International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 05 | May 2022 www.irjet.net p-ISSN: 2395-0072

EVALUTION OF CHURN PREDICTING PROCESS USING CUSTOMER BEHAVIOUR PATTERN

Kavita B Khatal, Dr. M.A. Wakchaure

Computer Department, Amrutvahini college of Engineering, Sangmner. Dist. Ahmednagar

---

Abstract – *Client churn is a prime problem and one of the maximum essential issues for massive corporations. because of the direct impact at the income of the agencies, specially withinside the telecom area, organizations are searching for to increase approach to are watching for capability purchaser to churn. consequently, finding factors that boom consumer churn is crucial to take crucial movements to lessen this churn. The primary contribution of our art work is to make bigger a churn prediction model which assists telecom operators to are looking ahead to customers who are most in all likelihood subject to churn. The version developed on this art work makes use of tool analyzing techniques on large statistics platform and builds a contemporary way of capabilitiesâ€™ engineering and selection. in an effort to diploma the general performance of the version, This art work additionally identified churn factors which are critical in identifying the premise motives of churn. by way of know-how the huge churn elements from clients' information, CRM can decorate productiveness, advise applicable promotions to the enterprise of probably churn clients primarily based totally on comparable conduct styles, and excessively enhance advertising campaigns of the business enterprise.*

***Key Words*: churn prediction, NLP,lexicacal evaluation,**

1. INTRODUCTION

Consumers these days undergo a complicated choice making manner earlier than subscribing to any individual of the severa Telecom provider options. The offerings furnished via way of means of the Telecom providers aren't relatively differentiated and wide variety portability is commonplace. The cellular smartphone enterprise churn is the same trouble [2] [9] [12]. Hence, it's far turning into more and more more vital for telecommunications groups to proactively perceive elements that have a propensity to unsubscribe and take preventive measures to preserve clients. To calculate your probably month-to-month churn, begin with the wide variety of customers who churn that month. Then divide via way of means of the whole wide variety of consumer days that month to get the wide variety of churns in step with consumer day. Then multiply via way of means of the wide variety of days withinside the month to get your ensuing month-to-month churn rate. It is observed that information mining strategies are extra powerful in predicting client churn from the studies performed over the last few years [17]. Creating an green churn prediction version is an vital pastime requiring a number of paintings proper from figuring out suitable predictor variables (features) from the massive quantity of to be had consumer information to deciding on an powerful predictive information mining approach appropriate for the function set.

The A multi-layer perceptron technique for consumer churn prediction has used in [14] for consumer-associated information inclusive of consumer profiling, calling pattern, and democratic information further to the community information they generate. Based at the consumer's records of calling behaviour and behaviour, there's a opportunity to categorise their mindset of both going away or not. Data mining strategies are observed to be extra powerful in predicting churn from the studies performed over the last decade. The predictive modelling strategies in churn prediction also are taken into consideration to be extra accurate. Churn prediction structures and sentiment evaluation the usage of category in addition to clustering strategies to categorise churn clients and the motives at the back of the churning of telecom clients [18]. In telecom enterprise have to we generate massive quantity of information on every day basis, it's far very tedious challenge to mine this kind of sort of closing information the usage of particular information mining strategies, even as tough to interpret the prediction on classical strategies. Sometime such telecommunication information can be containing a few churn and, it's far a lot essential to perceive seek problems. To a success identity of churn from massive information is presenting effectiveness to consumer courting management (CRM) [3]. Customer retention is one of the maximum vital troubles for groups. Customer churn prevention, as a part of a Customer Relationship Management (CRM) technique, is excessive at the agenda. Big groups put into effect churn prediction fashions as a way to stumble on feasible churners earlier than they efficiently depart the company [16].

2. LITERATURE SURVEY

Many methods like machine learning and data processing are used for churn prediction. The decision-tree algorithm could be a reliable method for churn prediction [6]. additionally, a neural network method [7], data certainty [8], and particle swarm optimization [9] are used for churn prediction.

According to system [2] a current collection of software to extend the quality of detecting possible churners. The roles are extracted from request information and client accounts and are classified as deal, request pattern and call pattern adjustments overview functionality. The characteristics are evaluated using two probabilistic data processing algorithms from Naïve Bayes and Bayesian Network, and their findings compared to those obtained by the utilization of C4.5 decision tree, an algorithm widely utilized in many classification and prediction tasks. Among other reasons these have led to the likelihood that customers will quickly communicate competitors. one in every of the techniques which will be wont to do that is to enhance churn prediction from great deal of knowledge with extraction within the near future.

According to [3] formalization of time-window of the gathering process, let alone literature review. Second, by expanding the duration of consumer events from one to seventeen years using logistic regression, classification trees and bagging along with classification trees, this analysis analyzes the increase in churn model accuracy. the sensible result's that researchers may substantially reduce the data-related pressures, like data collection, preparation, and analysis. the value customers are expected to pay depends on the length and therefore the pro-motional nature of the subscription. The newspaper business is sending a letter telling them that the service is ending. Then ask them if they require to renew their subscription, together with guidance on a way to do this. Customers are unable to cancel their subscription and have a grace period of 4 weeks once they need subscribed lapsed.

According to [4] the most efficient consumer engagement strategies can be used to high the client satisfaction level efficiently. The study indicates a Multilayer Perceptron (MLP) neural network method to estimate client turnover in one of Malaysia's leading telecommunications firms. The results were contrasted with the most traditional churn prediction strategies such as Multiple Regression Analysis and Analyzing Logistic Regression. The maximal neural network architecture includes 14 input nodes, 1 concealed node and 1 output node with the learning algorithm Levenberg Marquardt (LM). Multilayer Perceptron (MLP) neural network approach to predict client churn in one of the leading telecommunications companies in Malaysia compared to the most common churn prediction techniques, such as Multiple Regression Analysis and Logistic Regression Analysis. In system [5] on creating an efficient and descriptive statistical churn model utilizing a Partial Least Square (PLS) approach focused on strongly associated intervals in data sets. A preliminary analysis reveals that the proposed model provides more reliable results than conventional forecast models and recognizes core variables in order to better explain churning behaviors. Additionally, network administration, overage administration and issue handling approaches are introduced in certain simple marketing campaigns and discussed. Burez and Van den Poel [6] Unbalance data sets studies in churn prediction models, and contrasts random sampling performance, Advanced Under-Sampling, Gradient Boosting Method, and Weighted Random Forest. The concept was evaluated using Metrics (AUC, Lift). The study shows that the methodology under sampling is preferable to the other techniques evaluated.

Gavril et al. [7] Describes an innovative data processing method to elucidate the broad dataset form of consumer churn detection. About 3500 consumer details is analyzed supported incoming number likewise as outgoing input call and texts. Specific machine learning algorithms were used for training classification and research, respectively. The system's estimated average accuracy is about 90 percent for the whole dataset.

He et al. [8] during a with approximately 5.23 million subscribers, a significant Chinese telecommunications corporation developed a predictive model focused on the Neural Network method to handle the difficulty of consumer churn. the typical degree of precision was the extent of predictability of 91.1%.

Idris [9] suggested a gene-splicing solution to modeling AdaBoost-churning telecommunications problems. Two Standard Data Sets verified the series. With a precision of 89%, one from Orange Telecom and also the other from cell2cell and 63% for the opposite one.

Huang et al. [10] the customer churn studied on the massive data platform. The researchers ' aim was to indicate that big data significantly improves the cycle of churn prediction, supported the number, variety and pace of the information. A broad data repository for fracture engineering was expected to accommodate data from the Project Support and Business Support Department at China's biggest telecommunications firm. AUC used the forest algorithm randomly and assessed.

According to [11] with k-means and fuzzy c-means clustering algorithms are clustered input features to position subscribers in separate discrete groups. The Adaptive Neuro Fuzzy Inference System (ANFIS) is introduced using these classes to construct a predictive model for active churn management.

The first prediction step begins with Neuro fuzzy parallel classification. FIS then takes Neuro fuzzy classifiers outputs as input to decide on churners activities. Measurements of success can be used to recognize inefficiency problems. Churn management metrics are associated with customer service network services, operations, and efficiency. GSM number versatility is a vital criterion for churner's determination. In System [12] a New set of apps to improve the identification level of potential churners. The features are derived from call details and customer profiles and are

categorized as description features related to contract, call pattern, and call pattern changes. The features are tested using two Naive Bayes and Bayesian Network probabilistic data mining algorithms and their results compared to those obtained from the use of C4.5 decision tree, an algorithm commonly used in many classification and prediction tasks. These have contributed, among other factors, to the risk that customers can easily switch to competitors. One of the techniques that can be used to do this is to improve churn prediction from large amount of data with extraction in the near future. According to [13] Formalization of the selection method in time window, along with analysis of literature. Second, this study analyzes the increase in churn model consistency by extending the history of customer events from one to seventeen years using logistic regression, classification trees and bagging along with classification trees. The functional consequence is that researchers, such as data storage, planning and research, can significantly reduce data-related burdens. The amount that consumers have to pay depends on the subscription's duration and pro-motional sense. A letter is sent by the newspaper company to remind them that the subscription is coming to an end. Then ask them if they want to renew their subscription, along with guidance on how to do that. Customers are unable to cancel their subscription and have a grace period of four weeks once they have subscribed lapsed.

According to [14] the foremost effective customer retention techniques should be wont to effectively reduce customer turnover rates. The research suggests a neural network approach for Multilayer Perceptron (MLP) to predict customer churn in one in every of Malaysia's leading telecommunications firms. The findings were compared with the foremost common techniques of churn prediction like multiple correlation Analysis and Logistic multivariate analysis.

A preliminary experiment shows that the model presented provides more accurate performance than traditional models of prediction and identifies key variables to higher understand churning behaviors. Additionally, there's a spread of basic churn marketing strategies— system management, overage management, and complaint management strategies is presented and discussed.

Burez and Van den Poel [16] studied the matter of unbalance datasets in churn prediction models and compared performance of sampling, Advanced Under-Sampling, Gradient Boosting Model, and Weighted Random Forests. They used (AUC, Lift) metrics to judge the model. the result showed that undersampling technique outperformed the opposite tested techniques.

3. PROPOSED SYSTEM

In the proposed research work to design and develop an approach for churn prediction using NLP and machine learning approaches to enhance the system accuracy. Then we identify the customer changing behavior pattern during prediction [4]. We also evaluate the factor which mostly influences to reduce accuracy of churn prediction and finally evaluate and calculate churn rate for month wise as well as day wise, which useful for enhance the service quality of system. In this research we proposed churn prediction from large scale data, system initially deals with telecommunication synthetic data set which contains some imbalance meta data. To apply data preprocessing, data normalization, feature extraction as well as feature selection respectively [17]. During this execution some Optimization strategies have been used to eliminate redundant features which sometimes generate high error rate during the execution. The proposed system execution for training and testing. After completion both phases system describe classification accuracy for entire data set.

4. SYSTEM OVERVIEW

The churn rate, also called the speed of attrition or customer churn, is that the rate at which customers stop doing business with an entity [4]. it's most typically expressed because the percentage of service subscribers who discontinue their subscriptions within a given fundamental measure. The churn rate in developing markets ranges from 20% to 70%. In a number of these markets, over 90% of all mobile subscribers are on prepaid service. Some operators in developing markets lose during aggregate their entire subscriber base to churn in a year [5]. We identify some research gap supported entire literature review that are mention in below

- various researchers have already done the churn prediction model but most of the system having accuracy issues because of imbalanced data.

- Sometimes data contain mini miss-classified instances it's hard classifies by supervised learning algorithm.

- Most of the systems have used structured data so there's no scope for NLP feature extraction.

- Hard to detect consistent with user wise supported sentiment analysis.

- Low churn prediction accuracy

- Time complexity very high thanks to RF generate and extract redundant features.

• High error rate of system

It is our intention to gather data from the primary popular online Customer reviews website for churning predictions [3] [4]. Predict future Churn Prediction using machine learning algorithms. The system can add a stable and real time environment and might predict the most effective accuracy [5].

• In different category class labels to categorize online customer reviews.

• Identify customer-changing behavior patterns during forecasting

5. CONCLUSIONS

This research mainly focuses on identifying and detecting churn consumers from massive data set of telecommunications, state-of-the-art discusses churn prediction systems produced by different research. Some systems still face problems of conversion of linguistic data, which can occur at high error rate during execution. Many researchers have been putting forward Natural Language Processing (NLP) techniques as well as various machine learning algorithms such a combination is likely to generate good performance when structuring data. If any machine learning algorithm interacts with that kind of a method, it is necessary to test or confirm the entire data set with even sampling techniques that reduce data imbalance problems and provide reliable predictive flow of data. For future direction to implement a proposed system with various machine learning algorithm to achieve better accuracy, as well as the input data contains large size and volume, if we deal the proposed systems with HDFS framework and parallel machine learning algorithm which will provide better result in low computation cost.

REFERENCES

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] Lee, Hyeseon, et al. "Mining churning behaviors and developing retention strategies based on a partial least squares (PLS) model." Decision Support Systems 52.1 (2011): 207-216.

[6] Burez D, den Poel V. Handling class imbalance in customer churn prediction. Expert Syst Appl. 2009;36(3):4626–36.

[7] Brandusoiu I, Toderean G, Ha B. Methods for churn prediction in the prepaid mobile telecommunications industry. In: International conference on communications. 2016. p. 97–100.

[8] He Y, He Z, Zhang D. A study on prediction of customer churn in fixed communication network based on data mining. In: Sixth international conference on fuzzy systems and knowledge discovery, vol. 1. 2009. p. 92–4.

[9] Idris A, Khan A, Lee YS. Genetic programming and adaboosting based churn prediction for telecom. In: IEEE international conference on systems, man, and cybernetics (SMC). 2012. p. 1328–32.

[10] Huang F, Zhu M, Yuan K, Deng EO. Telco churn prediction with big data. In: ACM SIGMOD international conference on management of data. 2015. p .607–18

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] Lee, Hyeseon, et al. "Mining churning behaviors and developing retention strategies based on a partial least squares (PLS) model." *Decision Support Systems* 52.1 (2011): 207-216.

[16] Burez, Jonathan, and Dirk Van den Poel. "Handling class imbalance in customer churn prediction." Expert Systems with Applications 36.3 (2009): 4626-4636.

[17] Schouten, Kim, et al. "Supervised and unsupervised aspect category detection for sentiment analysis with co-occurrence data." IEEE transactions on cybernetics 48.4 (2017): 1263-1275.

[18] 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.

[19] Kamalraj, N., and A. Malathi. "A survey on churn prediction techniques in communication sector." International Journal of Computer Applications 64.5 (2013): 39-42.