



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 11, November 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.524

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com

Churn Prediction and Factor Identification using Machine Learning

Sanskar Rathod¹, Yogeshwar Sawdekar², Ajay Shinde³, Prerna Ahire⁴, Prof. A. V. Bendale⁵

UG Scholars, Department of Information Technology, Sandip Institute of Technology and Research Centre (SITRC),
Nashik, India^{1,2,3,4}

Professor, Department of Information Technology, Sandip Institute of Technology and Research Centre (SITRC),
Nashik, India⁵

ABSTRACT: Customer churn is a major problem and one of the most important concerns for large companies. Because customer churn directly impacts a company's revenues, particularly in the telecom sector, companies are actively developing methods to anticipate potential customer churn. Therefore, it is crucial to identify the factors that contribute to customer churn and implement the necessary measures to decrease this churn. The main contribution of our work is to develop a churn prediction model that assists telecom operators to predict customers who are most likely subject to churn. This work develops a model that employs machine learning techniques on a big data platform, introducing a novel approach to feature engineering and selection. In order to measure the performance of the model, this work also identified churn factors that are essential in determining the root causes of churn. By identifying the significant churn factors from customers' data, CRM can enhance productivity, recommend relevant promotions to the group of likely churn customers based on similar behaviour patterns, and significantly enhance the company's marketing campaigns.

KEYWORDS: Receiving Operating Characteristics, Machine learning, churn prediction Classification, Feature Selection.

I. INTRODUCTION

Consumers today go through a complex decision-making process before subscribing to any one of the numerous Telecom service options. The services provided by the Telecom vendors are not highly differentiated and number portability is commonplace. The mobile telephone industry churn is the similar problem [2] [9]. Customer loyalty becomes an issue. Hence, it is becoming increasingly important for telecommunications companies to proactively identify factors that have a tendency to unsubscribe and take preventive measures to retain customers. To calculate your probable monthly churn, start with the number of users who churn that month. Then divide by the total number of user days that month to get the number of churns per user day. Then multiply by the number of days in the month to get your resulting monthly churn rate. It is found that data mining techniques are more effective in predicting consumer churn from the research conducted over the past few years. Creating an efficient churn prediction model is an essential activity requiring a lot of work right from determining appropriate predictor variables (features) from the large volume of available customer data to choosing an effective predictive data mining technique suitable for the feature set.

The A multi-layer perceptron approach for customer churn prediction has used in for customer-related data such as customer profiling, calling pattern, and demographic data in addition to the network data they generate. Based on the customer's history of calling behaviour and behaviour, there is a possibility to classify their attitude of either going away or not. Data mining techniques are found to be more effective in predicting churn from the research done over the past decade. The predictive modelling techniques in churn prediction are also considered to be more accurate. Churn prediction systems and sentiment analysis using classification as well as clustering techniques to classify churn customers and the reasons behind the churning of telecom customers. In telecom industry should we generate large amount of data on daily basis, it is very tedious task to mine such a kind of last data using specific data mining techniques, while hard to interpret the prediction on classical techniques. Sometime such telecommunication data may be containing some churn and, it is much necessary to identify search problems. To successful identification of churn from large data is providing effectiveness to customer relationship management (CRM) [3]. Customer retention is one

of the most important issues for companies. Customer churn prevention, as part of a Customer Relationship Management (CRM) approach, is high on the agenda.

Big companies implement churn prediction models to be able to detect possible churners before they effectively leave the company.

II. LITERATURE SURVEY

According to [1] Clustered input features are clustered input characteristics that position subscribers in discrete groups using k-means and fuzzy c-means clustering algorithms. These classes are used to build the Adaptive Neuro-Fuzzy Inference System (ANFIS), a prediction model for active churn control. Neuro-fuzzy parallel categorization is the initial phase in the prediction process. FIS then uses the outputs of the Neuro-fuzzy classifier to determine the actions of the churners. An identifies inefficiency issues there might employ success metrics. The Customer service network services, operations, and efficiency are linked to churn management measures. The adaptability of GSM numbers is an essential consideration for cherner selection.

In System [2] A new collection of applications aimed at improving the detection of prospective churners The characteristics are obtained from call information and client profiles and grouped as contract, call pattern, and pattern change description features. The parts are examined using two probabilistic data mining techniques, Nave Bayes and Bayesian Network. The results are compared to those produced using a C4.5 decision tree, which is frequently employed in many classification and prediction applications. These have led to the possibility that consumers may readily migrate to rivals, among other things. Improving churn prediction from significant amounts of data using extraction is one strategy they may utilize to achieve this.

According to [3] it is the Formalization of the time window selection procedure and a literature review. Second, this research examines the rise in churn model consistency by utilizing logistic regression, classification trees, and bagging in conjunction with classification trees to extend the history of customer events from one to seventeen years. As a result, researchers may greatly minimize data-related difficulties in data storage, planning, and research. The amount that customers must pay is determined by the length of the subscription and the promotional value. The newspaper firm sends a letter to inform them that their subscription will expire. Then ask whether they want to renew their membership and provide instructions on how. Customers cannot cancel their subscriptions, although they have a four-week grace period once their membership has expired.

According to [4] to efficiently lower client turnover rates, they should implement the most effective customer retention tactics. In Malaysia's largest telecommunications companies, the study recommends a neural network technique for Multilayer Perceptron (MLP) to forecast customer turnover. They compared the results to the most popular churn prediction methodologies, including Multiple Regression Analysis and Logistic Regression Analysis. With the Levenberg Marquardt (LM) learning method, the best architecture of the neural network has 14 input nodes, one hidden node, and one output node. Compared to the most general churn prediction approaches, such as Multiple Regression Analysis and Logistic Regression Analysis, a Multilayer Perceptron (MLP) neural network methodology was used to forecast customer churn at one of Malaysia's largest telecoms businesses.

In system [5] they use a Partial Least Square (PLS) Methodology based on highly correlated data set intervals to build an accurate and straightforward predictive churn model. An early experiment demonstrates that the model given outperforms existing prediction models and highlights crucial characteristics for better understanding churning behaviour. Additionally, specific fundamental marketing approaches are provided and reviewed, including system management, overage management, and complaint management measures.

Burez and Van den Poel [6] it analyses the efficiency of random sampling, Advanced Under-Sampling, Gradient Boosting Model, and Weighted Random Forest in churn prediction models using unbalanced data sets. The model was evaluated using metrics (AUC, Lift). The test revealed that the under-sampling strategy outperformed the other methods.

Gavril et al. [7] Provides a sophisticated data mining method for identifying customer turnover in a vast dataset. They analysed the data of over 3500 customers based on the quantity of arriving and departing input calls and messages. For

the categorization of training and testing, they applied several machine learning techniques. The system's projected accuracy is about 90% on the total dataset.

He et al. [8] created a prediction model based on the Neural Network algorithm to handle the customer turnover issue in a big Chinese telecoms firm with around 5.23 million consumers. The total accuracy percentage was 91.1 %, indicating a high degree of predictability.

Idris [9] For AdaBoost, he proposed using genetic programming to mimic telecoms churning concerns. Two standard data sets have validated the sequence. One from Orange Telecom and the other from cell2cell had 89 %, while the other had 63%.

Huang et al. [10] the client turnover was investigated using a big data platform. The researchers wanted to show that, depending on the amount, diversity, and speed of the data, big data significantly improves the churn prediction process. A massive data store for fracture engineering was required to cope with data from China's biggest telecommunications company's Department of Operation Support and Business Support. AUC used the random forest method to evaluate.

Summary Table

Author and title	Features	Method	Evaluation Matrix	Limitations
Irfan Ullah et.al. [11] A churn prediction model using random forest: analysis of machine learning techniques for churn prediction and factor identification in telecom sector	Information Gain Values Correlation attribute ranking values	Various machine learning supervised algorithms	Accuracy ROC	It works only structured and relational data
Clement Kirui [12] Predicting customer churn in mobile telephony industry using probabilistic classifiers in data mining	Contract, usage pattern patterns, and calls pattern	Naïve Bayes, Decision Tree	Confusion matrix, accuracy, precision, recall	Some continuous attributes generate high time complexity.
Ballings, Michel [13] Customer event history for churn prediction: How long is long enough?	Demographic, Value added, usage pattern	Neural network, Regression	AUC	Many probabilities to churn threshold it is hard to detect to actual churn and accuracy issues
Ismail, Mohammad [14] A Multilayer Perceptron Approach for Customer Churn Prediction	Demographic, Billing data, usage pattern, customer relationship	Neural network, Regression	Confusion matrix, accuracy, precision, recall	Very small dataset has used, not enough information for identification churn it generate high error rate.
H Lee [15] Mining churning behaviours and developing retention strategies based on a partial least squares (PLS) model", Decision Support Systems	Behavioural information, Customer care and demographics	Stepwise variable selection partial least squares	Proportion of hit records	Low classification accuracy.

III. PROPOSED SYSTEM

This paper offers a method for predicting churn from big data sets. The system starts with a telecoms synthetic data set that includes imbalanced metadata. To make data preparation, data normalization, feature extraction, and feature selection, as needed. During this execution, they applied several optimization tactics to remove duplicate features that might cause a high error rate during the performance. The proposed system execution for training and testing is shown in Figure 1. After all, steps are completed; the system describes the categorization accuracy for the entire data set.

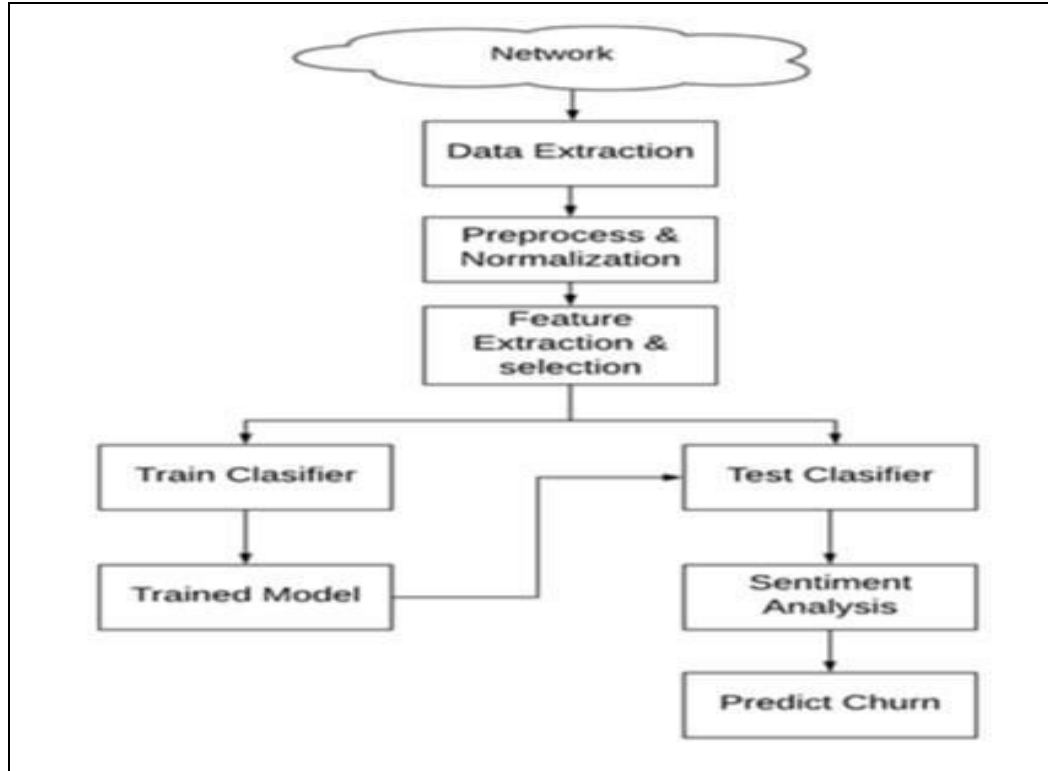


Figure 1: Research methodology of proposed system

IV. IMPLEMENT PROCESS

Data Pre-Processing:

Data Pre-processing – Before working on data it’s necessary to pre-process it. Data may contain missing values that leads to poor result to avoid this, pre-processing is necessary. Filtering and noise removal: In filtering process unwanted and unnecessary features are removed and only keep useful features. In noise removal, null values, space, missing characters are removed.

Feature Extraction:

Feature extractions remove the unwanted data from the data set and keep only accurate and complete data that should be processed.

Data Training: We compile artificial as well as real time using churn data and provide training with any machine learning classifier.

Testing with machine learning: We predict churn data using any machine learning classifier, weight calculator for real time or synthetic input data accordingly.

Analysis: We demonstrate the accuracy of proposed system and evaluate with other existing systems.

APPLICATIONS

BPO centre’s churn prediction systems.

Service application churns prediction systems.

Customer behaviours mapping system using churn prediction.

V. CONCLUSUION

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.

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 behaviours 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 System Appl.* 2009;36(3):4626–36.
7. Brandusoiu I, Todorean 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. ULLAH, Irfan, et al. A churn prediction model using random forest: analysis of machine learning techniques for churn prediction and factor identification in telecom sector. *IEEE Access*, 2019, 7: 60134-60149.
12. Kirui, Clement, Li Hong, Wilson Cheruiyot, and Hillary Kirui. "Predicting customer churn in mobile telephony industry using probabilistic classifiers in data mining." *IJCSI International Journal of Computer Science Issues* 10, no. 2 (2013): 1694-0784.
13. 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.
14. Ismail, Mohammad Ridwan, Mohd Khalid Awang, M. Nordin A. Rahman, and Mokhairi Makhtar. "A Multilayer Perceptron Approach for Customer Churn Prediction." *International Journal of Multimedia and Ubiquitous Engineering* 10, no. 7 (2015): 213-222.
15. H Lee, Y Lee, H Cho, Kim, YS Kim, "Mining churning behaviours and developing retention strategies based on a partial least squares (PLS) model", *Decision Support Systems*, Volume 52, Issue 1, 2011, Pages 207–216



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details