrates that most precise agitate forecast is given by Apriori calculation so telecom enterprises end up mindful of the requirements of high hazard clients and improve their administrations to upset the beat decision. We proposed a property choice strategy in view of Apriori procedure, which is called characteristic determination technique in view of casing level Apriori According to this calculation quality choice issue can be supplanted by the pruning inquiry of classifier mix and we structure a pointer arrangement of client stir.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

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

Visit: www.ijirset.com

Vol. 8, Issue 6, June 2019

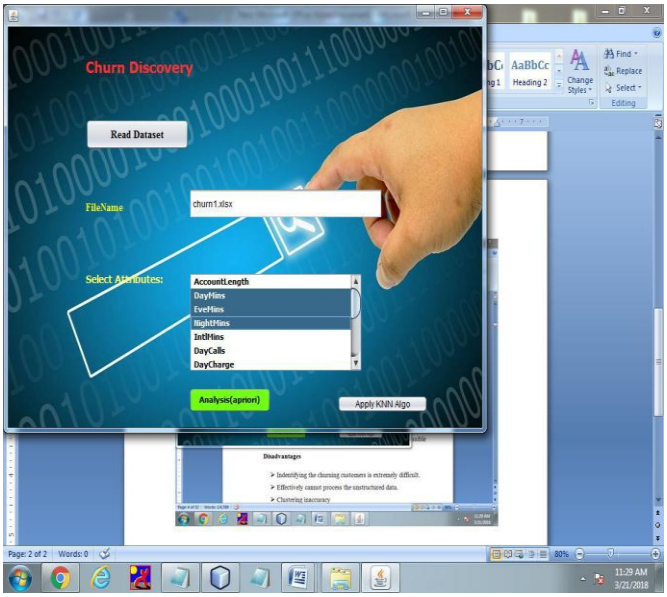

Figure 2: Reading of the dataset and showing the different attributes

Figure 3: Selecting the different attributes

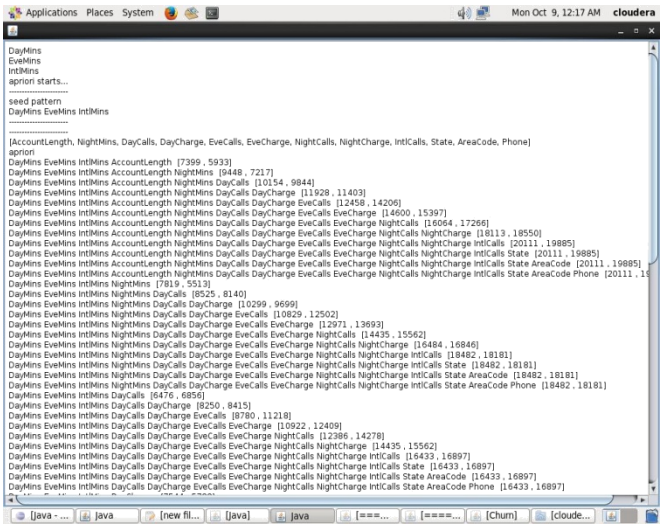

Figure 4: Different attributes present in the dataset

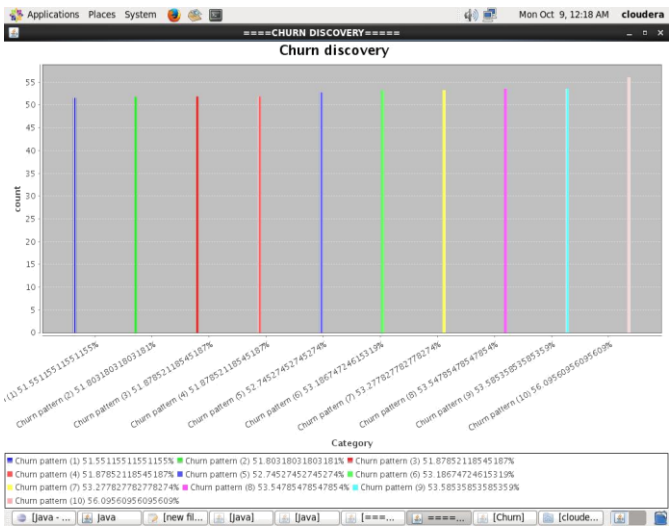

Figure 5: Representing graphical view of the different category of attribute.

VI. CONCLUSION AND FUTURE SCOPE

The main conclusions of the study may be presented in a short Conclusion Section. In this section, the author(s) should also briefly discuss the limitations of the research and Future Scope for improvement.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

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

Visit: www.ijirset.com

Vol. 8, Issue 6, June 2019

In this venture, at first the issue of customer beat and furthermore the endowments of predicting unsettle in media transmission undertakings trying to outfit associations with effective approaches to deflect blending customers in enormous information time, we tend to stomach muscle initio propose Apriori figuring upgrades clustering precision. Moreover, it reduces the peril of uncertain assignments organization. Investigation comes with respect to show that Apriori estimation has extra grounded bundling semantics quality and envisioning the beat precisely.

The second responsibility is that we tend to adjust the past Association oversee to substantial information, and execute it with a Map downsize structure. Thirdly, we tend to focus of the customer mix issue in China medium with great information Apriori and extensive information Association computations. For this case, we tend to use to look at the great execution of the arranged grand data. In addition, we tend to hold a duplicated publicizing push to locate the potential customers will's personality control with most stripped-down esteem.

Results exhibit that the promoting duplicate is critical to broaden boosting points of interest for wanders and furthermore the undertakings should offer cautious idea to the productive groups. From this examination, it are frequently construed that the principal real upset gauge is no heritable while exploitation Apriori estimation rather than single mulling over everything, the way toward dealing with customer mix issue in medium for loss to help the measure buyer beat organization inside the Brobdingnag data setting.

In the future enhancement of the scheme is to enter the multiple phone number of different mobile numbers based on the different mobile number we need to show the details associated with that particular mobile number.

In the scheme we need to find the any association present between the two different number and similarity attribute are there we need to check.

REFERENCES

[1] Wenjie Bi,MeiliCai,Mengqi Liu,Guo Li, "A Big Data Clustering Algorithm for Mitigating the Risk of Customer Churn", IEEE Transactions on Industrial Informatics, Volume.12 , Issue.3 , pp. 1270 – 1281, 2016.

[2] J. Hadden, A. Tiwari, R. Roy, D. Ruta, "Computer assisted customer churn management: State-of-the-art and future trends", *Comput. Oper. Res.*, vol. 34, no. 10, pp. 2902-2917, Oct. 2007.

[3] B. Q. Huang, T. K. Mohand, B. Brian, "Customer churn prediction in telecommunications", *Expert Syst. Appl.*, vol. 39, no. 1, pp. 1414-1425, Jan. 2012.

[4] W. H. Au, K. C. C. Chan, Y. Xin, "A novel evolutionary data mining algorithm with applications to churn prediction", *IEEE Trans. Evol. Comput.*, vol. 7, no. 6, pp. 532-545, Dec. 2003.

[5] E. G. Castro, M.S.G. Tsuzuki, "Churn prediction in online games using players' login records: A frequency analysis approach", *IEEE Trans. Comput. Intell. AI Games*, vol. 7, no. 3, pp. 255-265, Sep. 2015.

[6] B. Edelman, M. Ostrovsky, M. Schwarz, "Internet advertising and the generalized second-price auction: Selling billions of dollars worth of keywords", American Economic Review, vol. 97, no. 1, pp. 242-259, 2007

[7] T. Condie, N. Conway, P. Alvaro, J. M. Hellerstein, K. Elmeleegy, R. Sears, "MapReduce online", Oct 2009.

[8] R. G. Drozdenko and P. D. Drake, Optimal Database Marketing: Strategy, Development, and Data Mining. Newbury Park, CA, USA: Sage,2002.

[9] T. Zhang , R. Ramakrishnan, and M. Livny, " Birch: An efficient data clustering method for very large databases," in Proc. ACM SIGMOD Int. Conf. Manag. Data Rec., 1996, vol. 25, no. 2, pp. 103–114.

[10] A. L. A. Rodriguez , "Clustering by fast search and find of density peaks ," Sci., vol. 344, no. 6191, pp. 1492– 1496, Jun. 2014.

[11] R. Tibshirani , G. Walther, and T. Hastie, "Estimating the number of clusters in a data set via the gap statistic," J. Roy. Statist.Soc.: Series B Statist. Methodol., vol. 63, no. 2, pp. 411– 423, 2001.