|| Volume 7 || Issue 8 || October 2023 || ISSN (Online) 2456-0774
INTERNATIONAL JOURNAL OF ADVANCE SCIENTIFIC RESEARCH
AND ENGINEERING TRENDS

Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector

Chetana Vasant Patil[1], Prof. R. V. Patil[2]

Student, Department Computer Engineering, S.S.V.P.S.'s B.S. Deore College of Engineering, Dhule, India
Professor, Department of Computer Engineering, S.S.V.P.S.'s B.S. Deore College Of Engineering, Dhule, India

Abstract: - **Customer churn is a major problem and one of the most important concerns for large companies. Due to the direct effect on the revenues of the companies, especially in the telecom field, companies are seeking to develop means to predict potential customer to churn. Therefore, finding factors that increase customer churn is important to take necessary actions to reduce this churn. The main contribution of our work is to develop a churn prediction model which assists telecom operators to predict customers who are most likely subject to churn. The model developed in this work uses machine learning techniques on big data platform and builds a new way of features' engineering and selection. In order to measure the performance of the model, this paper also identified churn factors that are essential in determining the root causes of churn. By knowing the significant churn factors from customers' data, CRM can improve productivity, recommend relevant promotions to the group of likely churn customers based on similar behavior patterns, and excessively improve marketing campaigns of the company. The proposed churn prediction model is evaluated using metrics, such as accuracy, precision, recall, f-measure. Furthermore, it also provides factors behind the churning of churn customers through the rules generated by using the attribute-selected classifier algorithm**

Key terms*:* Receiving Operating Characteristics, Deep learning, Convolution Neural Network, churn prediction, Feature selection.

I INTRODUCTION

Telecommunications firms now produce a vast amount of data at a lightning pace. Associate editor Tariq Umer oversaw the review process and gave final approval before publishing. Range of telecom service providers compete in the market to expand their customer share. Services may be found elsewhere for less money or with higher quality. In order to survive in a cutthroat industry, telecom businesses must maximize their profits [21]. 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. Telecom Industries collect a large amount of customer-related data such as customer profiling, calling pattern, and democratic 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. Various researchers already described search a work to eliminate churn from large data sets fusion static as well as dynamic approaches, but still such systems are facing many problems actual identification of churn. 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). Because telecom operators need to keep their important clients and improve their Customer Relationship Management (CRM) administration, churn prediction is crucial in the telecom industry. Keeping current clients happy is CRM's toughest assignment. Customers may easily move to a different service provider because to the crowded and competitive industry. Because acquiring new clients is more costly for telecoms than keeping existing ones [5], the industry has created systems to better track and keep track of its clientele

II LITERATURE REVIEW

According to [1] process, most popular predictive models have been applied, namely, logistic regression, naive bayes, support vector machine, random forest, decision trees, etc. on train set as well as boosting and ensemble techniques are applied to see the effect on accuracy of models. In addition, K-fold cross validation has been used over train set for hyper parameter tuning and to prevent over fitting of models.

According to [2] the loss of customers is a huge issue and a top priority for every business. Companies, particularly in the telecommunications industry, are trying to find ways to forecast client turnover because of the direct impact on income. Therefore, it is crucial to identify the causes of client turnover in order to take measures to decrease it. Our primary contribution is a churn prediction algorithm that helps telecom providers identify which customers are at risk of leaving. This work's model presents a novel approach to features' engineering and selection by using machine learning methods on a large data platform. The model's efficacy is evaluated using the widely used Area Under Curve (AUC) metric, and a value of 93.3% is found. The extraction of Social Network Analysis (SNA) elements from consumer social networks for use in the prediction model is another significant advance. The model's performance relative to the AUC benchmark was improved by using SNA, from 84 to 93.3%. Through the use of the Spark environment, the model was developed and validated by processing a sizable dataset derived from the transformation of massive raw data given by the SyriaTel telephone firm. The dataset was utilized to train, test, and evaluate the system at SyriaTel and included data from all customers during a 9-month period.

According to [3] Clustering algorithms are clustered input functions with k-means and fuzzy c-means to position subscribers in independent, distinct classes. Using these groups the Adaptive Neuro Fuzzy Inference Framework (ANFIS) is implemented to construct a predictive model for successful churn management. The first step towards prediction starts with the parallel classification of Neuro soft. FIS then uses the outputs of Neuro fuzzy classifiers as feedback to settle on the behaviors of the churners. Progress metrics can be used to identify issues of inefficiency. Churn reduction indicators are concerned with the facilities, processes and performance of customer support network. Versatility of GSM numbers is a critical criterion for churners determination.

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 mining method to explain the broad dataset type of consumer churn detection. About 3500 consumer details is analyzed based on incoming number as well 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 entire dataset.

He et al. [8] in a with approximately 5.23 million subscribers, a major Chinese telecommunications corporation developed a predictive model focused on the Neural Network method to address the issue of consumer churn. The average degree of precision was the extent of predictability of 91.1%.

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

Huang et al. [10] the customer churn studied on the big data platform. The researchers ' aim was to show that big data significantly improves the cycle of churn prediction, based on the quantity, variety and pace of the data. 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 at random and assessed.

III Motivation

Having the ability to accurately predict future churn rates is necessary because it helps your business gain a better understanding of future expected revenue. Predicting churn rates can also help your business identify and improve upon areas where customer service is lacking. With this research we proposed a churn for telecom sector using machine learning to eliminate future revenue losses.

IV PROPOSED SYSTEM DESIGN

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. 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. 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

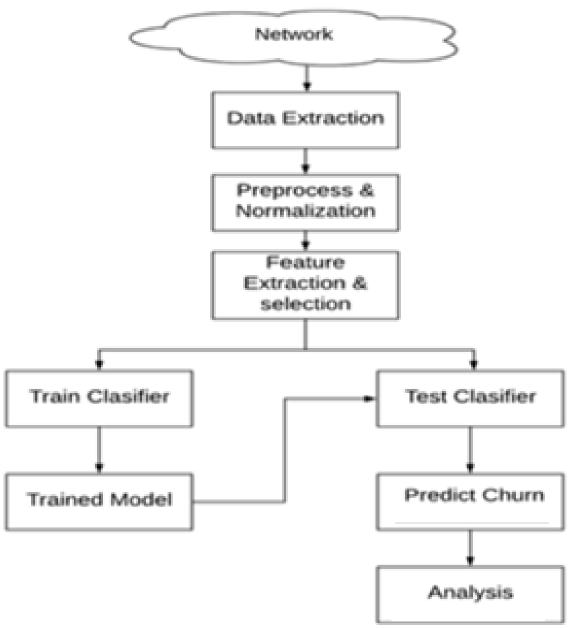

Figure 1: System Architecture

Problem Definition

In the proposed research work to measure identify the churn using text analysis using NLP and machine learning classifier. To identify the customer changing behavior pattern during prediction. To identify the factor which mostly influence to reduce accuracy of churn prediction? To evaluate and calculate churn rate for month wise as well as day wise, which useful for enhance the service quality of system.

Objectives

- To design and develop an approach for Churn Prediction with Sentiment Analysis on customer reviews large dataset.
- To implement proposed system with various feature extraction as well as selection techniques and evaluate the performance analysis of system.
- To validate the proposed system with respective machine learning algorithm and deploy on real time environment.
- To explore and validate the proposed system comparative analysis on various dataset with classification accuracy

Algorithm

Random Forest

Input: Selected feature of all test instances D[i….n], Training database policies {T[1]…………T[n]}

Output: No. of probable classified trees with weight and label.

Step 1: Read (D into D[i])

V ←Extract features (D)

Step 2: N ←Count Features(D)

Step 3: for each (c into Train DB)

Step 4: Nc[i] ←Ext Features(c)

Step 5: select relevant features of w= {Nc[i], N}

Step 6: Statement (w>t)

Step 9: Return Tree Instance { Nc[i], N, w, label}

Artificial Neural Network

Artificial Neural Network

Input: Traininput TrF[], Testfeatures TsF[], Threshold T, Feedback_count n

Output: Refine weight for each object.

Step 1: Read Train_feature TrF

Step 2: Read Train_feature TsF

Step 3: for each (tsf into TsF)

Step 4: for each (trf into TrF)

If (feed-back_count != n)

Step 5: send feed layer to tsf again

tsF ← feed-Layer [] execute for all neurons early stodolgy

Step 6: optimized feed-Layer weight

Step 7:weight = FeedLayer[0]

Step 8 : return current_weight

NB

Input: Train features TF [], Test features Ts[], threshold T,

|| Volume 7 || Issue 8 || October 2023 || ISSN (Online) 2456-0774
INTERNATIONAL JOURNAL OF ADVANCE SCIENTIFIC RESEARCH
AND ENGINEERING TRENDS

Output: Classification of weight

Step 1: vector is given as an input

Step 2: Each values in the given vector is extracted

Step 3: The extracted values is searched in the dataset

Step 4: for each (x[] into TF[] when!=null)

Step 5: Get all features x1[]=Ts[]

Step 6: w=CalDistance(x[], x1[])

Step 7: evaluate w with T

Step 8: Classify weight

V RESULT

The implementation process was completed in a Java open-source setting. The device operates on the Java 3-tier analytics platform with a distributed INTEL 3.0 GHz i5 CPU and 4 GB RAM. Whether an email is spam or not has been determined uses the APK dataset. We have performed experiment analysis on ensemble machine implementation to verify the outcomes**.**

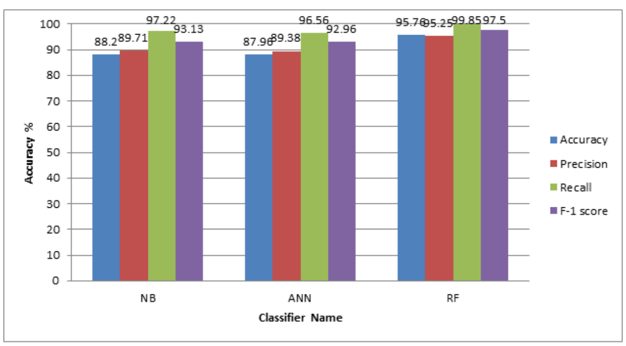

Figure 2: Accuracy of system analysis

Figure 2 shows the suggested system's classification accuracy and a comparison to several state-of-the-art systems. The figure above shows the detection accuracy of dataset in churn or not churn detection using different machine learning and deep learning classifications. The suggested classifier has been used to identify churn data in malicious or not, with a high accuracy rate of up to 95.40%.

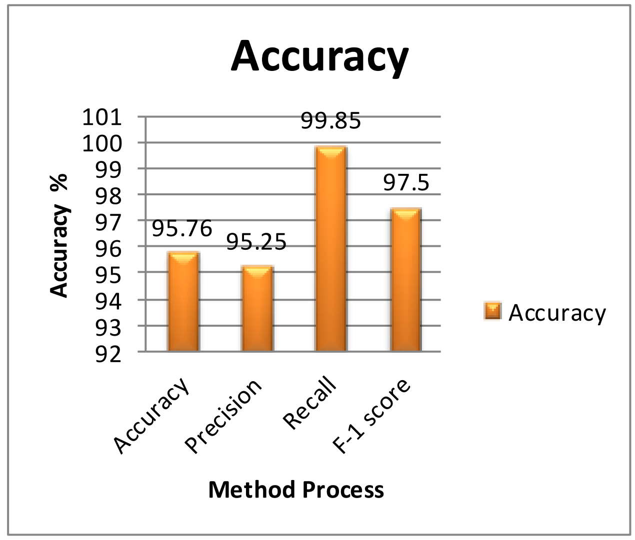

Figure 3: Churn Prediction System accuracy using RF

The Figure 3 shows the classification accuracy of the system; this system achieves around 95.76% accuracy for Churn Prediction using RF algorithm.

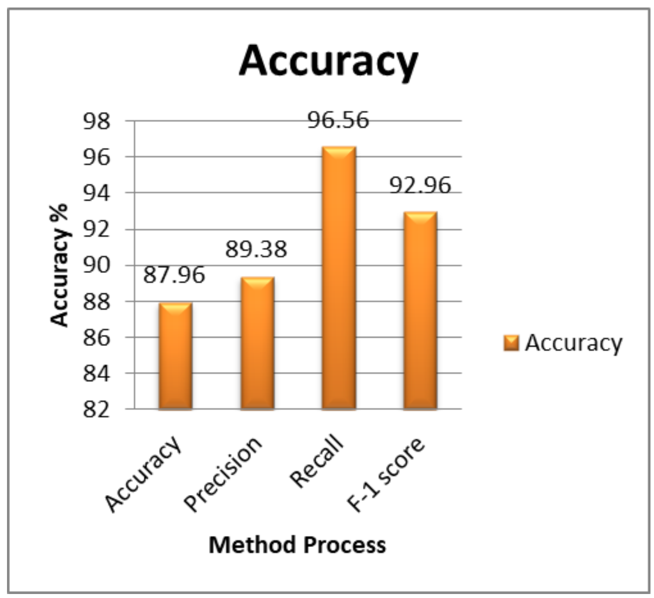

Figure 4: Churn Prediction System accuracy using ANN

The Figure 4 shows the classification accuracy of the system; this system achieves around 87.96 % accuracy for Churn Prediction using ANN algorithm.

|| Volume 7 || Issue 8 || October 2023 || ISSN (Online) 2456-0774
INTERNATIONAL JOURNAL OF ADVANCE SCIENTIFIC RESEARCH
AND ENGINEERING TRENDS

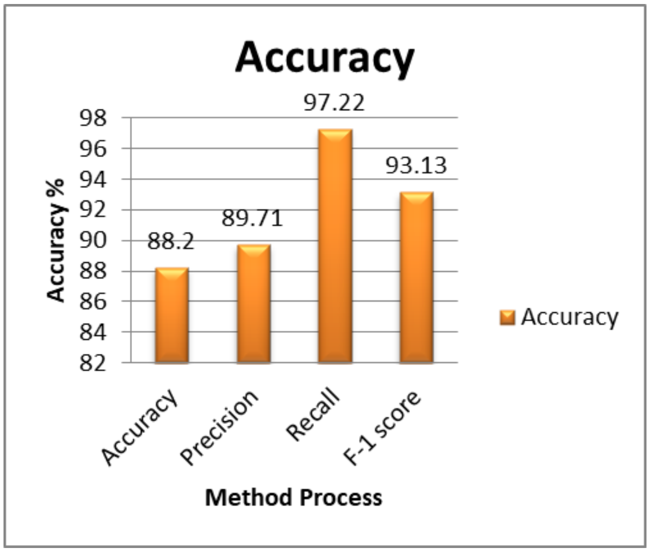

Figure 5: Churn Prediction System accuracy using NB

The Figure 5 shows the classification accuracy of the system; this system achieves around 88.20 % accuracy for Churn Prediction using NB algorithm.

VI CONCLUSION

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] 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.

[2] 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.

[3] 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.

[4] 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.

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

[6] 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.

[7] 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.

[8] 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.

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