International Journal of Future Generation Communication and Networking
Vol. 13, No. 4, (2020), pp. 344–363

Customer Churn Prediction Using Machine Learning Technique With Certainty Computation

[*1]S.Mahalakshmi,

Research Scholar, Department of Computer Science, BharathiarUniversity,Coimbatore.Email ID : mahakanishka23@gmail.com

[2]Dr. M. HemaLatha,

Professor, Department of Computer Science, Sri Ramakrishna College of Arts &Science ,Coimbatore.Email ID : hema.bioinf@gmail.com

Abstract

Rapid growth of the digital data, industries become hastily digitalized. The Telecom industry is one of the most revenue-generating ones. Here the telecom industry faces customer churn problems. Customers who were wanted to switch their network to another network. Predicting these customers and providing a better option to customers will help balance the revenue of their industry. Customer Churn Prediction (CCP) is a confronting activity for machine learning and decision maker's community. Non-churning and churning customers hold some similar kind of resembling features. With various experimentation over customer churn and corresponding data, it is identified that classifier provides various accuracy levels for various dataset zones. In some situations, the correlation among features is observed easily in various classifier accuracy and prediction certainty. In this case, if any mechanisms are defined to evaluate classifiers certainty in various zones along with data, then predictable classifier accuracy could be evaluated before classification. Here, a novel prediction model is provided based on validating classifier certainty using a distance measure. Datasets are clustered into various zones based on distance measures that are classified as two categories: data with higher certainty level and data with lower certainty level for predicting behavioral nature of churn and non-churn customers. A distance measure will perform with the gaming theory. Performance measures like accuracy, precision, recall, and f-measure on diverse publicly available telecom industry datasets are computed to demonstrate the classifier's certainty level. Classifier acquires higher accuracy with distance measure, that is, churn/non-churn with higher and lower certainty.

Keywords- *Churn/Non-Churn Customers, Prediction Model, Distance Measure, Gaming Theory, Certainty Computation.*

I. INTRODUCTION

Telecommunication services usually suffer a customer churn problem, hence the existing customer may get unsatisfied from their telecom service and switch to another company. It provides a loss to their company while holding the present customer is better than receiving a new customer. Every telecom industry has faced these kinds of problems. In past years' telecom companies have made many changes due to the competition of companies; those companies provide new services and technologies [1]. So people were interacted with that and get interested in the new offers and services, so they were a churned customer. In another way, churners are those persons who may quit a company's service due to several reasons. Hence such industries provide better services, like price plan, data plan, call plan, etc., they also provide responses to their complaints, and it will provide a better approach for the customer who can switch to the other network. For these new companies and new services, today, industries face churn problems [2]. That churn problem is especially in the service industries; the here most affected industry is the telecom industry. Hence early detection of such customers will help to retain the customer for the telecom industries. Prediction of the data mining approach will help to preserve the customers.This research work provides attention to predicting customer churn in Telecom industries [3].

Hence, the process of churn management is pre-essential to the development of the organization and assists with getting to the present circumstance of the organization and set designs for the advancement of the organization. The pace of customer churn in an association is viably overseen by utilizing Customer Relationship Management (CRM) persons. The early acknowledgment of customer churn empowers the association to focus more on such customers and build up certain particular maintenance activities to know the explanation and hold productive and loyal customers. In telecom services, service provider affects a lot because of this customer churn. Customer may change their operator voluntarily and un-voluntarily. Involuntary churn, the provider not able to ask them to stay at all. Here providing the best offers they the customer may be in a stable to their network; otherwise, it cannot be, in some cases, by providing offers also they cannot stay with that operators. This prediction method will help to analyze and find the churned customer [4]. Classification also one of the approaches to find the churned customer, but prediction provides more helpful to find exact churn customers. Churn prediction used to perceive customers who are well on the way to churn. Churn prediction and examination can assist an organization with developing a manageable methodology for customer maintenance programs [5]. By getting familiarity with the level of churners, we can undoubtedly define point by point examination, reasons for the churn, and customer maintenance programs. Having the option to foresee customer churn ahead of time gives an organization a high significant knowledge to hold and increment its customer base.

Prediction of telecom churners has always been an area of researcher's interest, and hence many researchers have worked in this direction to predict telecom customer churn. Prediction of churn is a challenging task for telecom services. Hence they have a large number of customers [6]. And also high imbalanced data between the churners and non-churners the occurrence of high cardinality definite features and the fact that the dataset will be profound and large to be attempted by the typical machine learning algorithms. The opportunity of customers to churn from the telecommunication companies has materialized as to be a real risk for telecommunication companies because of the lost income and the high cost of attainments of new customers [7]. And so, now predicting the churn is significant for telecommunication companies, even though that churn is unavoidable, it can be accomplished and predicted so a corrective measure can be taken to keep it at a safe level. This research paper studied and analyzed the machine learning approach, which is well-upgraded technology in many areas and recommended to use in this research work for providing the best solutions for churn prediction in telecoms services. Machine learning is regularly begat with the terms design in data and utilizing them to settle on predictions or choices [8]. Machine learning includes building algorithms that can gain from data accessible and can be utilized to make predictions on data. Here a model is worked from the given information data, which at that point used to make predictions on new data. As the data being gathered is radically expanding every day, this requires the requirement for machine learning [9].

Churn prediction has generally concentrated in the previous decade. Keaveney [10], in 1995, distributed an early and compelling investigation about customer churn in the specialist organization division by reviewing to discover the reasons customers switch service. Many machine learning-based approaches implemented to predict churn in recent days to tackle the issue of customer churn. To name a few linear regressions [11], Association rules [12], Decision tree [13], classification based on churn analysis using Support Vector Machine (SVM) [14, 15].Customers' churn is a significant worry in administration segments with high serious administrations. They were numerous methodologies used to actualize the prediction of churn in telecom. Then again, anticipating the customers who are probably going to leave the organization will speak to conceivably enormous extra income sources on the off chance that it has done in the early stage. Moreover, by applying the proposed way to deal with the freely accessible dataset, the results show that the proposed strategy can completely anticipate every one of those customers that will churn or conceivably may churn and gives valuable data to key chiefs also.

II. RELATED WORK

Amin, Adnan, SaeedShehzad, Changez Khan, Imtiaz Ali, and Sajid Anwar [16] discover the Customer churn is a significant action in quickly developing and develop serious telecommunication part and is one of the best significance for an undertaking director. This examination is another endeavor that utilizes a harsh set hypothesis, a standard based decision-making technique, to extricate rules for churn prediction. Trials performed to investigate the presentation of four distinct algorithms (Exhaustive, Genetic, Covering, and LEM2). It has seen that harsh set characterization dependent on the hereditary algorithm; rules age yields the most appropriate execution out of the four rules age algorithms. Their proposed technique can completely anticipate every one of those customers that will churn or potentially may churn and gives helpful data to vital decision producers.

Praveen Asthana [17] presents a close to write about the most standard machine learning methods applied to the troublesome issue of a customer churning prediction in the broadcast communications business. In the essential time of our assessments, all models were applied and surveyed using cross-approval on a standard, open space dataset. We played out a movement of Monte Carlo entertainments for each methodology and a wide extent of parameters to choose the most capable parameter blends. Our results show away from the models' helped variations against the plain (non-upheld) interpretations. The best all around classifier was the SVM-POLY using AdaBoost with the exactness of essentially 97% and F-measure over 84%.

Prashanth, R., K. Deepak, and Amit Kumar Meher [18] separate the churn prediction is a noteworthy factor to consider for Customer Relationship Management (CRM). In this examination, quantifiable and data were burrowing strategies used for churn prediction. The maker uses straight (vital backslide) and non-direct strategies of Random Forest and Deep Learning structures, including Deep Neural Network, Deep Belief Networks, and Recurrent Neural Networks for prediction. It is seen that non-straight models played out the best. Such farsighted models can be used in the telecom business for settling on better choices and customer management.

Lu, Ning, Hua Lin, Jie Lu, and Guangquan Zhang [19] find the improvement of mechanized frameworks and related data headways. There is a developing example in the overall economy to build modernized customer relationship management (CRM) frameworks. This article coordinates an authentic report on customer churn prediction and proposes the use of boosting to update a customer churn prediction model—this paper utilize a boosting algorithm. Determined backslide used in this investigation as a reason understudy, and a churn prediction model dependent on each pack, independently. The result differentiated and a singular key backslide model. Exploratory evaluation reveals that boosting in like manner gives a better than average division of churn data; thusly, boosting recommended for churn prediction assessment. Bi, Wenjie, MeiliCai, Mengqi Liu, and Guo Li [20] publicize competition fortifies; customer churn management is logically transforming into a noteworthy strategy for high ground for associations. In light of these difficulties, another gathering algorithm called a semantic-driven subtractive grouping methodology (SDSCM) is proposed. Preliminary results show that SDSCM has more grounded gathering semantic quality than subtractive grouping method (SCM) and soft c-infers (FCM). By then, an equivalent SDSCM algorithm realized through a HadoopMapReduce structure. For the circumstance study, the proposed equivalent SDSCM algorithm acknowledges a fast running rate when differentiated and various methodologies. Gaur, A. also, Dubey, R. [21], discover the data mining techniques and algorithm assumes a significant job for organizations in the present business conditions because increasing another customer's expense is more than holding the current ones. In this paper, we can center around different machine learning techniques for foreseeing customer churn through which we can manufacture the characterization models, for example, Logistic Regression, SVM, Random Forest, and Gradient supported tree and look at the presentation of these models.

Ahmad, Abdelrahim et al [22] utilize machine learning methods on big data stage and assembles another method for highlights' building and determination. The area under curve (AUC) normal measure

ISSN: 2233-7857 IJFGCN

contained, and the AUC esteem acquired is 93.3% to quantify the presence of the model. Another fundamental commitment is to utilize customer interpersonal organization in the estimate model by removing Social Network Analysis (SNA) highlights. The utilization of SNA upgraded the model's exhibition from 84 to 93.3% beside the AUC standard.Kim, Kyungtae, and Jee-Hyong Lee. [23] enhances the customer churn prescient model utilizing Bayesian Optimization. The customer churns prescient model utilized as a significant instrument in Customer Relationship Management. Be that as it may, hyper-parameters must be set fittingly for high precision. In this paper, we streamline seven hyper-parameters of the customer churn prescient model utilizing the Recurrent Neural Network. The analysis shows that the exactness of the prescient model can be fundamentally improved. Ullah, Irfan et al [24] proposed model initially characterizes churn customer's data utilizing order algorithms, in which the Random Forest (RF) algorithm performed well with 88.63% accurately grouped examples. Making successful maintenance arrangements is a fundamental assignment of the CRM to forestall churners. After characterization, the proposed model sections the churning customer's data by classifying the churn customers in bunches utilizing cosine likeness to give bunch based maintenance offers. This paper additionally distinguished churn factors that are fundamental in deciding the underlying drivers of churn.

Dulhare, Uma N., and IfrahGhori [25] examine the customer churns can be limited by investigating the previous history of the potential customers of an organization systematically. Also, decision creators constantly confronted with a loose activity management issue. A crossbreed technique made out of Axiomatic Fuzzy Set (AFS) and equal DBSCAN grouping thought of managing these issues. The test results show that Parallel DBSCAN takes substantially less running velocity at that point, equal K-implies bunching. Mitkees, et al [26]. Find a big issue that experiences organizations, particularly telecommunications businesses, are 'customer churn.' This occurs when a customer selects to leave an administration's landline business for additional link candidate. Our point past this examination is to fabricate a model that will foresee churn customers by characterizing the customer's exact conduct and qualities. We will utilize data mining techniques, for example, clustering, characterization, and relationship rule. All things measured, if the group doesn't know about a customer who is going to consent their business, no legitimate move can be made by that organization towards that customer. ShrishaBharadwaj ; B.S. Anil ; AbhirajPahargarh ; AdhirajPahargarh [27] propose two models which predict customer churn with a high level of precision. Their first method is a calculated relapse model, a non-direct classifier with sigmoid as its initiation work. The model's exactness increased by normalizing it with the legalizing parameter set to 0.01, which gives a precision of 87.52% on our test dataset. Our subsequent model is an undeniable Multilayer Perceptron(MLP) Neural Network with a standardized information highlight vector that weighted with three concealed layers. It utilizes double cross-entropy as the misfortune work with a erudition pace of 0.01. This model is part of a test-train set and accomplishes an exactness of 94.19%.

Ahmed, Ammara, and D. Maheswari Linen [28] intend to concentrate the absolute most significant churn prediction techniques created over the ongoing years. The essential goal is to churn in telecom industries to precisely evaluate the customer endurance and customer danger capacities to pick up the total information on churn over the customer residency. Another goal is the recognizable proof of the customers who are at the flicker of churn and approximating the time they will churn. This paper centers around breaking down the churn prediction techniques to recognize the churn conduct and approve the explanations behind customer churn. Saini, Ms Nisha, Dr Monika, and KanwalGarg [29] Telecommunication industry gives customers a chance to browse different specialist co-ops. Customer churn can, in this manner, characterized as the exchanging of a customer from administrations of one supplier to another. Since it very well may be an expensive hazard, it should be overseen appropriately. They introduced paper targets foreseeing customer churn utilizing Decision Trees, one of the most broadly utilized characterization techniques dependent on data mining. Mishra, Abinash, and U. Srinivasulu Reddy [30] Customer maintenance in telecommunication is one of the prime issues in customer relationship management (CRM). The essential focal point of CRM is on the existing customer as it is hard to get new customers. The fundamental objective of churn guess is to characterize customers into churners and non-churner. Towards this, deep learning as they are outfitted with enormous expanding data estimates and reveal shrouded design bits of knowledge, distinguishes design, fundamental dangers, and alarm the Telecom Industry about customer conduct with a superior exactness when contrasted with the conventional machine learning techniques. In this paper, Deep learning by Convolutional Neural Network (CNN) actualized for churn prediction, and it demonstrated great execution as far as precision.

III PROPOSED WORK

Churn is a loss of a customer for any industry, and recently telecom industries face these issues. And it is a necessity to find the churned customer in telecom industries. In the research field, most of the research is an ongoing concept to find and analyze the churn. Many approaches, like classification, decision making, and prediction approaches, were used to find churn customers. Prediction pays the main attention to this approach, and this research work applies prediction methods. In our approach, the churn model's prediction viewed as a classification problem; hence, the customer's behavior is to known churner or non-churner using the training dataset previously defined in our constraints. Enduring the prediction phase if the new customer also to be calculated, because the new customer also an important one for improving the revenue, they may be a churn or non-churn, hence it is to be noticed one. In churn, trained data is an essential one. Hence it has to predict by using the trained dataset. The categories here applied is based on the customer performance it categorized to (i) Customer is new, (ii) Customer inactive, (iii) Customer in inactive (churn). An inactive stage always precedes churn customers. Here our main goal is to predict the churned customer by using distance measure and game theory. And pay a way to provide the threshold level from active to inactive.

Dataset

To validate the results, the perfect or correct dataset needed for the implementation. Here dataset was used from publically available data, which provided in [31]. With understanding the context, it is possible to identify the right data sources. An important asset for a company is the valuable data of the customer. Dataset play in many roles, such as providing new data offers, call offers based on their records, and getting new customers by promoting the old customers, so many things may happen with these data. Churner data is an essential one to know that the customer will be churn or not. Some of the datasets shown in Table 1 and Table 2, which are extracted from the cell 2 cell dataset. Here cell 2 cell telecom industries data were used to process our method. The dataset link referred to in [32].

Data processing to filter the accurate data

Churn models expect to distinguish early churn flags and perceive customers with an improved probability to leave voluntarily. For this accurate data is needed, accurate data means clear data of the customer, which contains no repeated data, no invalid data. Data handling are significant strides in the information revelation process that recognize those applicable factors or properties from the enormous number of qualities in a dataset that are excessively pertinent and diminish the computational expense. Data processing includes data cleaning, duplicate removal, extracting features, remove invalid data. From our dataset, it has many extra fields; to reduce the computation cost, we basically taken an important field that is relevant to our research work. That was, we analyzed by ↵→using a data processing phase. Customer churns are those particular customers who have chosen to leave the utilization of administration, item, or even organization and to move to the following rival in the market. So the important fields are given in table 1 and table 2. While processing the data repeated and duplicate data, also we got to extract those data we used data processing the steps is shown below.

Step 1: Start
Step 2: Check all the field from CusID
Step 3: If any field is empty from one CusID then remove entire details of that CusID
Step 4: Check whether repeated CusID Found, IF Yes then remove entire details of that CusID
Step 5: Continue the process until accurate data found
Step 6: End

ISSN: 2233-7857 IJFGCN

Feature/variable extraction

Extraction is one of the important terms of machine learning concepts. If we want important data by searching a term is a type of extraction. Here customer data were extracted by specifying the field of data here; we applied variables/fields like Customer ID, Multiple Lines, Internet Service, Tower problem, Recharge problem, and Customer Care Support. And other fields also included, which shown in table 2. One important factor is feature extraction can stimulus the performance of the predictive model to increase the prediction rates, which is related to high TRUE POSITIVE and low FALSE POSITIVE. Here for the feature extraction, this research work uses the nominal variable. The nominal variable is the type of the variable used to label or categorize, and name the particular attributes measured. It cannot be performed on numbers. In this case, it can categorize the fields like service area, call response services, customer care supports. These parameters can calculate by its variable by applying a nominal variable. For feature extraction here the class label of X is,

$$P\ (X \mid Ci)\ \rightarrow P\ (Cj)$$

Here the x denotes the specific variable, and the P denotes the probability values, and C denotes the categories variables. Hence we got an output of categories variable j of the specific fields.

Handling class imbalance

Class imbalance is a major problem often experienced when dealing with rare events, such as churn recognition in the mobile telecommunications industry. Imbalance data is an important task for the telecom industry.Datasets with imbalanced class distributions are quite common in many real-life applications. For example, in a customer churn prediction system, it is common to find that churning customers are significantly fewer than active customers. The latter are willing to continue using the services of a service provider. In telecom churn prediction, the ratio of churner is very low as compared to non-churners, which lead towards class imbalance problem.

$$Balance\ Data\ Count\ N\ = \frac{Total\ number\ of\ customer\ data}{possible\ No.of\ churn\ data -\ Possible\ No.of\ Nonchurn\ data}$$

From the above equation, it calculates the churn customer data. And N denotes that number of data count.

Prediction using Game theory

Prediction of telecom churners has always been an area of researcher's interest, and hence many researchers have worked in this direction to predict telecom customer churn. Customer churn prediction aims to detect customers with a high tendency to leave a company. One of the crucial key purposes of churn prediction is to discover what reasons expand churn chance. The purpose of churn administration is to keep present customers so long as the corporation is alive. Additionally, classify consumers into churners and non-churner. Measuring the effectiveness of a prediction model depends additionally on how good the outcome will also be interpreted for inferring the feasible reasons for churn. The purpose of prediction is to count on the worth that a random variable will expect at some point or to estimate the likelihood of future routine.

Game Theory is worried about the proper investigation of circumstances called "games," where specialists can pick various methodologies that decide their activities under specific conditions. Conditions and results unfurl through the associations of the specialists' procedures. Game Theory can offer solid scientific confirmation methods to critical thinking approaches with high computational multifaceted nature. This helps to improve the machine learning technique. In this research work, a new approach that integrates machine learning with game theory attempt to discover and extract a churned customer from the telecom dataset. For this extraction game theory method is used. Game Theory tries to analyze the outcomes of "perfect-information." The telecom customer data normally stored in a database. They can sort it all by the terms such as CusID, Income call pack, outgoing call pack, data pack, Director Assisted Calls, Unique

Subs, Active Subs, response to Mail Offers, other country travel, Retention Calls, Retention Offers Accepted, Referrals Made by Subscriber, Adjustments to Credit Rating, Made Call to Retention Team, Credit Rating, Prizm Code, Occupation sorted. Then applied game theory, here the learned knowledge is characterized in forms of rules, which is prediction rules are applied to predict the customer churn possibilities. These results can often be used in support of decision making for prediction improvement.

In-game theory, because of the uncertainty of the opposite team actions, it to be on the safe side, which leads to strategy selection matrix, and then the person will choose to declare. In the same way, both people are balanced or rational, then both should declare it. By this method can predict the customer that would be churn or non-churn. By playing the activity, we can guess the next step by the opponent person; by this manner, by getting a customer activity, plan, recharge, call, data details, we can easily predict the person will be churn or not. If there is perfect knowledge about the customer, then the telecom agent has control of the customer to not churn.This knowledge gathered from records; the following step must be followed to predict the churned customer

Step 1: Observation, Customer Data is detected and gathered; here, the significant data will be taken into account.

Step 2: Analysis, Significant data is put into the extraction phase to get interesting rules to play the game.

Step 3: Choosing Strategy from these rules derives an improved knowledge of the game environment. The technique can create a better selection of the strategies available, i.e., selecting the churned customer.

Step 4: Testing, after choosing a strategy, testing would be made to find out the exact result of that customer surely would be churn or not. Finally, updating would be made in the database.

In this work, the accuracy of the prediction results inspected using a distance measurement method. The effects of distance measurement depending on the attribute types. Therefore, the influence of the distance measurement method depending on the type of attribute needs to considered when calculating the similarity distance between a given churn case and previous churn cases. The effects of the distance measurement methods as these have important roles for predicting the customer. Game theory is employed to determine the overall effect of distance measurement methods.

Algorithm: Game theory and distance measure to predict the churn data

INPUT: *Dataset D, Feature set P, Class C, CusVoice, CusData, and CusSms ∀ Customer.*

OUTPUT: *Possible_Churners_Data PCD, NonChurners_Data NCD.*

PCD = ϴ (unknown churn data)

FOR Each*CusID ==1 To End of the CusID*

Check All Customer Details from table 1 and 2 // Data processing stage

Extract the Details of all the customer // Feature extraction stage

End

W(C): = ***INSERT*** *Data from field 1 through CusID,* **ADD** *Entire details from the specific CusID*

UPDATE *data count as 1 updation completed*

DO *till the Last CusID*

Check Customer Complains (Recharge issue R, Tower issue T, Internet Issue I)

IF *(Complaint = = R)*

Check the previous recharge for the specific customer

Else IF *(Complaint = = T)*

Check with the nearest tower related issues

Else

Check the internet connection

Form this analyze the customer activity //Applying the game theory of machine learning approach of boosting method

Check Previous Activity from 3 month

Measure the activity // distance measure applies here

Calculate $C_n(p)$ *and* $A_n(p)$ *for all* $p \in P$

boosting proceeds in rounds; On_Round t=1…. T // Applying machine learning technique

$C_{sum}(p)=C(p)+PCD$ ∀ p Є P
***For** J= 1 to N do // Last CustomerID*
***Foreach** p Є P do*
Check the PCD status
End For
Select PCD data and store it in as a churned customer balance data store as a Non-Churn customer.
validation_predictions = TRUE
End For
End IF
End IF
END

The proposed work shows the effectivity of the similarity of the distance measurement for predicting the results. This method should improve the prediction accuracy in the preconstruction phase, which will help in the management of prediction performance and the establishment of churn related measures in advance. The outcome of this research can be useful not only to predict the churn but also to predict the non-churn customer.

IV EXPERIMENTAL RESULT

This section shows the implemented results and alsoevaluates the performance of our proposed work. Each model prepared utilizing the preparation data, and their exhibition tested utilizing the test data. Specifically, we thought about the prediction execution concerning the Decision tree and slope boosting model. We have investigated the results of the proposed exact examination and assessed these results through cutting edge assessment measures (i.e., precision, review, f-measure, and accuracy).

The scientifically communicates the precision measure used to assess the right level of prediction intensity of the proposed model. Here the accuracy parameter is calculated by using the below method.

$$Accuracy\ calculation = \frac{(predicted\ data\ -\ measured\ data)}{Total\ data}$$

We expect the test set is new data where the estimation of the class mark acquired from the proposed prescient model. We at that point gathered the predictions result from the prepared classifier on the contributions from the test sets of the Telecom business. As a result, on each iteration, we obtained two predictive sets, which is Churn customer and non-churn customer.

Dataset

Here, data gathered from online accessible Telecom ventures of cell 2 cell dataset for approval. For the telecom industries the data of the customer which contains all data associated to customer's services and contact information like CusID, Income call pack, outgoing call pack, data pack, Director Assisted Calls, Unique Subs, Active Subs, responds to Mail Offers, other country travel, Retention Calls, Retention Offers Accepted, Referrals Made by Subscriber, Adjustments to Credit Rating, Made Call to Retention Team, Credit Rating, Prizm Code, Occupation was sorted, which shown in Table 2 for reference.

Table 1: Parameter of Telecom Customer Dataset

CustomerId	multiple lines	Internet Service	Tower problem	Recharge problem	Customer Care Support
7590-VHVEG	No phone service	No	NA	NA	Yes
5575-GNVDE	No	Yes	No	No	Yes
3668-QPYBK	No	Yes	No	Yes	No

7795-CFOCW	No phone service	No	NA	NA	NA
9237-HQITU	No	No	No	No	Yes
9305-CDSKC	Yes	No	No	No	No
1452-KIOVK	Yes	Yes	No	Yes	Yes
6713-OKOMC	No phone service	No	NA	NA	No
7892-POOKP	Yes	No	No	No	Yes
6388-TABGU	No	Yes	No	Yes	No
9763-GRSKD	No	Yes	Yes	No	Yes
7469-LKBCI	No	No	No	No	Yes
8091-TTVAX	Yes	No	No	No	Yes
0280-XJGEX	Yes	No	No	Yes	Yes
5129-JLPIS	No	Yes	Yes	No	Yes
3655-SNQYZ	Yes	No	No	Yes	Yes
8191-XWSZG	No	No	No	No	No
9959-WOFKT	Yes	No	No	No	No
4190-MFLUW	No	Yes	No	No	Yes
4183-MYFRB	No	No	No	Yes	No
8779-QRDMV	No phone service	No	NA	NA	NA

Discussion

"Prediction" alludes to an algorithm's yield after it has been prepared on a verifiable dataset and applied to new data when determining the probability of a specific result, for example, regardless of whether a customer will churn. Machine learning model predictions permit organizations to make profoundly precise. These furnish the business with experiences that result in substantial business esteem. For instance, if a model predicts a customer is probably going to churn, the business can target them with explicit interchanges and effort that will forestall the loss of that customer. To evaluate the effect of the newly presented features, three sets of experiments were carried out independently in this paper. Issue related things such as tower issue, recharge issue, internet issue. The feature set taken in different types of details like call details, SMS, and data details for the experiments.For the second set of experiments, the new feature set presented in table 2. Hence this research work was used for churn prediction. After the three sets of experiments completed, we obtained the prediction results in terms of the prediction rates – false churn rate (FP) and true churn rate (TP). Based on these pairs of FP and TP, the AUC plotted to measure the result with Distance measure.

ISSN: 2233-7857 IJFGCN

Evaluation measure

The AUC values by utilizing the full set of the new highlights are most noteworthy among the element subsets. This shows the full set is the best element for the churn prediction. To all the more vitally assess the impact of capabilities and modelling strategies, the AUC method utilized. In the methodology of the choice tree, the churn prediction of FP is extremely high. It gives wrong results. A choice framework is any information system of the form

S = (U, C ∪ {d}), where C is a conditional attribute, and d ∉ C is the decision attribute. To assess the classification results, we count the number of True Positives (TP), True Negative (TN), False Positive (FP) and False Negative (FN). The FN value belongs to Positive P (e.g. TP + FN = P) but wrongly classified as Negative N (e.g. TN + FP = N) while FP value part of N but wrongly classified as P.

Table 2: Dataset for Identifying Churn Customer

Customer ID	Director Assisted Calls	Unique Subs	Active Subs	Responds To Mail Offers	other country travel	Retention Calls	Retention Offers Accepted	Referrals Made By Subscriber	Adjustments To Credit Rating	Made Call To Retention Team	Credit Rating	Prizm Code	Occupation
3000470	24.01	3	3	No	No	0	yes	2	0	No	3-Good	Town	Other
3000482	0.25	1	1	No	No	0	no	0	0	No	4-Medium	Other	Other
3000486	1.49	2	1	Yes	No	0	no	3	7	No	1-Highest	Suburban	Other
3000498	0	1	1	No	No	0	no	2	3	No	2-High	Rural	Other
3000542	0	1	1	Yes	No	0	no	0	0	No	2-High	Other	Professional
3000546	3.71	1	1	No	No	0	no	0	0	No	1-Highest	Suburban	Professional
3000566	3.96	3	3	No	No	0	yes	1	0	No	1-Highest	Other	Other
3000594	4.21	1	1	No	No	0	yes	0	3	No	1-Highest	Town	Other

3000606	0.25	1	1	No	No	0	yes	0	0	No	1-Highest	Town	Other
3000610	4.21	3	3	No	No	0	yes	1	0	No	1-Highest	Other	Other
3000630	0	2	2	Yes	Yes	0	no	0	0	No	1-Highest	Town	Retired
3000662	0.25	1	1	Yes	No	0	yes	0	0	No	3-Good	Suburban	Clerical
3000666	0	1	1	Yes	No	0	no	2	0	No	1-Highest	Other	Other
3000686	0	2	2	No	No	0	yes	0	0	No	1-Highest	Town	Other
3000698	0.25	1	1	Yes	No	0	yes	2	0	No	6-Very Low	Suburban	Professional
3000702	4.95	2	1	No	No	0	yes	0	1	No	3-Good	Rural	Other
3000746	0	1	1	Yes	No	0	no	0	0	No	1-Highest	Suburban	Clerical
3000754	0.25	3	2	No	No	0	no	1	0	No	2-High	Rural	Other
3000758	0	2	2	No	No	0	no	0	0	No	1-Highest	Town	Other
3000798	2.48	2	2	No	No	0	no	2	2	No	3-Good	Other	Other
3000846	6.43	2	1	No	No	0	no	0	2	No	3-Good	Suburban	Other
3000862	2.97	2	2	Yes	No	0	no	0	0	No	1-Highest	Town	Professional

3000866	0	1	1	No	No	0	no	0	0	No	6-Very Low	Suburban	Other

Customer Churn is a key challenge in competitive markets and highly observed in the telecommunication section.It is important to clarify that each instance in the training and validation data may refer to different customers. Applying our approach obtained high accuracy in predicting the data (i.e., customer churn and non-churn with high certainty). As we aim to predict the customer churn correctly; hence we need the sensitivity to be high.

Sensitivity=True Positive/Positive, where Positive=True.

The model was prepared and tested through the Area Under Curve (AUC)**.** There are two datasets used in this work.The first dataset consists of 1000 samples and 6 attributes, while the second dataset contains 1500 samples and 14 attributes. Dataset's details are as shown in Tables 1 and 2.Each customer details stored in a database. It contains who disconnect our service and who own our services, every detail will be stored in the database so that we can predict the revenue of the telecom services. Here we use the cell 2 cell dataset.It is the 6th largest wireless company in the US. Some of the extracted data shown in Tables 1 & 2, shown above.

Table 3: Certainty Calculus

Certainty Calculus	**Certainty churn**	**Certainty non-churn**
Distance Measure	45	60
Gaming Theory	34	45
Similarity measure	67	79

Here the certainty estimation using distance measure is applied. Certainty used to both churn and non-churn with high and low certainty.

Table 4: Types of complaints

Complaints	**Coverage**	**Recharge issues**	**less offers**
Complaints in %	80%	45%	77%

Table 4 shows the type of complaint given by the customer.

Table 5: Correlation analysis

Max-Cosine-Sim	**Min-Cosine-Sim**
56	40

Here Cosine is a similarity with other operators' customers

Table 6: Performance measure with AUC

Co-relation analysis	**Churn**	**Non-churn**
Distance Measure	56%	88%
Gaming Theory	45%	78%
Similarity measure	32%	45%

Table 7: Prediction calculus

Prediction calculus	Distance Measure & Game theory	Decision tree	Gradient Boosting Model
	95%	76%	82%

Table 8: True positive

True positive calculus	Distance Measure & Gaming theory	Decision tree	Gradient Boosting Model
	0.9	0.4	0.6

Table 9: False positive

False-positive calculus	Distance Measure & Gaming theory	Decision tree	Gradient Boosting Model
	0.1	0.6	0.8

Table 10: Accuracy Calculus

Method	Recall
Distance Measure & Gaming theory	92.426
Decision tree	72.644
Gradient Boosting Model	84.780

Table 11: Precision Calculus

Method	Precision
Distance Measure & Gaming theory	89.842
Decision tree	64.822
Gradient Boosting Model	74.980

Table 12: Recall Calculus

Method	Precision
Distance Measure & Gaming theory	89.842
Decision tree	64.822
Gradient Boosting Model	74.980

Table 13: F-Measure Calculus

Method	Precision
Distance Measure & Gaming theory	94.842
Decision tree	74.982
Gradient Boosting Model	84.684

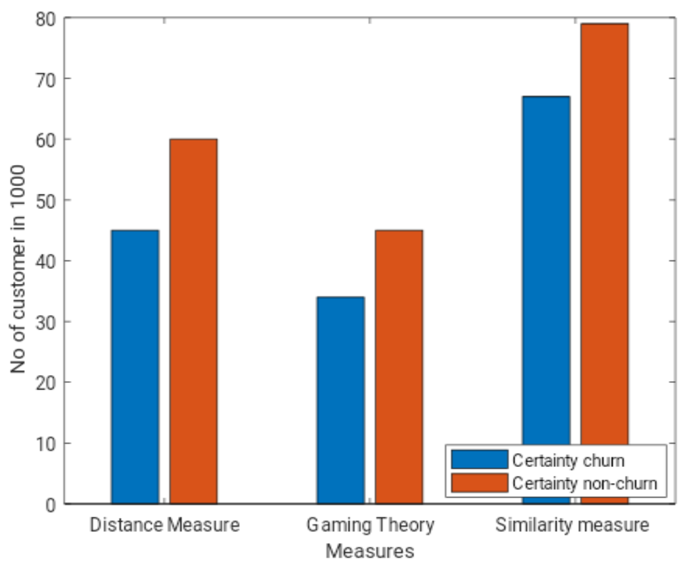


Chart 1: Certainty Calculus

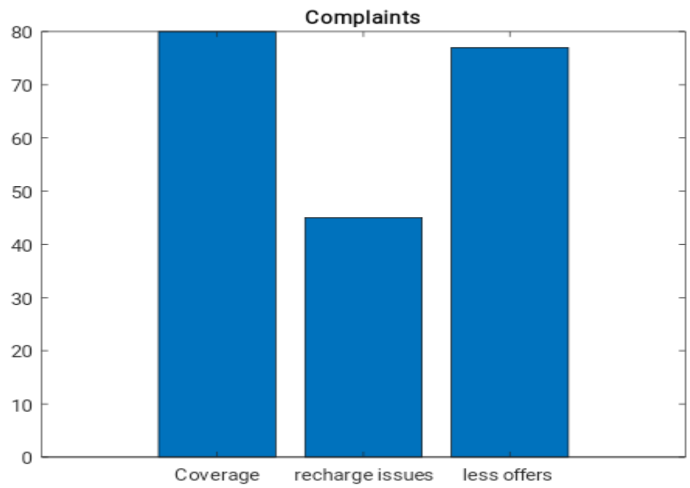


Chart 2: Types of Complaints

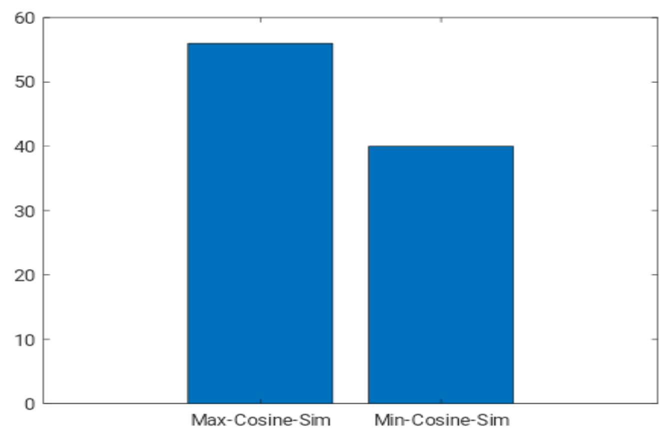


Chart 3: Co-relation analysis

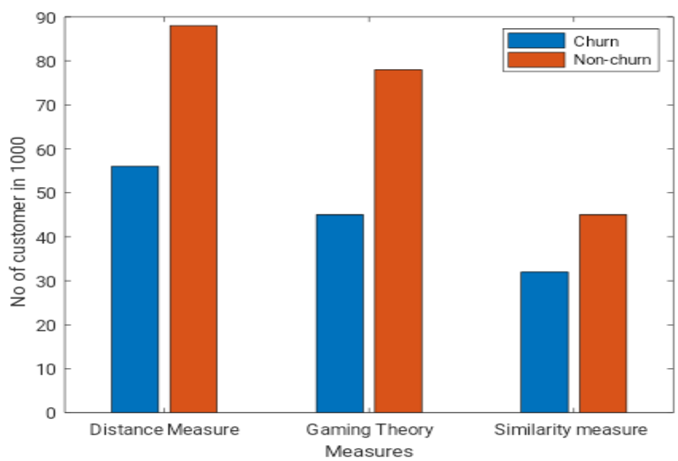


Chart 4: Performance measure with AUC

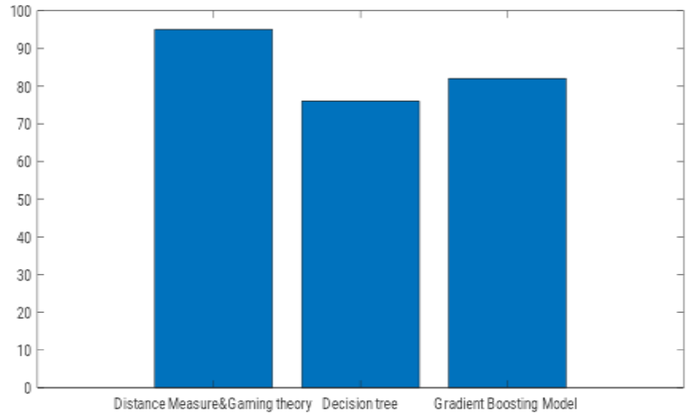


Chart 5: Prediction calculus

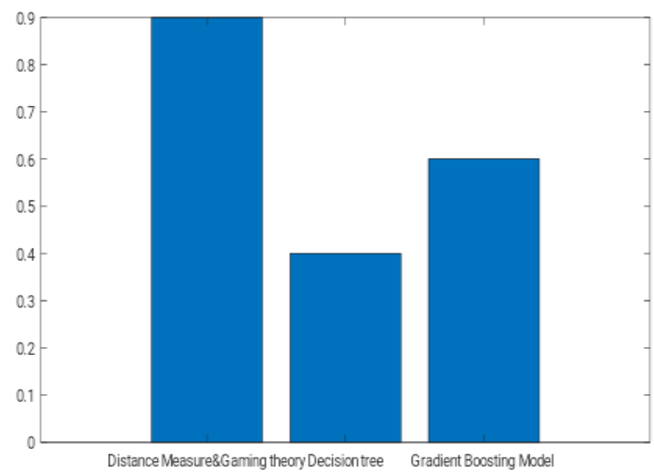
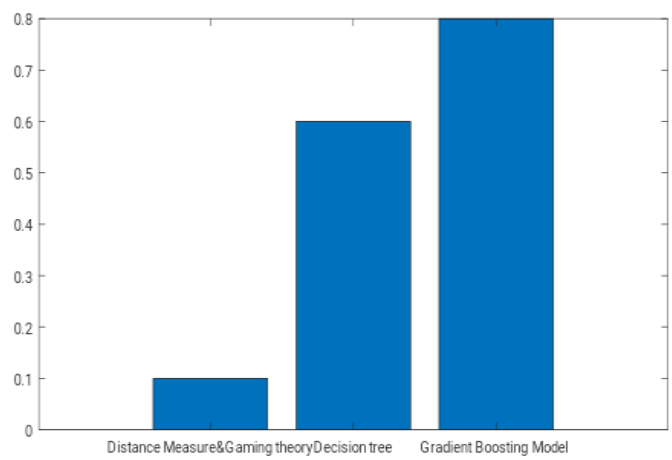


Chart 6: True positive



ISSN: 2233-7857 IJFGCN

Chart 7: False positive

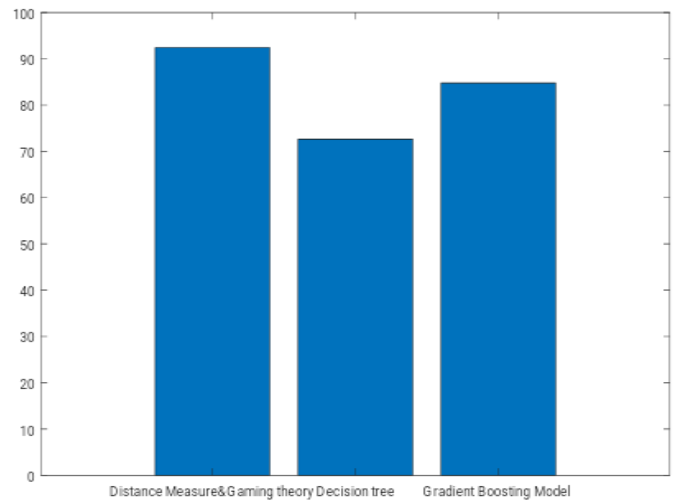


Chart 8: Accuracy

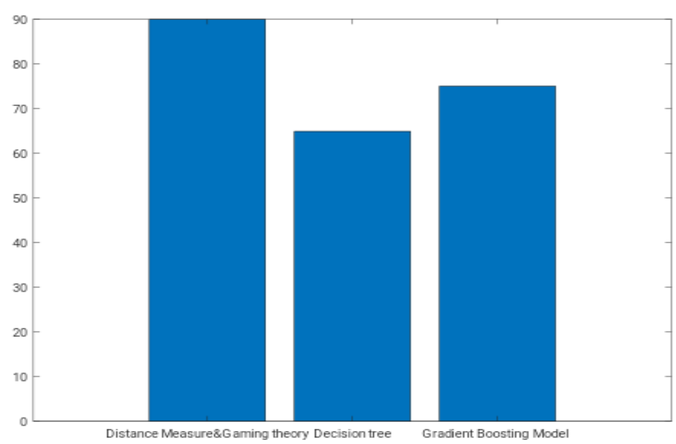


Chart 9: Precision Calculus

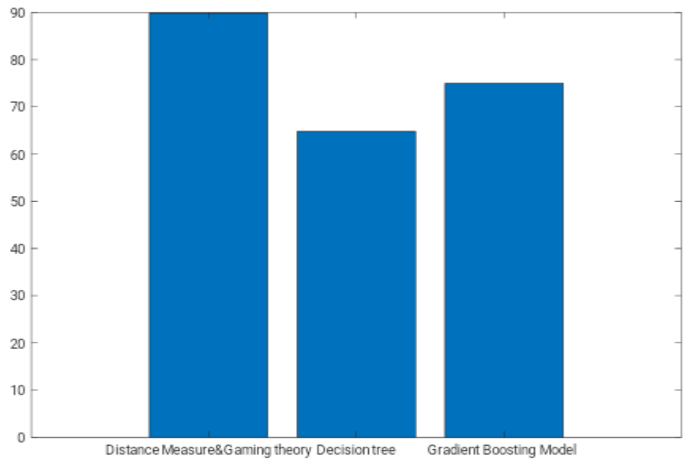


Chart 10: Recall Calculus

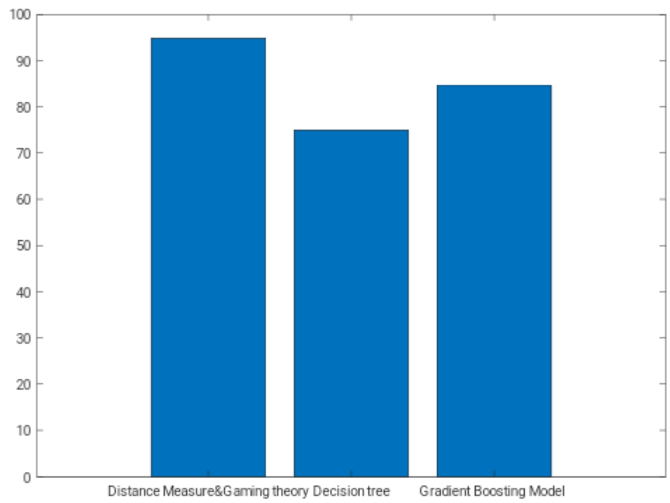


Chart 11: F-Measure Calculus

We have removed the accuracy of the classifiers, True Positive and False Positive rates, F-measure, MCC score, Area Under Curve (AUC) area, and Precision-Recall (PRC) area. These measures helped us know which algorithm is progressively productive and gave us bits of knowledge about the differing execution. It saw that, in the greater part of the cases, the classifiers, when joined with both of the outfit strategies, yield better results. The exploratory results uncover that the accuracy of the prediction improves when combined with boosting method of machine learning technique.

V. CONCLUSION

Customer churn prediction is an important practice in rapidly growing and competitive telecom industries. Customer churn prediction has emerged as an essential part of a considered decision-provided approach for telecom services due to the high cost associated with acquiring new customers.In our approach, prediction of churn model is viewed as a classification problem, hence the behavior of the customer is to known churner or non-churner by using the training dataset of previously defined in our constraints. In this work, the Distance measure is performed with the gaming theory. Our approach works about predicting churn in telecommunication networks use Distance Measure and Gaming Theory. It is to identified already churned customers and analyzed a graph model to infer interactions with the current customers to predict new churners based on these interactions. Our method provides an accuracy of 92.426% for predicting the customer who will be churn.

REFERENCES

[1]. Mozer, M.C., Wolniewicz, R., Grimes, D.B., Johnson, E., Kaushansky, H.: Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry. IEEE Trans. Neural Netw. 11, 690–696 (2000)

[2]. Huang, B., Kechadi, M.T., and Buckley, B., 2012. Customer churn prediction in telecommunications. Expert Systems with Applications, 39(1), pp.1414-1425.

[3]. Amin, A., Shehzad, S., Khan, C., Ali, I., Anwar, S.: Churn Prediction in Telecommunication Industry Using Rough Set Approach. In: Camacho, D., Kim, S.-W., Trawiński, B. (eds.) New Research in Multimedia and Internet Systems. SCI, vol. 572, pp. 83–96. Springer, Heidelberg (2015)

[4]. Wangperawong, Artit, CyrilleBrun, Olav Laudy, and RujikornPavasuthipaisit. "Churn analysis using deep convolutional neural networks and autoencoders." arXiv preprint arXiv:1604.05377 (2016).

[5]. L. Ning, L. Hua, L. Jie, Z. Guangquan, "A customer churn prediction model in the telecom industry using boosting," IEEE Trans. Ind. Inform. 10 (2014) 1659– 1665.

[6]. T. Verbraken, W. Verbeke, B. Baesens, "A novel profit is maximizing metric for measuring classification performance of customer churn prediction models," IEEE Transaction on Knowledge and Data Engineering 25 (2013) 961–973.

[7]. Ning Lu, Hua Lin, Jie Lu, Guangquan Zhang "A Customer Churn Prediction Model in Telecom Industry Using Boosting," IEEE Transactions on Industrial Informatics, vol. 10, no. 2 May, 2014.

[8]. Qureshi, S.A., Rehman, A.S., Qamar, A.M., Kamal, A., Rehman, A.: Telecommunication subscribers' churn prediction model using machine learning. In: Eighth International Conference on Digital Information Management (ICDIM 2013), pp. 131–136. IEEE (2013)

[9]. Wang, C., Li, R., Wang, P., & Chen, Z. (2017, July). Partition cost-sensitive CART based on customer value for Telecom customer churn prediction. In Control Conference (CCC), 2017 36th Chinese (pp. 5680-5684). IEEE. July 2017

[10]. Keaveney, S.M.: Customer switching behavior in service industries: An exploratory study. J. Mark. 59, 71–82

[11]. K Coussement and D Van den Poel, "Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques," Expert systems with applications, Vol. 34, No. 1, 2008, pp. 313-327.

[12]. W Verbeke, D Martens, C Mues, and B Baesens, "Building comprehensible customer churn prediction models with advanced rule induction techniques," Expert Systems with Applications, Vol. 38, No. 3, pp. 2354-2364.

[13]. KW De Bock and D Van den Poel, "An empirical evaluation of rotation-based ensemble classifiers for customer churn prediction," Expert Systems with Applications, Vol. 38, No. 10, 2011, pp. 12293-12301.

[14]. S Maldonado and C Montecinos, "Robust classification of imbalanced data using one-class and two-class SVM-based multi classifiers", Intelligent Data Analysis, Vol. 18, No. 1, 2014, pp. 95-112.

[15]. G-E Xia and W-D Jin, "Model of customer churn prediction on support vector machine," Systems Engineering-Theory & Practice, Vol. 28, No. 1, 2008, pp. 71-77.

[16]. Amin, Adnan, SaeedShehzad, Changez Khan, Imtiaz Ali, and Sajid Anwar. "Churn prediction in telecommunication industry using the rough set approach." In New trends in collective computational intelligence, pp. 83-95. Springer, Cham, 2015.

[17]. Praveen Asthana, "A comparison of machine learning techniques for customer churn prediction." Simulation Modelling Practice and Theory 55 (2015): 1-9.

[18]. Prashanth, R., K. Deepak, and Amit Kumar Meher. "High accuracy predictive modeling for customer churn prediction in the telecom industry." In International Conference on Machine Learning and Data Mining in Pattern Recognition, pp. 391-402. Springer, Cham, 2017.

[19]. Lu, Ning, Hua Lin, Jie Lu, and Guangquan Zhang. "A customer churn prediction model in the telecom industry using boosting." IEEE Transactions on Industrial Informatics 10, no. 2 (2012): 1659-1665.

[20]. Bi, Wenjie, MeiliCai, Mengqi Liu, and Guo Li. "A big data clustering algorithm for mitigating the risk of customer churn." IEEE Transactions on Industrial Informatics 12, no. 3 (2016): 1270-1281.

[21]. Gaur, A., and Dubey, R., 2018, December. Predicting Customer Churn Prediction In Telecom Sector Using Various Machine Learning Techniques. In 2018 International Conference on Advanced Computation and Telecommunication (ICACAT) (pp. 1-5). IEEE.

[22]. Ahmad, AbdelrahimKasem, AssefJafar, and KadanAljoumaa. "Customer churn prediction in telecom using machine learning in big data platform." Journal of Big Data 6, no. 1 (2019): 28.

[23]. Kim, Kyungtae, and Jee-Hyong Lee. "Bayesian Optimization of Customer Churn Predictive Model." In 2018 Joint 10th International Conference on Soft Computing and Intelligent Systems (SCIS) and 19th International Symposium on Advanced Intelligent Systems (ISIS), pp. 85-88. IEEE, 2018.

[24]. Ullah, Irfan, BasitRaza, Ahmad Kamran Malik, Muhammad Imran, SaifUl Islam, and Sung Won Kim. "A churn prediction model using random forest: analysis of machine learning techniques for churn prediction and factor identification in the telecom sector." IEEE Access 7 (2019): 60134-60149.

[25]. Dulhare, Uma N., and IfrahGhori. "An efficient hybrid clustering to predict the risk of customer churn." In 2018 2nd International Conference on Inventive Systems and Control (ICISC), pp. 673-677. IEEE, 2018.

[26]. Mitkees, Ibrahim MM, Sherif M. Badr, and Ahmed Ibrahim BahgatElSeddawy. "Customer churn prediction model using data mining techniques." In 2017 13th International Computer Engineering Conference (ICENCO), pp. 262-268. IEEE, 2017.

[27]. ShrishaBharadwaj ; B.S. Anil ; AbhirajPahargarh ; AdhirajPahargarh, "Customer Churn Prediction in Mobile Networks using Logistic Regression and Multilayer Perceptron(MLP)", IEEE, 2018.

[28]. Ahmed, Ammara, and D. Maheswari Linen. "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, 2017.

[29]. Saini, Ms Nisha, Dr Monika, and KanwalGarg. "Churn prediction in telecommunication industry using decision tree." Streamed Info-Ocean 1, no. 1 (2016).

[30]. Mishra, Abinash, and U. Srinivasulu Reddy. "A novel approach for churn prediction using deep learning." In 2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), pp. 1-4. IEEE, 2017.

[31]. https://www.kaggle.com/jpacse/datasets-for-churn-telecom

[32]. https://www.kaggle.com/kkairu/telecom-churn-data

[33]. Suznjevic, M., Stupar, I., Matijasevic, M.: MMORPG Player Behavior Model based on Player Action Categories. IEEE (2011)

[34]. Kwon, Nahyun, Joosung Lee, Moonsun Park, Inseok Yoon, and YonghanAhn. "Performance evaluation of distance measurement methods for construction noise prediction using case-based reasoning." *Sustainability* 11, no. 3 (2019): 871.

[35]. Koo, C.; Hong, T.; Hyun, C.; Koo, K. A CBR-based hybrid model for predicting a construction duration and cost based on project characteristics in multi-family housing projects. Can. J. Civ. Eng. 2010, 37, 739–752.

ISSN: 2233-7857 IJFGCN