

A Review on Customer Churn Detection Techniques Using Map Reduce

Adarsh Mishra¹

Department of Computer Science and Engineering
Vishwavidhyalaya Engineering College, Lakhanpur
Sarguja, Chhattisgarh – India

Prof. Mohan Rao Mamdikar²

Department of Computer Science and Engineering
Vishwavidhyalaya Engineering College, Lakhanpur
Sarguja, Chhattisgarh – India

Asst. Prof Dipti Verma³

Department of Computer Science and Engineering
Vishwavidhyalaya Engineering College, Lakhanpur
Sarguja, Chhattisgarh – India

Abstract: *In this competitive world, business is becoming highly saturated. Particularly, the field of telecommunication faces complex challenges because of various vibrant competitive service providers. Subsequently, it has turned out to be exceptionally troublesome for them to retain existing customers. Since the cost of getting new customers is considerably higher than the cost of retaining the current customers, it is the ideal opportunity for the telecom ventures to find a way to retain the customers to balance out their market value. In the previous decade, a few data mining strategies have been proposed in the literature for foreseeing the churners utilizing heterogeneous customer records. This paper reviews the diverse classifications of customer data accessible in open datasets, predictive models and execution metrics utilized as a part of the writing for beat expectation in telecom industry.*

Keywords: *Customer relationship management (CRM), Data mining, Customer churn prediction, Predictive models, and Performance metrics.*

I. INTRODUCTION

With the growth of companies, needs of customer changes rapidly and every customer is currently confronting free products (discount) and number of decisions. So as to get all profit and services, customers change their service provider to other, where they discovered some additional advantage. Changing of service provider by the clients is called "Churning" or "Attrition". The idea of churn prediction turns into the crucial issue in each organization. Acquiring and retention of new clients ends up real issue in the business growth. Attraction of new customers dependably put much pressure on the company. While new companies focus on acquiring new clients, old one focus on retention of existing clients. Retention of old customer makes profit to the company.

Acquiring new customers in an organization takes twelve times more than retaining them. An organization growth relies on number of customers routinely visited, or on the loyal customers. In this way, company constantly utilized DM procedures in making qualification amongst churn and loyal clients with the goal that proper techniques would be connected in future for holding churn clients in future.

Churn customers [6] are the customers who tend to move to another company. Diverse Data Mining systems have been use to isolate the churn and loyal customers for the profit of company by diminishing churn customers in a company. Churn prediction is a technique for separating churners and non-churners, with the goal that fitting advances can be taken to hold them. To control the churn customers in company, it ends up important to build up a successful model for churn prediction. DM

procedures have been utilized to find fundamental data from distribution centers or from other data resources. NN [8], DT, SVM [7] are the most mainstream DM strategies for churn prediction.

II. CHURN CUSTOMERS

Churn customers prompts the loss of company as they are moving starting with one company then onto the next, where they discovered some additional profit. It is difficult to distinguish the customers who are tending to move or leaving the company. In this way, company wants to utilize different DM procedures so as to hold the churn customers as customers are the important people for them. Churn demonstrates the part of consumers that will move their use of support to a competitive support agency.

Churn forecast strategies gives the estimate with respect to purchasers who liable to churn presently while churn tasks, helps on the other give which tries to recognize such corners and to play out some valuable exercises to decrease the churn impact. Shows the part of consumers that will move their use of support to a competitive support agency. Churn forecast procedures gives the conjecture with respect to shoppers who prone to churn in no time though churn tasks helps on the other give which tries to recognize such churners and to play out some advantageous exercises to diminish the churn impact Fig 1:.

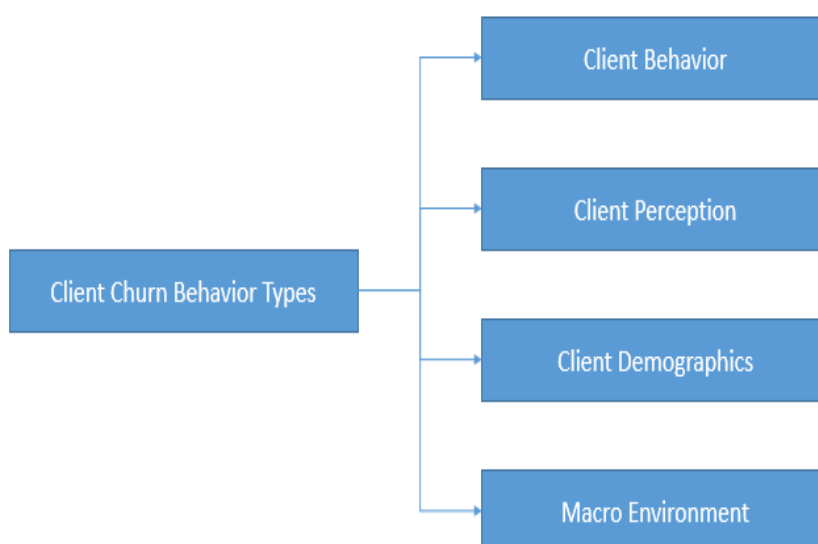


Fig. 1. Churn Behaviour Types

1. Client Behavior-refers to the treatment by the clients. These additionally identifies with the primary parts that are used by the company and how the customers are managing them.
2. Client Perception-it identifies with the connection of the customers to the company. It is fundamentally identifies with the actuation and deactivation of the relation with the company.
3. Client statistic incorporates individual data of customers, likewise utilized for churn count.
4. Macro environment-variable has a place with the progressions accessible on the world and the diverse perspectives of the customers about the product characteristics. Macro climate variable finds the progressions on earth and the different experience of customer which influence the bearing they utilize the service. For example, in the telecommunication exchange the individuals who have survived a natural disaster and could rely upon the cellphones all through it are inclined to keep using the service.

III. CHURN TAXONOMY

Fig. 2. Shows the flow of churn taxonomy.

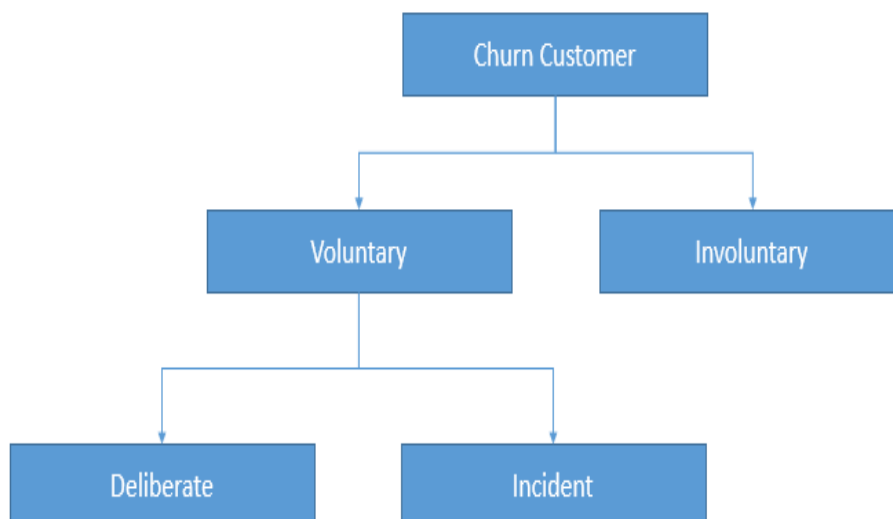


Fig. 2. Churn Taxonomy

Churn customers are to be divided into two kinds: - voluntary and involuntary as shown in Fig: 2. In involuntary churners customers are the most effortless to distinguish as the company disconnect the services to the customers, because of a few reasons, as for fraud, non-payment and so forth. It turns out to be difficult to recognize voluntary churner as it happens when customers naturally terminate his/her service with the provider as they found some profit in other service provider. Voluntary churn can be sub-divided into 2 primary classes, unforeseen churn and furthermore strategic churn. Incidental churn comes about, positively not chiefly on the grounds that the shoppers designed into it albeit primarily on the grounds that a thing happened of their lives.

For instance: distinction in financial ailment churn, contrast in place churn, and so on. Arranged churn happens for good reasons including know-how (customers requiring more complex or maybe prevalent innovation), prompt and continuous costs (value affectability), company top notch viewpoints, sociable or maybe inner angles, and furthermore ease reasons. Arranged churn is unquestionably the test that a ton of churn supervision arrangements endeavor to disentangle non-reflex churn can be sub-divided into 2 primary classifications incidental churn and furthermore key churn.

Incidental churn happens Voluntary churn can be sub-divided into 2 primary classes, unforeseen churn and furthermore key churn. Incidental churn comes to fruition, absolutely not for the most part on the grounds that the consumer designed into it albeit basically on the grounds that a thing happened of their lives. For instance: distinction in financial ailment churn, contrast in place churn, and so on. Arranged churn happens for good reasons including know-how (customers requiring more advanced or maybe unrivaled innovation), quick and continuous costs (value affectability), company great perspectives, sociable or maybe inside viewpoints, and furthermore ease reasons. Prepared churn is unquestionably the test that a considerable measure of churn supervision arrangements endeavor to unwind non-reflex churn can be sub-divided into 2 primary classes' incidental churn and furthermore key churn. Incidental churn comes to fruition, when customer disconnects the service because of a few issues, for instance: financial issue of customers, change in location churn, and so forth. Think churn happens where customer tends to move other company as they discovered advantages or offers in competitor service.

IV. LITERATURE SURVEY

In this section we presents existing work done in the field of customer churn detection.

[1], Churn management is an imperative and critical problem for Global Services of Mobile Communications (GSM) operators to create systems and tactics to keep its subscribers to pass other GSM operators. The primary period of churn management begins with profile creation for the subscribers. Profiling process assesses call detail data, financial information, calls to customer benefit, contract details, market details and geographic and population data of a given state. In this examination, input features are clustered by x-means and fuzzy c-means clustering algorithms to put the subscribers into various discrete classes. Adaptive Neuro-Fuzzy Inference System (ANFIS) is executed to build up a sensitive prediction demonstrate for churn management by utilizing these classes. First prediction step begins with parallel Neuro-fuzzy classifiers. After at that point, FIS takes Neuro-fuzzy classifiers' yields as contribution to settle on a decision churners' exercises.

[2], The key question of this examination is: How long should customer occasion history be for customer churn prediction? While most investigations in predictive churn displaying expect to enhance models by data augmentation or improvement of algorithm, this examination focuses on another measurement: time window optimization with respect to predictive execution. This paper first shows a formalization of the time window selection technique, alongside a literature survey. Next, utilizing logistic regression, classification trees and bagging in combination with classification trees, this investigation examines the change in churn display execution by broadening customer occasion history from one to sixteen years. The outcomes demonstrate that, after the fifth extra year, predictive execution is just possibly expanded, implying that the company in this investigation can dispose of 69% of its data with no diminishing in predictive execution. The down to earth suggestion is that examiners can considerably diminish data-related weights, for example, data stockpiling, planning, and investigation. This is especially profitable in the midst of huge data while diminishing computational intricacy is fundamental.

[3], In this article, we test the helpfulness of the mainstream data mining models to foresee churn of the customers of the Polish cellular telecommunication company. When contrasting with past investigations on this point, our examination is novel in the accompanying regions: (1) we manage prepaid customers (past examinations managed postpaid customers) who are significantly more inclined to churn, are less steady and considerably less is thought about them (no application, demographical or individual data), (2) we have 1381 potential factors got from the customers' utilization (past investigations managed data with no less than many factors) and (3) we test the stability of models crosswise over time for every one of the percentiles of the lift curve – our test is collected six month after the estimation of the model. The principle finding from our examination is that direct models, particularly calculated relapse, are a decent decision when displaying churn of the prepaid customers. Choice trees are flimsy in high percentiles of the lift curve, and we don't suggest their use.

[4], In this competitive world, business is becoming profoundly immersed. Particularly, the field of telecommunication faces complex difficulties because of various competitive service providers. Along these lines, it has turned out to be extremely troublesome for them to hold existing customers. Since the cost of gaining new customers is significantly higher than the cost of holding the current customers, it is the ideal opportunity for the telecom businesses to find a way to hold the customers to balance out their market value. This paper investigates the use of data mining systems in foreseeing the imaginable churners and quality determination on recognizing the churn. It likewise thinks about the productivity of a few classifiers and records their exhibitions for two genuine telecom datasets.

[5], the penetration of mobile telephones is almost immersed in both developing and developed districts. In such a situation, how to anticipate subscribers churn has turned into a vital issue for the present telecom operators, as the cost to secure another supporter is considerably higher than that to hold a current endorser. In this paper, we propose to use the power of huge data to relieve the issue of supporter churn and improve the administration nature of telecom operators. As the information center point, telecom operators have amassed a tremendous volume of important data on endorser practices, benefit utilization, and system

tasks. To empower productive huge data handling, we first form a devoted appropriated cloud framework that incorporates both on the web and disconnected preparing capacities. Second, we build up a total churn investigation demonstrate in light of profound data mining strategies, and use between endorser impact to enhance prediction accuracy. At long last, we utilize genuine datasets acquired from a substantial telecom operator in China to check the accuracy of our churn investigation models. The dataset contains the information of more than 3.5 million subscribers, which create more than 600 million call detail records (CDRs) every month. The exact outcomes exhibit that our proposed strategy can accomplish around 90% accuracy for T + 1 testing periods and distinguish subscribers with high negative impact effectively.

TABLE I. Shows comparison between various existing approaches

| Ref. | Image Dataset | Metrics | Finding | Technique Used | Extracted Features |
|------|---|---|--|---|--|
| [1] | GSM operator, Turkey 24,900 customers 22 attributes | Precision and Recall | Proposed integrated diagnostic system for the churn management application presented is based on a multiple Adaptive Neuro Fuzzy Inference System with fuzzy c-means. | K-Means clustering, Adaptive Neuro Fuzzy Inference System | Demography, Usage pattern, Value added services |
| [2] | Unknown 129,892 customers 113 attributes | AUC | Author have used logistic regression, classification trees and classification trees in combination with bagging to study the relation between the length of customer event history and classification performance. | Logistic regression, Bagging, Decision Tree | Demographic, Value added, usage pattern |
| [3] | Polish mobile operator 122098 customers 1381 attributes | Lift curves | Author evaluated usefulness of regression and decision trees approach to the problem of modelling churn in the prepaid sector of the cellular telecommunication company. | Logistic regression Decision tree | Demographic, call data records, customer care services |
| [4] | Cell2Cell Dataset 100,000 customers 171 attributes | Confusion matrix, accuracy, precision, recall, F1-score | Author explores the application of data mining techniques in predicting the likely churners and attribute selection on identifying the churn. | Gradient Boosting, Decision Tree, Support Vector Machine, Random Forest, K-NN, Ridge Regression and Logistic Regression | Behavioral information, Customer care and demographics |
| [5] | 600 million call detail records | Accuracy, precision, recall, F1-score | Author proposed a systematic methodology to predict the probability of subscriber churn, identify users with high negative influence, and help the selection of correlation attributes. | Logistic regression, decision tree methods. | Behavior information is extracted from the custom relationship management system |

V. CONCLUSION

Customer churn has been recognized as a noteworthy issue in Telecom industry and aggressive research has been directed in this by applying different data mining strategies. Numerous data mining methods are usually applied in customer churn. This paper aims at surveying on the current literature in the field of telecom customer churn.

References

1. Adem Karahoca, Dilek Karahoca, "GSM churn management by using fuzzy c-means clustering and adaptive neuro fuzzy inference system", *Expert Systems with Applications* 38 (2011) 1814–1822.
2. Ballings, Michel, and Dirk Van den Poel. "Customer event history for churn prediction: How long is long enough?." *Expert Systems with Applications* 39, no. 18 (2012): 13517-13522.
3. Marcin Owczarczuk, "Churn models for prepaid customers in the cellular telecommunication industry using large data marts", *Expert Systems with Applications*, 37, 2010, pp. 4710–4712.
4. V. Umayaparvathi, K. Iyakutti,, "Attribute Selection and Customer Churn Prediction in Telecom Industry", *Proceedings of the IEEE International Conference On Data Mining and Advanced Computing*, 2016 (to be appeared).
5. Li H, Wu D, Li GX et al. Enhancing telco service quality with big data enabled churn analysis: Infrastructure, model, and deployment. *JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY* 30(6): 1201–1214 Nov. 2015. DOI 10.1007/s11390-015-1594-2
6. P. K. Dalvi, S. K. Khandge, A. Deomore, A. Bankar and V. A. Kanade, "Analysis of customer churn prediction in telecom industry using decision trees and logistic regression," 2016 Symposium on Colossal Data Analysis and Networking (CDAN), Indore, 2016, pp. 1-4.
7. Yujun Yang, Jianping Li and Yimei Yang, "The research of the fast SVM classifier method," 2015 12th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), Chengdu, 2015, pp. 121-124.
8. Jurgen Schmidhuber, "Deep Learning in Neural Networks: An Overview", 2014, *Neural Networks* 61 (2015) 85–117.