

# Prediction and Minimizing of Churn Rate using Machine Learning based on user financial attributes

C K Vanamala

Dept. of ISE

National Institute of Engineering  
Mysuru, India

Veena Mohan

MDept. of ISE

National Institute of Engineering  
Mysuru, India

**Abstract**—Client churn forecast models plan to recognize clients with a high inclination to draw in. Precise precision, intelligibility, and reasonability are three vital parts of a churn forecast model. An exact model grants to accurately target future churners in a maintenance showcasing effort, while a fathomable and natural rule-set permits to distinguish the fundamental drivers for clients to churn, and to foster a compelling maintenance methodology as per space information. The client churn expectation (CCE) is one of the difficult issues in the telecom business. With the progression in the field of AI and man-made brainpower, the potential outcomes to foresee client churn has expanded altogether. This paper gives a drawn out outline of the writing on the utilization of AI in client churn expectation and minimization displaying. This paper sums up the churn expectation strategies to have a more profound comprehension of the client churn and it shows that most precise churn forecast is given by the mixture models as opposed to single calculations so telecom enterprises become mindful of the requirements of high gamble clients and improve their administrations to topple the churn choice.

**Keywords**—Churn, Network, Financial attributes, churn rate

## I. INTRODUCTION

The turn of events and digitalization of the world has prompted better approaches for carrying on with work and organizations all around the globe have been compelled to adjust [1]. Membership based administrations are one of the results of the unstable digitalization that has surprised the world and with this comes the two prospects and difficulties that require current arrangements [2]. Digitalization has impacted how business is directed as well as the overflow of data accessible has likewise prompted buyers confronting a higher stock of membership based administrations. This can be seen as quite difficult for organizations since holding clients might possibly turn out to be more troublesome. Digitalization inside organizations can prompt a diminishing in labor costs, an expansion in productivity and a superior outline of the organization's tasks inside the association [1].

This is all fundamental for remaining serious, and to acquire an edge over other companies. Customer beating alludes to the activity of when a client decides to leave their specialist organization [3]. The term is somewhat new and has acquired pertinence with the development of online administrations. Firms across the globe perceive client beating as an extraordinary misfortune since they have proactively put resources into drawing in these clients. This is one of the significant reasons that client maintenance is helpful for a firm. Clients can churn for some reasons and it is difficult to pinpoint a general justification for churning.

The accessibility of data has given buyers a bartering power, and these days clients can undoubtedly find the specialist co-op, which furnishes a similar item with a seriously fulfilling bargain [4]. To deal with this, organizations put resources into churn forecast, and that implies that organizations attempt to foresee which of their clients will churn, so they can apply precaution measures. These preventive measures could contrast relying upon the explanation a client could churn, and could be for instance, offering a lower cost or including an additional help. As referenced before, dissecting client conduct fills in as the reason for anticipating clients who could churn, which is significant for some reasons. One explanation is that for organizations who depend on membership based pay, it can have a major effect on whether they can keep a consistent pay level or on the other hand assuming they need to make changes to their administrations to keep clients. That's what one more explanation is, contrasted with holding clients, drawing in new ones is costlier and firms can set aside cash by holding their current client base [5].

## II. LITERATURE REVIEW

According to Horia Beileu et al. [6], they used three machine learning techniques to predict customer churn: neural networks, support vector machines, and bayesian networks. Principle component analysis (PCA) is used in the feature selection process to condense the dimensions of the data. Yet, an optimization technique can be used to enhance the feature selection process and boost classification accuracy. Gain measure and ROC curve were employed in the performance evaluation.

The authors J. Burez et al. [7] attempted to define the issue of class imbalance. They employed logistic regression and random forest with re-sampling technique. Boosting algorithms were additionally used. AUC and Lift are taken into account in the performance analysis. They also looked at the impact of cutting-edge sampling methods like CUBE, but the results showed no improvement in performance. The problem of class imbalance can still be better resolved by adopting optimization-based sampling strategies, though.

The authors of K Coussement et al. [8] used support vector machines, logistic regression (LR), and random forests to attempt to capture the churn prediction problem (RF). SVM initially performed about as well as LR and RF, but when the best parameter choices were taken

into account, SVM outperformed both of them in terms of PCC and AUC.

The two machine learning models, decision trees and logistic regression, were used on the churn prediction data-set by K. Dahiya et al. [9] researchers. WEKA tool was employed during the trial. However, by utilising further machine learning approaches, the aforementioned issue can be effectively resolved.

J. Hadden et al. [10] reviewed all machine learning models taken into account and provided a thorough study of the methods currently used for feature selection. They discovered that decision trees outperformed the competition in the prediction models. The improvement of the prediction algorithms in feature selection is greatly aided by optimization techniques. After comparing and contrasting the various methods already in use, the authors suggested a course for further research.

According to Y. Huang et al. [11], the authors used a variety of classifiers on the churn prediction dataset, and the findings showed that random forest outperforms the competition in

terms of AUC and PR-AUC analyses. But, accuracy can be increased much further by employing feature extraction optimization approaches. For instance, a higher volume of customer service contacts should result in a higher likelihood of turnover. To our knowledge, the only other discussion of domain knowledge in churn prediction modelling is in [12].

The researchers used support vector machines, decision trees, naive bayes, and logistic regression in their comparative investigation of machine learning models for customer churn prediction. After that, they looked at how boosting methods affected classification accuracy [13].

The implementation of two unique data mining algorithms for predicting customer attrition is presented by the authors in this research. The first method, AntMiner+, expressly aims to provide precise, understandable, and intuitive classification rule-sets by using Ant Colony Optimization (ACO) to infer rules from data [14].

TABLE I. TABLE TYPE STYLES

Ref	Techniques used	Dataset used	Extracted features	Research gap
[13]	Random forest algorithm	South asian GSM's data	Calling services, value added services	Time consuming, high error rate
[14]	Decision tree, SVM, Random forest, K-NN and logistic regression	Cell2cell dataset 100000 customer 171 attribute	Behavioral information, customer care	Behavioral information generate the churn possibility sometimes it generate false ratio
[15]	Naive Bayes, Decision tree	European operator 106,405 customer 112 attribute	Contract usage pattern, calls pattern	High error rate to detect actual churn
[16]	Neural Network, Regression	European operator 106,405 customer 112 attribute	Contract usage pattern, calls pattern	Heterogeneous dataset tedious to handle in similar pattern environment
[17]	X-means clustering algorithm and neuro fuzzy algorithm	GSM operation data 24900 customers 22 attribute	Value added services	System reflects good accuracy on structured dataset only

### III. METHODOLOGY

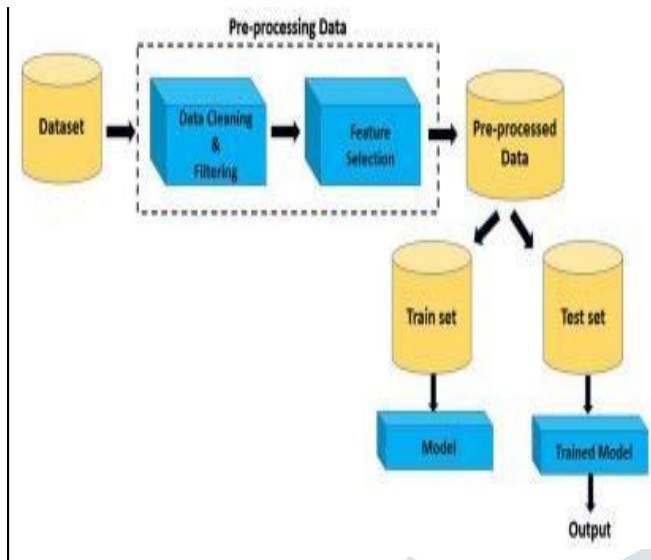
#### A. Machine Learning Models

##### 1) Regression analysis

Logistic Regression Analysis Relapse is one of the factual cycles for assessing how the factors are connected with one another. It incorporates adequate measures of methods for laying out the model and examining a few factors, when the focal point of significance is on the security which is divided among a reliant variable and one or numerous free factors. In the radiance of client stirring, relapse examination isn't extensively utilized on the grounds that direct relapse models are helpful for anticipating ceaseless qualities. Be that as it

may, Calculated Relapse or Logit Relapse investigation (LR) is a probabilistic factual characterization model. It is likewise utilized for twofold characterization or double expectation of an all out esteem (e.g., house rate expectation, client stir) which relies on at least one boundary (e.g., house highlights, client highlights). In resolving the perplexing issue of client agitate forecast issue, information initially must be casted under legitimate information change from the underlying information to accomplish great execution and some of the time it performs [18] as great as Choice Trees.

Fig. 1. Architectural Diagram



### 2) Naïve Bayes

Naïve Bayes classifier is a probabilistic methodology where every vector include is thought of as free of one another. Naïve Bayesian classifiers expect that the worth of each component impacts a given class, and this supposition that is called class contingent freedom that is utilized to work on the calculation

### 3) Support Vector Machine

In AI, Backing Vector machines otherwise called Help Vector Organizations presented are regulated learning models with related learning calculations that examine information utilized for order and relapse examination. What support vector machine is attempting to do is, it separates the expectation into two sections +1 that is correct side of the hyperplane and - 1 that is left half of the hyperplane. The hyperplane is of width two times the length of edge. Contingent upon the kind of information for example (dispersed on the chart) tuning boundary like pieces are utilized like straight, poly, rbf, callable, pre- determined [19]. Support Vector machine gives high exactness than Naïve Bayes and Strategic Relapse.

### 4) Decision Tree

It deals with the voracious methodology and utilizations a progression of rules for characterization. On the other hand, this approach clarifies the high order precision rate it neglects to answer information having commotion. The fundamental boundary to conclude the root hub boundary of choice tree is gain. The choice trees created by C4.5 can be utilized for characterization and hence C4.5 is frequently alluded to as a measurable classifier.

### 5) XG Boost Classifier

XGBoost carries out choice tree calculation with angle supporting [20]. The slope helping follows a methodology where new models are utilized to process the blunder or residuals of recently applied model and afterward both are joined to make the last expectation. It likewise utilizes inclination drop to find the minima or decrease the worth of misfortune capability.

## IV. CONCLUSION

Stir rate expectation addresses one of the main parts of Client Relationship The executives (CRM). Membership Items frequently are the primary wellspring of income for organizations across all ventures. organizations quite often attempt to limit client beat (a.k.a. membership undosings). In the reason for holding clients and keeping up with their fulfillment, specialists of many fields including business knowledge, advertising and data innovation are persuaded to examine the best strategies that convey the best administrations for clients. This venture presents an examination investigation of the presentation towards beat forecast between probably the most impressive AI calculations like SVM, Irregular Backwoods and Strategic Relapse. The outcomes are contrasted with reach inferences and examine agitate designs

## ACKNOWLEDGMENT

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g”. Avoid the stilted expression “one of us (R. B. G.) thanks ...”. Instead, try “R. B. G. thanks...”. Put sponsor acknowledgments in the unnumbered footnote on the first page.

## REFERENCES

- [1] R. Shkurti and A. Muça, "AN ANALYSIS OF CLOUD COMPUTING AND ITS ROLE IN ACCOUNTING INDUSTRY IN ALBANIA," Romanian-American University, vol. 8, no. 2, pp. 219- 229, 2014.
- [2] C. Blank and T. Hermansson, "A Machine Learning approach to churn prediction in a subscription based service," KTH, Stockholm, 2018.
- [3] D. L. Garcia, A. Nebot and A. Vellido, "Intelligent data analysis approaches to churn as a business problem: a survey," Knowledge and Information Systems, vol. 51, no. 3, pp. 1-56, 2017.
- [4] N. Gordini and V. Veglio, "Customers churn prediction and marketing retention strategies. An application of support vector machines based on the AUC parameter-selection technique in B2B e-commerce industry," Industrial Marketing Management, vol. 62, pp. 100-107, 2017.
- [5] I. Ullah, B. Raza, A. K. Malik, M. Imran, S. U. Islam and S. W. Kim, "A churn prediction model using random forest: analysis of machine learning techniques for churn prediction and factor identification in telecom sector," IEEE Access, pp. 60134-60149, 2019.
- [6] Brândușoiu, I., Todorean, G., Beleiu, H.: Methods for churn prediction in the pre-paid mobile telecommunications industry. In: 2016 International conference on communications (COMM), pp. 97– 100. IEEE (2016).
- [7] Burez J, Van den Poel D (2009) Handling class imbalance in customer churn prediction. Expert Systems with Applications 36(3):4626–4636.
- [8] Coussement K, Van den Poel D (2008) Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. Expert systems with applications 34(1):313–327.
- [9] Dahiya, K., Bhatia, S.: Customer churn analysis in telecom industry. In: 2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions), pp. 1–6 (2015).
- [10] Hadden J, Tiwari A, Roy R, Ruta D (2007) Computer assisted customer churn management: State-of-the-art and future trends. Computers & Operations Research 34(10):2902–2917.
- [11] Huang, Y., Zhu, F., Yuan, M., Deng, K., Li, Y., Ni, B., Dai, W., Yang, Q., Zeng, J.: Telco churn prediction with big data. In: Proceedings of the 2015 ACM SIGMOD international conference on management of data, pp. 607–618 (2015).
- [12] Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., & Vanthienen, J. (2003). Benchmarking state-of-the-art classification

- algorithms for credit scoring. *Journal of the Operational Research Society*, 54(6), 627–635.
- [13] Irfan Ullah, Basit Raza , Ahmad Kamran Malik , Muhammad Imran, Saif Ul Islam And Sung Won Kim "A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector" *IEEE ACCESS*.2019.2914999, VOLUME 7, 2019
- [14] Karahoca Adem, and Dilek Karahoca. "GSM churn management by using fuzzy c-means clustering and adaptive neuro fuzzy inference system." *Expert Systems with Applications* 38.3 (2011): 1814-1822.
- [15] Kirui, Clement, et al. "Predicting customer churn in mobile telephony industry using probabilistic classifiers in data mining." *International Journal of Computer Science Issues (IJCSI)* 10.2 Part 1 (2013): 165.
- [16] Ballings, Michel, and Dirk Van den Poel. "Customer event history for churn prediction: How long is long enough?." *Expert Systems with Applications* 39.18 (2012): 13517-13522.
- [17] Ismail, Mohammad Ridwan, et al. "A multi-layer perceptron approach for customer churn prediction." *International Journal of Multimedia and Ubiquitous Engineering* 10.7 (2015): 213-222.
- [18] 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).
- [19] Ahmed, A., & Linen, D. M. (2017, January). A review and analysis of churn prediction methods for customer retention in telecom industries. In *2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS)* (pp. 1-7). IEEE.
- [20] Zadoo, A., Jagtap, T., Khule, N., Kedari, A., & Khedkar, S. (2022, May). A review on Churn Prediction and Customer Segmentation using Machine Learning. In *2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON)* (Vol. 1, pp. 174-178). IEEE.

