iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** VI **Month of publication:** June 2018

DOI: http://doi.org/10.22214/ijraset.2018.6007

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue VI, June 2018- Available at www.ijraset.com

Customer's Mitigation Rate Detection Using MapReduce in Parallel Environment

Pallavi Ghanshani[1], Mr. Vaibhav Chandrakar[2]
[1, 2]*Central College of Engineering and Management, Dept. of Computer Science and Engineering, Raipur, Chhattisgarh, India*

***Abstract:** As market rivalry strengthens, customer mitigation administration is progressively turning into an imperative methods for upper hand for organizations. When managing enormous information in the business, existing mitigation prediction models can't work extremely well. In this paper we proposes a method in combination with Hadoop and MapReduce to predict the customer mitigation rate at an early stage. We have considered parallel fuzzy c-means algorithm to predict the customer mitigation rate.*
***Keywords:** Customer Mitigation. Clustering, Telecom Industry, Substructive Clustering, Axiomatic Fuzzy Sets.*

I. INTRODUCTION

NOWADAYS, expanding market shares has become more and tougher for service industry, such as telecommunications industry, as the competition is fierce and market is increasingly saturated. Thus, these companies pay more attention to the existing customers so as to avoid customer mitigation. Customer mitigation refers to the loss of customers who switch from one company to another competitor within a given period [2]. Industrial practice has shown that customer mitigation can lead to huge economic losses and even hurt the company's public image. Hence, customer mitigation management is extremely important especially for the service industry. In the present business world associations are seeing that customer esteem and extended pay will presumably begin from their present customer base than from new customer securing. The reasoning behind this is associations know their present customers, starting at now have a relationship with them, and adequacy of data on them [9]. In like way affirmation of this issue, industry has seen an ascent of client relationship administration (CRM) things. Programming associations have comprehended that CRM has transformed into an open topic inside industry and thusly a region of potential wealth and open entryway for the change and progression of things that certification to lift customer upkeep, and additionally to expand offering openings [10].

A. Customer Mitigation

Client mitigation insinuates when a customer (player, supporter, customer, et cetera.) stops his or her association with an association. Online associations ordinarily view a customer as disturbed once a particular measure of time has slipped by since the customer's last joint effort with the website page or organization. The full cost of client agitate fuses both lost pay and the displaying costs required with supplanting those customers with new ones. Diminishing client beat is a key business target of each online business [11] [12].

B. Importance of Predicting Mitigation

The ability to anticipate that a particular customer is at a high peril of agitating, while there is still time to make a move, addresses a gigantic additional potential salary hotspot for each online business. Other than the quick loss of pay that results from a customer surrendering the business, the costs of at first understanding that customer won't not have starting at now been secured by the customer's spending to date. (In that capacity, understanding that customer may have truly been a losing wander.) Furthermore, it is continually more troublesome and expensive to secure another customer than it is to hold a current paying customer [13] [14].

II. HADOOP AND ITS COMPONENTS

A. Apache Hadoop

The Apache Hadoop programming library is a structure that takes into consideration the disseminated handling of expansive informational indexes crosswise over bunches of PCs utilizing basic programming models. It is intended to scale up from single servers to a huge number of machines, each offering nearby calculation and capacity.

B. Hadoop MapReduce

A MapReduce work more often than not parts the info informational collection into autonomous pieces which are handled by the guide undertakings in a totally parallel way. The structure sorts the yields of the maps, which are then contribution to the lessen

©IJRASET: All Rights are Reserved

errands. Normally both the info and the yield of the activity are put away in a record framework. The structure deals with planning errands, checking them and re-executes the fizzled undertakings.

C. Mapper

The Mapper maps input key-esteem sets to an arrangement of middle key-esteem sets. It changes input records into middle of the road key-esteem sets. Utilizing map() strategy one can compose their own particular execution of guide errand to create transitional key-esteem combine.

D. Reducer

Hadoop system calls reduce() techniques for every novel key in the sorted request. The reduce work is to emphasize through the key and create at least zero yields. In this manner the delivered yield is put away in HDFS in different parts as gathered by different reducers.

III. LITERATURE SURVEY

Ngai E.W.T. et al. [1] investigation gives a guide to guide future look into and encourage learning amassing and creation concerning the use of information mining methods in CRM. Their investigation recognized arrangement and affiliation models are the two regularly utilized models for information mining in CRM.

Rekha [2] have utilized Naïve Bayesian classifier and Decision Tree-Based classifier to anticipate and break down fraudlent in accident coverage claims. The execution of the model is analyzed with confusion matrix.

Keyvan [3] connected three regulated classifiers specifically choice tree, bayesian arrange and neural systems for forecast of client conduct in an insurrance industry in Iran. He presumed that choice tree assessment strategies appear great execution contrasted and different strategies. Horng et al. [4] have centered clustering and association rule methods to map star items to anticipated potential clients. Clustering examination done to find potential clients utilizing key attributes of steadfast clients individual data. Association manage investigation used to distinguish potential clients' not so distant future enthusiasm for a star item utilizing information of steadfast clients' buying conduct.

Razieh et al. [5] performed client segmentation utilizing RFM method also, grouping calculations in light of client's esteem, to indicate steadfast and beneficial customers. The creator considered the model for a supermarket's information. The creator utilized a mix of behavioural and demographical attributes of people to evaluate steadfastness.

Baesens et al., [6] distinguished the slant of the client lifecycle in view of Bayesian system classifier. The author showed Bayesian system classifiers as a valuable device in the tool kit of CRM examiners in use of recognizing the incline of the client lifecycle of long-life clients.

A. Proposed Methodology

In this section we present our proposed architecture in detail. Fig. 1. Shows proposed system architecture.

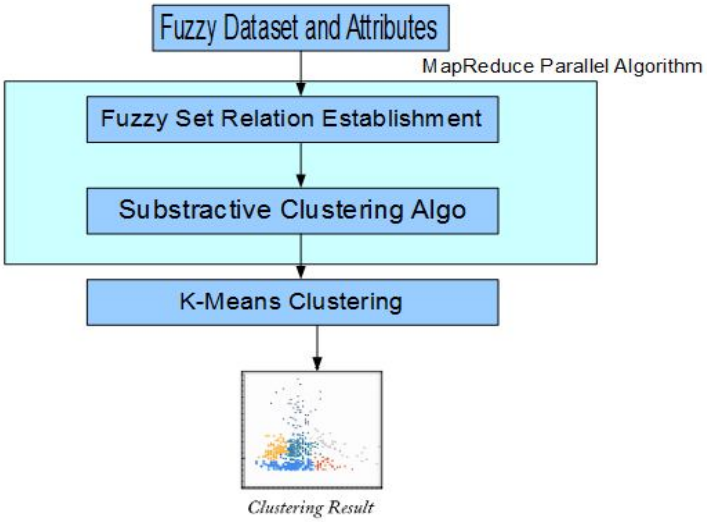


Fig. 1. Shows the proposed system architecture

©IJRASET: All Rights are Reserved

B. Axiomatic Fuzzy Set

AFS is a most proper system to represent the fuzzy idea. The participation work that a class of a specific tuple has a place that some class or not can be effectively determined by the AFS calculation. This determination did with the assistance of some coherent activities. Numerous straightforward connection can be communicated by means of AFS polynomial math.

C. Subtractive Clustering Method

In this area, the SCM calculation and its working is talked about. It is utilized to discover the centroid of the groups. In view of the raw information, the SCM strategy is sufficiently shrewd to decide the cluster and centroid of the groups. In SCM, every one of the information point in the dataset is taken as potential cluster centroid and it is computed as:

$$M_l(x_i) = \sum_{j=1}^{n} \exp\left(-\frac{\|x_i - x_j\|^2}{(\tau_1/2)^2}\right)$$

Where, τ1, is the radius of the neighbors. This radius effect the centroid cluster position. The larger τ1 is the greater will be effect on formula.

D. K-Means Clustering

K-means is the most direct one among all these clustering methods. Therefore, we show it here for figuring the gatherings. Take note of that the clustering comes about are likely going to be free if having dishonorably regarded early on parameters in k-means. In spite of what may be normal, SCM can make more correct data parameters in perspective of unrefined data, including bunch centroids and group number. In this manner, in this paper, the parameters created by SCM go to k-infers keeping in mind the end goal to make walks the precision of k-means.

IV. RESULTS

To assess our framework execution we have utilized Hadoop with Parallel usage subtractive clustering algorithm. Telecom dataset are utilized as a part of our assessment. We have considered different fuzzy parts of the telecom businesses dataset.

This paper, presents work done in field of customer mitigation analysis utilizing clustering algorithm. The Telecom dataset is taken which is accessible on the web.

The MapReduce algorithm are executed for find the customer mitigation rate and results are presented in table I.

TABLE I. Rate of Churn and Number of Customers present in each cluster

Cluster	Number of Customers	Churn Rate
Cluster-0	808	.027
Cluster-1	878	0
Cluster-2	816	.091
Cluster-3	831	.077

The table I. shows the clusters and its customer churn rates. Cluster 0 has 808 customers and its churn rate is 2.7 %. Whereas cluster-1 has 878 customers and its churn rate is 0%. Similar with Cluster-2 and Cluster-3. The highest churn rate is with Cluster-2 and Cluster-3 among all, which is 9.1% and 7.7% respectively. The Churn rate for Cluster-1 is 0%, which is great for any company. But churn rate for cluster-2 and cluster-3 are serious issue for the telecom company.

V. CONCLUSION

We proposed parallel subtractive clustering calculation to execute in parallel in Hadoop with MapReduce. After that K-Means calculation is connected over the dataset. The dataset comprises of 3300 endorsers. The calculation is executed in pseudo-distributed method of Hadoop design.

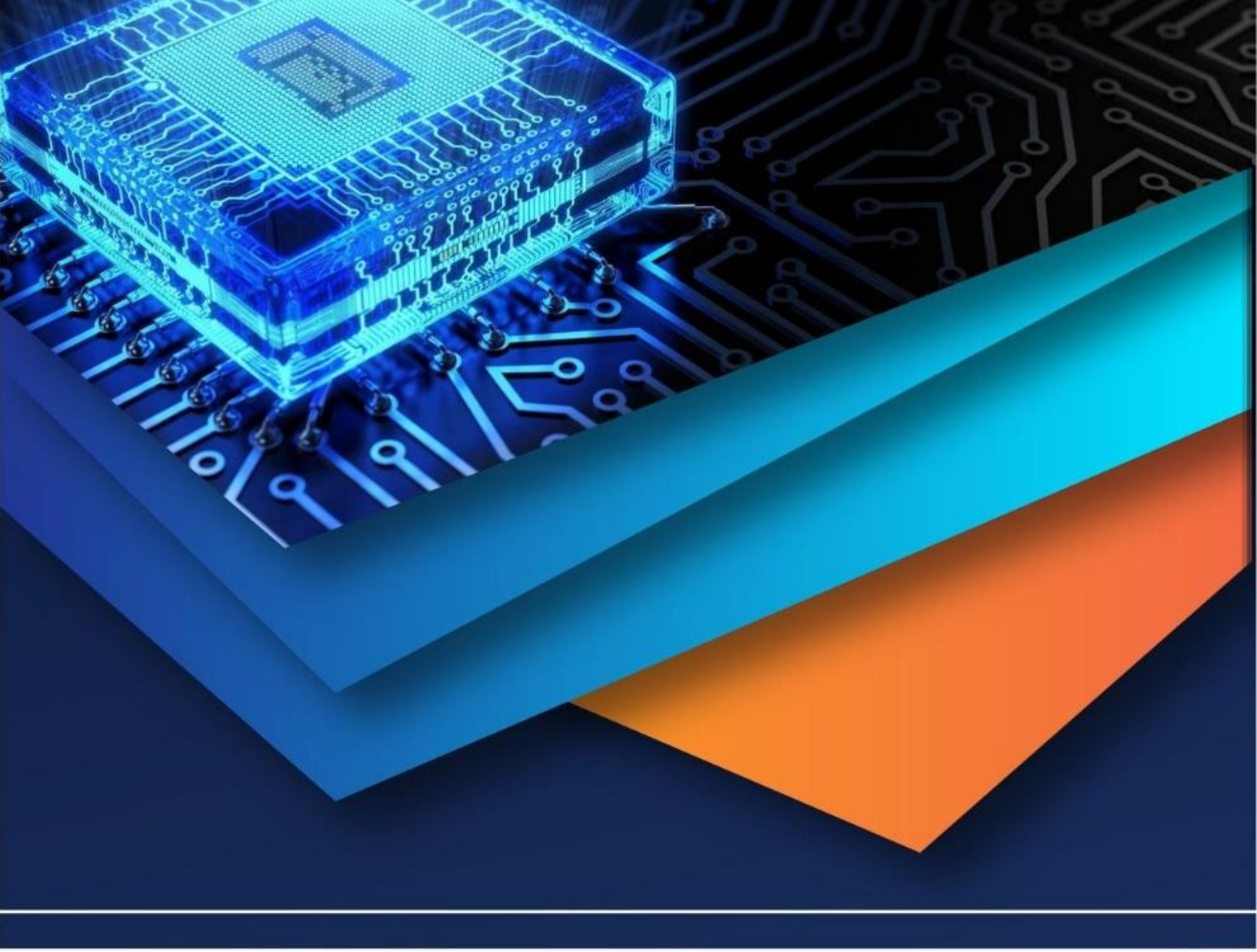
The outcomes demonstrates that, cluster 0 has 808 customers and its mitigation rate is 2.7 %. Though cluster 1 has 878 customer and its mitigation rate is 0%. The most noteworthy mitigation rate is with Cluster-2 and Cluster-3 among all, which is 9.1% and 7.7% individually. The Churn rate for Cluster-1 is 0%, which is great for any company. But churn rate for cluster-2 and cluster-3 are serious issue for the telecom company.

©IJRASET: All Rights are Reserved

REFERENCES

[1] Ngai E.W.T., Li Xiu and Chau D.C.K. (2009), "Application of data mining techniques in customer relationship management: A literature review and classification", Elsevier Expert Systems with Applications, Vol.36, No.3, pp.378-386.
[2] Rekha Bhowmik (2011), "Detecting auto insurance fraud by Data mining techniques", Journal of Emerging Trends in Computing and Information Sciences, Vol.2, No.4, pp.156-162.
[3] Keyvan Vahidy Rodpysh (2012), "Model to Predict the Behavior of Customers Churn at the Industry", International Journal of Computer Applications,Vol.49, No.15, pp.12-16.
[4] Horng-Jinh Chang, Lun-Ping Hung and Chia-Ling Ho (2007), "An anticipation model of potential customers' purchasing behavior based on clustering analysis and association rules analysis", Expert Systems with Applications,Vol. 32, pp.753-764.
[5] Razieh Qiasi, Malihe Baqeri-Dehnavi, Behrooz Minaei-Bidgoli and Golriz Amooee (2012), "Developing a model for measuring customer's loyalty and value with RFM technique and clustering algorithms", The Journal of Mathematics and Computer Science, Vol.4, No.2, pp.172-181.
[6] Baesens B., Verstraeten G., Dirk V. D. P., Michael E. P., Kenhove, V. K. and Vanthienen J. (2004), "Bayesian network classifiers for identifying the slope of the customer-lifecycle of long-life customers", European Journal of Operational Research,Vol.156, pp.508-523.
[7] Sheu J.J., Su Y.H. and Chu K.T. (2009), "Segmenting Online Game Customers - the Perspective of E45xperiential Marketing", Expert Systems with Applications, Vol.36, pp.8487-8495.
[8] Guha S., Rastogi R. and Shim K. (1998), "CURE: An efficient clustering algorithm for large databases", In Proceeding of the 1998, ACM SIGMOD International Conference on Management of Data Engineering, Seattle, WA USA, pp. 512-521.
[9] Hruschka H. (1996), "Market definition and segmentation using fuzzy clustering methods", International Journal of Research in Marketing, Vol. 3, No.2, pp.117-134.
[10] Pramod Prasad and Latesh G. Malik (2011), "Generating customer profiles for retail stores using clustering techniques", International Journal on Computer Science and Engineering, Vol.3, pp.2506-2510.
[11] Hwang H., Jung T. and Suh E. (2004), "An LTV model and customer segmentation based on customer value: A case study on the wireless telecommunication industry", Expert Systems with Applications, Vol. 26, No.2, pp.181-188.
[12] Miguéis V.L., Camanho A.S., João Falcão e Cunha (2012), "Customer data mining for lifestyle segmentation", Expert Systems with Applications, Vol.39, pp. 9359-9366.
[13] Chiu C.C. and Tsai C.Y (2004), "Purchase-based market segmentation methodology",Expert Systems with Applications,Vol.27, No.2, pp-265-276.
[14] Helsen K. and Green P.E. (1991), "A computational study of replicated clustering with an application to market-segmentation", Decision Sciences, Vol.22, No.5, pp.1124-1141.
[15] Chu Chai Henry Chan (2008), "Intelligent value-based customer segmentation method for campaign management: A case study of automobile retailer", Expert Systems with Applications, Vol.34, pp.2754-2762.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)