Indonesian Journal of Electrical Engineering and Informatics (IJEEI)
Vol. 10, No. 2, June 2022, pp. 431~440
ISSN: 2089-3272, DOI: 10.52549/ijeei.v10i2.2985

Customer Churn Prediction in Telecommunication Industry Using Classification and Regression Trees and Artificial Neural Network Algorithms

Sulaiman Olaniyi Abdulsalam[1], Micheal Olaolu Arowolo[2,3] Yakub Kayode Saheed[4], Jesutofunmi Onaope Afolayan[5]

[1]Department of Computer Science, Kwara State University, Malete, Nigeria.
[2,5]Department of Computer Science, Landmark University, Omuaran, Nigeria.
[3]Landmark University SDG 9 (Industry, Innovation and Infrastructure).
[4]School of Information Technology & Computing, American University of Nigeria.

Article Info

Article history:

Received Jan 25, 2021
Revised Mar 2, 2022
Accepted Mar 11, 2022

Keyword:

Telecoms
Relief-F
ANN
CART
Churn

ABSTRACT

Customer churn is a serious problem, which is a critical issue encountered by large businesses and organizations. Due to the direct impact on the company's revenues, particularly in sectors such as the telecommunications as well as the banking, companies are working to promote ways to identify the churn of prospective consumers. Hence it is vital to investigate issues that influence customer churn to yield appropriate measures to diminish churn. The major objective of this work is to advance a model of churn prediction that helps telecom operatives to envisage clients that are most probable to be subjected to churn. The experimental approach for this study uses the machine learning procedures on the telecom churn dataset, using an improved Relief-F feature selection algorithm to pick related features from the huge dataset. To quantify the model's performance, the result of classification uses CART and ANN, the accuracy shows that ANN has a high predictive capacity of 93.88% compared to the 91.60% CART classifier.

Corresponding Author:

Micheal Olaolu Arowolo
Department of Computer Science,
Landmark University, Omuaran Nigeria.
Email: arowolo.olaolu@gmail.com

1. INTRODUCTION

Data-driven sectors have been able to carry out analysis of data and fetch out extensive knowledge through technology advances. Methods of data mining has helped in achieving the prediction of certain future customer behaviors [1]. Customer churn, is categorized as customer attrition, it is amongst the most critical issues that reduces a company's profit. The procedures of business intelligence for locating customers who wants to change from a company to its other competitor can be described as customer churn [2]. The telecoms industry is a highly technological industry that has grown enormously in the past, as a consequence of the emergence and commercial success of both mobile telecommunications, these two decades [3-4].

For many telecoms' firms, customer churn or customer attrition is a major problem, it occurs when a customer terminates his subscription and moves to another rival. There are several variables that impact the decision of the client to turn to another rival. In general, these variables were related to the high cost, bad jobs, fraud and privacy issues related to customer service[5]. Customer turnover causes significant loss of profit when those thresholds are surpassed. Companies know that gaining fresh clients can be costly than retaining old ones [6].

In several sectors, such as telecom providers, credit cards, Internet service providers, e-commerce, newspaper publishing firms, banking sectors, among others, Consumer Churn Prediction (CCP) has been raised as a key concern in telecommunication firms [7].

In recent years, Consumer Churn Prediction has become an increasingly common research problem and therefore, telecom suppliers have commonly used strategies to classify potential churn customers based on their historical records, previous behaviors and offering some services to convince them to live[8]. Long-term customers, on the other hand, are more lucrative for service providers because they are more focused on purchasing additional goods and spreading the satisfaction of the customer within their radius, thereby attracting more and more customers indirectly [9].

Businesses must have a thorough knowledge of why churn emerges to retain their clients. There are certain factors to be discussed, such as organization discontent, certain businesses' competitive costs, customer migration, the necessity for better services for clients that can encourage users switching to their present service provider and moving to a different one [4]. Companies however, understand that winning new customers is a great deal More costly than current ones being retained [10].

In general, churn prediction obtained data are imbalanced, instances in non-churner customer may outstrip churners class instances. Typical classification techniques seem to achieve relevant accuracy results for huge classes and miss smaller ones, this is regarded one of the most challenging and significant issues. Different methods have been suggested to handle the issue of imbalanced churn prediction data. These techniques comprises of sufficiently common evaluation metrices, use of cost-sensitive learning, modification of training set distributions by method sampling, the use of approaches to reduce dimensionality, among others [11][12].

Reducing dimensionality is of essence in data mining, it is motivated by developing feature dimensionality in specified concerns and increasing interest in innovative yet costly computational methodologies capable of modeling complex associations. Feature Selection is one of the methods of preprocessing to classify the data sub-set from large-dimensional data. In particular, feature selection techniques such as Relief-F, Genetic Algorithm, among others, are computationally efficient, but responsive to complex association patterns, such as associations, so that prior to downstream modeling, informative features are not mistakenly excluded [13]. Relief-F-based algorithms, a distinct category of filter-based feature selection algorithms which have received attention by achieving an efficient balance among these goals optimally adapting to different data features [14][15][16].

In attempt to fine-tune developed models, recent investigations for churn analysis have proposed and suggested that methods such as SVM, ANN, CART, among others classification methods are the most commonly utilized. Numerous optimization techniques and strategies have been explored and recommended to have been identified to operate the best and make studies in several fields such as telecommunications companies, banking, business and insurance, among others improved in productivity of these sectors [17][18].

In this study, the major contribution is adopting an enhanced Relief-F feature selection algorithm, is created with an innovative learning method by using the subsets of the relevant churn prediction method data based on CART and ANN classifiers using optimal predictors that increase the predictive output of variables. Using diverse evaluation metrics, and related to traditional prediction methods as well as other relevant processes presented in the literature, the presented methodology will be evaluated.

2. LITERATURE REVIEW

Classification of churners and non-churners are considered to be a predominant problem for telecom providers, it is characterized as losing clients as they flee for contenders. To be able to pre-classify customer churn offers the telecom business an appreciated insight into retaining its customer base. In recent years, large ranges of churn classification methods have been explored. Most creative models uses advanced machine learning classifiers and have found that the roots of customer churn are reviewed in relation to service quality, customer satisfaction/dissatisfaction and economic value variables.

Prediction investigation utilizing Random Forest with discriminating features method of analysis for prediction of churners in the telecommunications industry was proposed [3], grounded on developmental search Random Forest, they predicted churners and non-churners in the telecommunication sector that use discriminant feature investigation as innovation postponement of the traditional Random Forest to learn tilted Developmental Detection tree. The suggested approach controls the benefit of two methods of discriminant investigation to measure the project index used in PPtree construction. they used Support Vector Machines with Linear Discriminant Analysis to obtain linear division of variables and developed specific classifiers that are stronger and more flexible than traditional Random Forest in oblique PPtree development. The detection techniques are proven to outperform in terms of Accuracy. The prediction model, PPForest based on LDA delivers efficient evaluators.

A comparative investigation of customer churn prediction by means of Negative Correlation Learning have been suggested [4], by utilizing an ensemble based Multilayer Perceptrons. Training for predicting consumer churn in a telecom market is gained using negative correlation learning. The test findings suggested that the NCL-MLP-ensemble can improve overall classification efficiency (high churn rate) compared to the non-NCL-MLP-ensemble and other traditional data mining strategies utilized in churn study.

Investigating synthetic and ensemble approaches in the telecommunications industry for users churn prediction system was suggested [19], by presenting a detailed study on churn-based machine learning prediction of the telecommunications industry for 8 years. The issues and problems in telecommunication churn were were observed with eliminated predictions and the suggestions and solutions published. The study and overview enable researchers or data experts in telecommunication fields to fetch optimum and suitable techniques with design methods to improve novel models for future churn prediction.

A findings of consumers' purchase decision-making using churn prediction support framework was carried out [20], their investigation showed a wide range of characteristics used to generate consumer churn model by many scholars. It demonstrates specific methodologies used in churn prediction up to date. Methods of modeling, for instance; Neural Network, Logistic Regression, Decision Tree, Random Forest, Support Vector Machine among other methods are employed for churn discovery. The discoveries indicate that through predictive analytics, the customer churn forecast can achieve more precise results as compared with other related prediction approaches. In Customer churn prediction and customer retention by predictive analytics, there is a wide scope of research.

Churn analysis for telecommunications Sector utilizing Decision Tree was proposed [21]. Decision Tree classification method used a large client dataset evaluated for churn. After the implementation of all possible decision tree variants in SPSS, it was noted that Exhaustive CHAID method demonstrated to be further consistent and reliable to envisage the likely customers to churn rather than impending ones.

Telecom consumer analysis churn prediction with machine learning in large amount of data was proposed [22], they constructed a new approach to develop and choose features. The Area Under Curve sole criterion is utilized to determine the model's efficiency, and the obtained AUC value is 93%. The use of customer communal networks in the prediction model by mining features of social network study is another key contribution. The use of system network study increased the model's output against the AUC benchmark from 84 to 93%. Via Spark setting, the model was systematized and verified by employing a huge dataset generated by transforming large raw information obtained from SyriaTel telecommunication corporation. The database comprises certain customer data across a span of nine months which was used by SyriaTel to train, test and assess the classification. Four algorithms were experimented with in the methodology: Decision Tree, Random Forest, Gradient Boost Tree and Extreme Gradient Boost. The better outcomes, therefore, are obtained by implementing the XGBOOST algorithm. In this churn-predictive framework, this algorithm was used for classification.

An innovative attribute selection approach and framework telecommunications churn evaluation correlation was proposed [23], the information used for the analysis contains information of actual customer phone metadata collected from a large telecom industry in Turkey for the years 2013 and 2014. In addition to an overall attribute selection procedure, for the development of five datasets, four various methodologies called R-correlation coefficient-based feature selection, ¥-correlation coefficient-based feature selection, Relief-F, and Gain Ratio were utilised. Four classifier algorithms were subsequently implemented, with Random Forest, Decision Tree, Naive Bayes and AdaBoost. Evaluation criteria consisting Accuracy, Sensitivity, Specificity, F-score, and run-time were used to evaluate the results obtained. The findings of the correlations indicate that on consumer churn estimation, the projected feature selection algorithm outdoes the state-of-the-art approaches.

Telecom market consumer churn prediction analysis using the CART algorithm was proposed [24], The estimation of customer attrition in the telecommunications industry has been the most relevant subject for research in recent years. Since it helps detect which client is likely to change or cancel their service subscription. Review of data collected from telecommunications providers can help find the consumer churn reasons and also use the information to attract customers. Thus, for telecommunication companies to maintain their customers, predicting churn is extremely important. This study developed the call tree classification model, evaluated the output indicators, and compared its performance with the model of logistic regression.

3. RESEARCH METHOD

The goal of the proposed study is to construct a classification model to indicate that the customer in Telecom datasets is a likely churner or non-churner. By implementing the key retention policies that are likely to retain and attract consumers who have the most propensity to churner and pursue them to stay, this procedure would aid customer relationship management. The feedback for suggesting the customer churn prediction model includes information for each mobile subscriber from past calls, along with all the person and business information held by the provider of telecom services. Fully trained with the training dataset after the prediction model, the test dataset and the model have to be able to predict churners. Figure 1 shows the technique for the prediction of churners and the description of the steps proposed.

Figure 1. Customer churn prediction Approach using Relief-F with CART and ANN models

Machine learning is a method of understanding strategies from big data to find useful knowledge. To obtain and analyze beneficial information from various huge datasets, it uses analytical tools, arithmetic, artificial intelligence, and data science, it presents it for advanced, valuable knowledge and information. Machine learning can solve problems relating to data learning theory of classification, regression, clustering, and correlation depending on the intent of research. The pattern of data is In this method, descriptively and intelligently presented.

3.1. Datasets

Telecom datasets produced by Telecom Industry operators collected from the Francisco gallery of bigml.com are the realistic part of this analysis, it comprises of 20 attributes and 3333 instances. A dataset pertaining to functionality and use of telephony account features and whether or not the customer has churned [25]. The main characteristics of the dataset attributes comprises of; name, account length, zone code, global plan, voicemail, number vmail messages, entire day minutes, entire day calls, entire day charge, entire eve minutes, churn, among others [25].

3.2. Feature Selection based on enhanced Relief-F

Identifying attributes that are definitely applicable to the target variable is the most critical step in data pre-processing. Not all features, nevertheless, are well-contributed to the classifier learner model. The feature selection method became important to improve efficiency and make the customer churn prediction model easier to interpret, minimize overfitting, remove variables that are redundant and do not provide any information or contribution to the model's production because of the wide-scale datasets in telecom provider services. In addition, it decreases the size of the prediction problem and allows classification algorithms to generate results as quickly as possible [3].

Instance-based learning inspired the initial Relief algorithm. Relief computes a deputation statistic for respective feature that can be utilized to approximate the eminence or significance of the feature to the target definition as a distinct assessment filtering feature selection process. Such feature information is represented as feature weights (feature weight 'A' = W[A]), or informally as 'scores' features that can vary from -1 (worst) to +1 (best). Particularly, the unique Relief algorithm was restricted to problems with binary classification and had no method for handling misplaced data. Approaches to extend relief to issues with multi-class or continuous endpoints are required [13]. The Pseudocode for the traditional Relief-F Algorithm emphasized the sequences for training selected instances with no substitutes with user-defined parameters [13]. Relief-F have proven to be the best known variant and most utilized, it relies on number of neighbors, increasing weight estimate reliability and noisy problems, it can handle missing data values, and handle multi-class endpoints [26].

Algorithm1: Pseudocode for the enhanced Relief-F Algorithm
***n* = numbers of the trained instances**
***a* = feature numbers (attributes)**
m* = parameter numbers random training instances out of n used to update *W
c = constant

set all feature weights *W[A]* = 0.0
for i: = 1 to *m* do
arbitrarily pick a ‘target’ instance R_i
discover an adjoining hit ‘*H*’ and adjoining miss ‘*M*’ (instances)
for A: = 1 to a do

$W[A] = W[A]\ d_i ff\ (A, R_i, H)/m + d_i ff\ (A, R_i, M)/m$
$W[A] = W[A]\sum_{j=1}^{k} di \frac{ff(A,ri,hj)}{m.c} + O\ (n^2.a)\ (m+c_3{}^2c.n.\ a)$
end for
end for
return vector *W* of feature scores that estimate the feature values

The Relief computes the ratings of features based on modifications in feature and class values among neighbor instances. If a set of neighbor instances has positive variations for a feature but the same class value, then ReliefF reduces the score of that feature. Additionally, ReliefF improves the score of the function if adjacent instances have positive variations for a feature and different class values. For a set of experimented instances and their nearest neighbours, this is repeated to determine an average score for each characteristic [16][27]. In this study an enhanced Relief-F for fetching the missed fits and best fits is suggested to fetch relevant information from the churned telecom dataset. The results of the relief-f fetch a relevant subset of the data and it is used as a reduced preprocessed data for classification.

3.3. Classification based on CART (Classification and Regression Tree)

For the continuous dependent variable and categorical predictor variable, the CART approach is more appropriate. CART recursively divides the function space into non-overlapping areas. A classification tree is generated to predict the value of a dependent categorical variable. Incorporates CART to determine the goodness of fit more reliably, including checking with a reference data set and cross-validation. In distinct sections of the tree, CART can use the same variables more than once. This ability can reveal complex interdependencies of variables between sets. CART may be used to pick the input set of variables in combination with other prediction techniques [21].

Pruning is performed after the CART algorithm is qualified. As the basis for pruning, the total error rate is used. The smallest tree provides the most effective classification (trees with least number of layers). For a target variable with constant and definite data, the CART algorithm is applicable. If continuous data is represented by the goal variable, then the regression tree can be used. A classification tree may be used if the target variable includes unconditional data [28]. An active threshold value is calculated as a state for each node in the CART algorithm. At each node, a sole input variable function splits the data and constructs a binary DT. To estimate the metrics, the Gini index is used. The presence of several groups in the data is demonstrated by a high level of dispersed indicators. By comparison, the presence of a single group is suggested by a low level of indicators.

3.4. Classification based on ANN

The Neural Networks Model is used to build features such as non-linear features. Due to its comparable data processing system, the model retains the capacity to learn. After applying several concerns, such as grouping, these methods provide good results. Due to its possible range estimation, the model is dissimilar to the classification model and decision tree. Has the neural network multiple approaches of merits and demerits. The investigator claims that the deep neural network is stronger than the churn prediction model of decision tree and regression analysis [29].

Neural Networks is a methodology for datamining that has the capacity to learn from mistakes. The brain stimulates Neural Networks. This happens in the sense that a few new things are learned by the brain, which will then be transmitted through neurons. The neural network neuron may also learn from training data with learning algorithms; this makes them denoted to as Artificial Neural Neurons [30].

It is possible to distinguish neural networks into single-layer perception and multilayer perception (MLP) networks. The perception of multiple layers comprises of multiple layers of plain, two-state, sigmoid transfer mechanism with processing element or neurons that communicate using weighted links. In reality, the neural network involves one or more intermediaries' secret layers of neurons in between the layers of input and output.

These intermediate layers are known as hidden layers, and nodes embedded in these layers are known as hidden nodes because they do not take inputs directly from outside [31].

3.5. Evaluation Criteria and Experimantal Setup

In this study, modifications of analytical data have been implemented in MATLAB data mining tool [32][33]. In order to find a definitive decision, effective functional set-up and use of study variables and effective performance metrics are important. To determine the efficiency of the churn prediction model, the telecom industry represents various methods of performance measures [34].

Accuracy: Calculate the right predictions made over all sorts of predictions made by the prediction model. In general, how frequently is the classifier model exact.

$$Accuracy = TP+TN/TP+TN+FP+FN$$

Precision: The number of confirmed samples that have been identified properly.

$$Precision = TP/TP+FP$$

Sensitivity: The amount of real positive instances that have been identified right.

$$Sensitivity = TP/TP+TN+FP+FN$$

Specificity: The number of real negative instances accepted appropriately.

$$Specificity = FP/FP+FN$$

F-Score: Precision is vital for evaluating the efficiency of datamining classifiers, but it definitely leaves out details and will also be complicated for that purpose. The Recall is a part of the true optimistic predictions in the dataset for overall positive observations. Calculate the proportion of the churn rate that is correctly labeled as churn/non-churn. The low-recall prediction models indicate that a significant number of positive cases are miss-classified.

$$F\text{-}Score = 2 \text{ X } precision * sensitivity / precision + sensitivity$$

4. RESULTS AND DISCUSSION

In order to create a friendly user experience, this study was carried out and built using Matlab programming (MATLAB 2016A) with connected components of the MATLAB graphical user interface framework. In order to develop an output result of the data mining task with data filtering, feature selection using Relief-F, classification using ANN and CART, and performance evaluation, the built systems used different component environments in Matlab. Figure 2 shows the user interface and the loaded telecom dataset used in this syudy, 3333 samples with 21 attributes were loaded, Relief-F ranking was used as a feature selection technique to select relevant information from the data and passed into ANN and CART classifier separately.

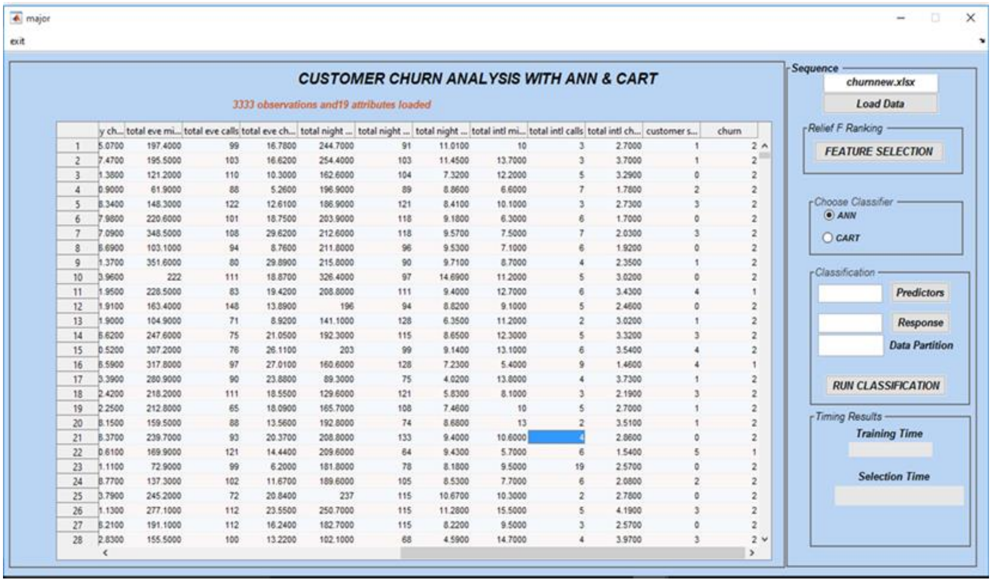

Figure 2. Loaded Telecom Customer Churn Data

For the input data matrix and response vector, the relief-f computes ranks and weights of attributes (predictors), the Releif-f filter selection method was able to identify the predicting variables according to their respective weight score with respect to the class mark. The characteristics on the positive response variable scale were chosen as the characteristics, totaling fourteen characteristics. Figure 3 shows the selected features using Relief-F algorithm, 14 features were selected from the given data as a subset dataset.

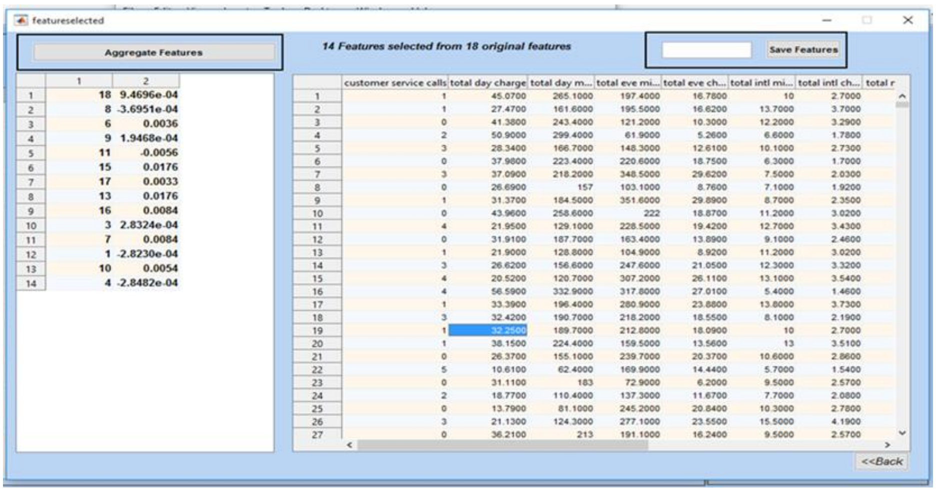

Figure 3. Selected Features Using Relief-F Algorithm

The selected data were passed to the training and testing set, the data was splitted into the training set and the data set was tested. For both the ANN and the CART classification algorithm, the system used 75% of the data for training. The loaded class mark indicates the split rate set at 0.25, which is an indicator of the data for both algorithms being 25% kept out.

4.1. ANN Training Approach

The 14 inputs, 10 neurons in the hidden layer, 1 neuron in the output layer with an activation role with 1 output were used for artificial neural network architecture. The 14 inputs reflect the churn dataset input data provided to the ANN with adjustable weight and bias (W,b), with 10 neurons the hidden layer was processed while the output layer was processed with one neuron in order to predict a single churners or non-churners outcome. 42.2313Sec the training computational time was used in processing the ANN for training the dataset, it is measured in terms of the total seconds used for executing the training process.

Based on each classification algorithm, as well as the comparative assessment of the two algorithms, the experimental results are mentioned. The evaluation parameter displays the KNN and SVM classifier combination results. With the True Positive rate (TP), False Positive (FP), True Negative (TN) and False Negative (FN), accuracy and error rate as shown in Table 1, the research (probing) assessment was achieved. Using False Acceptance Rate (FAR), False Rejection Rate (FRR), Accuracy (Recognition Rate), and Error Rate, the assessment parameters for classification rate were achieved.

Table 1. ANN Analysis per class

Analysis per class.	**True Positive**	**True Negative**	**False Positive**	**False Negative**
Class 1	88	694	18	33
Class 2	694	88	33	18

The confusion matrix is used as a description of the prediction results of this study on a classification issue. The number of correct and incorrect predictions is summarized and broken down by each class by counting values. Class 1 is true, which is the consumer who is likely to churn, while class 2 is false, which is the class of non-churners. Class 1 gives a total of 121 out of the test observation set, a total of 88 were correctly classified and 33 were misclassified, while the class of non-churners described by mark 2 gives a total of 712 out of the test observation set, a total of 694 were correctly classified and 18 were misclassified. Table 2 shows the Confusion matric used in ANN, with 88 =TP, 694=TN, FP= 18 and FN=33.

Table 2. ANN Confusion Matrix

	ANN Confusion Matrix	
	1	2
1	88	33
2	18	694

4.2. CART Training Approach

The CART analysis per each class based on the churners and non-churners class is shown in the table 3.

Table 3. CART Analysis per class.

Analysis per class.	True Positive	True Negative	False Positive	False Negative
Class 1	90	673	39	31
Class 2	673	90	31	39

To summarize the prediction outcomes on a classification problem, the Confusion matrix is used. The number of correct and incorrect predictions is summarized and broken down by each class by counting values. Class 1 is true, which is the consumer who is likely to churn, while class 2 is false, which is the class of non-churners. Class 1 gives a total of 121 out of the test observation set, a total of 90 were correctly classified and 31 were misclassified, while the class of non-churners described by mark 2 gives a total of 712 out of the test observation set, a total of 673 were correctly classified and 39 were misclassified. Table 4 shows the CART confusion matrix wherr TP=90, TN=673, FP=39 and FN=31. The real computing time used in the processing of the CART for the dataset training is taken and 7,811 seconds are used, which is calculated in terms of the cumulative usage time of seconds for the training phase to be performed.

Table 4. CART Confusion Matrix

	ANN Confusion Matrix	
	1	2
1	90	31
2	39	673

Table 5 shows a comparative result of the evaluation performance metrics for the classification of telcom churn prediction using ANN and CART classifier. The comparative results for the Artificial Neural Network and CART are shown in Table 5, which indicates that the ANN classification algorithm exceeded the CART classification algorithm for the telecom churn dataset, as shown in the table, as it gives a higher classification accuracy of 93.88% compared to 91.6% of the CART.

Table 5. Performance Metrics for ANN and CART

Performance Metrics (%)	ANN (%)	CART (%)
Accuracy	93.88	91.6
Sensitivity	72.7	74.38
Specificity	97.47	94.52
Precision	83.02	69.77
F-Score	77.53	72
Matthews Correlation Coefficient	74.22	67.11

In this study a feature selection approach using Relief-F was used to select relevant features from a huge churn telecom dataset, the relevant features were classified using ANN and CART, however the results of the classification show that ANN outperformed CART approach and suggested that this approach is an efficient one for this study compared with other existing works from literature, Table 6 compares the work with existing works.

Table 6. Comparison with Existing Works

Authors and Years	Work Done	Results (%)
Khalid et al., 2021 [35]	FE + Random Forest	91
Saini et al, 2017 [36]	CHAID + DT	91
Ahmad et al, 2019 [37]	X-Boost	89

The comparative analysis using uncertainty matrix research was conducted between Relief-F-ANN and Relief-F-CART. In order to verify the achievement, the assessment highlighted the accuracy

About R-F-ANN. Finally, device architecture that adopted MATLAB execution was then protected by the RF-ANN prediction procedures mechanism. In order to provide a better overview of telecommunications decision-making activities, the R-F-ANN prediction method was developed for data mining.

5. CONCLUSION

This research applied to the selection algorithm of a Relief-F function with ANN and CART classifiers on telecom customer churn prediction results. The issue of customer churn prediction is simultaneously important and difficult. In order to assist them in developing successful customer retention strategies, telecommunications companies invest more in creating accurate churn prediction model. An analysis of the application of Relief-F with ANN and CART was tested in this study and trained to predict customer churn in a telecommunications business. Experimental findings confirm that, compared to Relief-F-CART machine learning models, Relief-F-ANN achieves better generalization efficiency in terms of churn rate prediction with a highly reasonable precision rate.

REFERENCES

[1] U. Sivarajah, M. M. Kamal, Z. Irani, and V. Weerakkody, "Critical analysis of Big Data challenges and analytical methods," *J. Bus. Res.*, vol. 70, pp. 263–286, Jan. 2017, doi: 10.1016/j.jbusres.2016.08.001.

[2] B. He, Y. Shi, Q. Wan, and X. Zhao, "Prediction of Customer Attrition of Commercial Banks based on SVM Model," *Procedia Comput. Sci.*, vol. 31, pp. 423–430, 2014, doi: 10.1016/j.procs.2014.05.286.

[3] A. M. Naser alzubaidi and E. S. Al-Shamery, "Projection pursuit random forest using discriminant feature analysis model for churners prediction in telecom industry," *Int. J. Electr. Comput. Eng.*, vol. 10, no. 2, p. 1406, Apr. 2020, doi: 10.11591/ijece.v10i2.pp1406-1421.

[4] A. Rodan, A. Fayyoumi, H. Faris, J. Alsakran, and O. Al-Kadi, "Negative Correlation Learning for Customer Churn Prediction: A Comparison Study," *Sci. World J.*, vol. 2015, pp. 1–7, 2015, doi: 10.1155/2015/473283.

[5] K. O. Kadiri and S. O. Lawal, "Comparative Analysis of Per Second Billing System of GLO, MTN, Etisalat, Airtel and Visafone in Nigeria," *Curr. J. Appl. Sci. Technol.*, pp. 1–8, Apr. 2019, doi: 10.9734/cjast/2019/v34i230125.

[6] P. K. Banda and S. Tembo, "Factors Leading to Mobile Telecommunications Customer Churn in Zambia," *Int. J. Eng. Res. Africa*, vol. 31, pp. 143–154, Jul. 2017, doi: 10.4028/www.scientific.net/JERA.31.143.

[7] M. Singh, S. Singh, N. Seen, S. Kaushal, and H. Kumar, "Comparison of learning techniques for prediction of customer churn in telecommunication," in *2018 28th International Telecommunication Networks and Applications Conference (ITNAC)*, Nov. 2018, pp. 1–5, doi: 10.1109/ATNAC.2018.8615326.

[8] K. Kim, C.-H. Jun, and J. Lee, "Improved churn prediction in telecommunication industry by analyzing a large network," *Expert Syst. Appl.*, vol. 41, no. 15, pp. 6575–6584, Nov. 2014, doi: 10.1016/j.eswa.2014.05.014.

[9] A. Keramati and S. M. S. Ardabili, "Churn analysis for an Iranian mobile operator," *Telecomm. Policy*, vol. 35, no. 4, pp. 344–356, May 2011, doi: 10.1016/j.telpol.2011.02.009.

[10] T. Hennig-Thurau and U. Hansen, Eds., *Relationship Marketing*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2000.

[11] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, "A Review on Ensembles for the Class Imbalance Problem: Bagging-, Boosting-, and Hybrid-Based Approaches," *IEEE Trans. Syst. Man, Cybern. Part C (Applications Rev.*, vol. 42, no. 4, pp. 463–484, Jul. 2012, doi: 10.1109/TSMCC.2011.2161285.

[12] J. Brank *et al.*, "Feature Selection," in *Encyclopedia of Machine Learning*, Boston, MA: Springer US, 2011, pp. 402–406.

[13] R. J. Urbanowicz, M. Meeker, W. La Cava, R. S. Olson, and J. H. Moore, "Relief-based feature selection: Introduction and review," *J. Biomed. Inform.*, vol. 85, pp. 189–203, Sep. 2018, doi: 10.1016/j.jbi.2018.07.014.

[14] R. P. L. DURGABAI and R. B. Y, "Feature Selection using ReliefF Algorithm," *IJARCCE*, pp. 8215–8218, Oct. 2014, doi: 10.17148/IJARCCE.2014.31031.

[15] D. M. D. Raj and R. Mohanasundaram, "An Efficient Filter-Based Feature Selection Model to Identify Significant Features from High-Dimensional Microarray Data," *Arab. J. Sci. Eng.*, vol. 45, no. 4, pp. 2619–2630, Apr. 2020, doi: 10.1007/s13369-020-04380-2.

[16] R. Spencer, F. Thabtah, N. Abdelhamid, and M. Thompson, "Exploring feature selection and classification methods for predicting heart disease," *Digit. Heal.*, vol. 6, p. 205520762091477, Jan. 2020, doi: 10.1177/2055207620914777.

[17] T. Vafeiadis, K. I. Diamantaras, G. Sarigiannidis, and K. C. Chatzisavvas, "A comparison of machine learning techniques for customer churn prediction," *Simul. Model. Pract. Theory*, vol. 55, pp. 1–9, Jun. 2015, doi: 10.1016/j.simpat.2015.03.003.

[18] Y. Qu, Y. Fang, and F. Yan, "Feature Selection Algorithm Based on Association Rules," *J. Phys. Conf. Ser.*, vol. 1168, p. 052012, Feb. 2019, doi: 10.1088/1742-6596/1168/5/052012.

[19] J. Pamina, T. Dhiliphan, Rajkumar, S. Kiruthika, T. Suganya, and F. Femila, "Exploring Hybrid and Ensemble Models for Customer Churn Prediction in Telecom Sector," *Int. J. Recent Technol. Eng.*, vol. 8, no. 2, pp. 299–308, Jul. 2019, doi: 10.35940/ijrte.A9170.078219.

[20] J. Britto and Gobinath, "A Detailed Review For Marketing Decision Making Support System In A Customer Churn Prediction," *Int. J. Sci. Technol. Res.*, vol. 9, no. 4, pp. 3698–3702, 2020.

[21] Nisha Saini, Monika, and Dr. Kanwal Garg, "Churn Prediction in Telecommunication Industry using Decision Tree," *Int. J. Eng. Res.*, vol. V6, no. 04, Apr. 2017, doi: 10.17577/IJERTV6IS040379.

[22] A. K. Ahmad, A. Jafar, and K. Aljoumaa, "Customer churn prediction in telecom using machine learning in big data platform," *J. Big Data*, vol. 6, no. 1, p. 28, Dec. 2019, doi: 10.1186/s40537-019-0191-6.

[23] F. Kayaalp, M. S. Basarslan, and K. Polat, "TSCBAS: A Novel Correlation Based Attribute Selection Method and Application on Telecommunications Churn Analysis," in *2018 International Conference on Artificial Intelligence and Data Processing (IDAP)*, Sep. 2018, pp. 1–5, doi: 10.1109/IDAP.2018.8620935.

[24] S. Rai, N. Khandelwal, and R. Boghey, "Analysis of Customer Churn Prediction in Telecom Sector Using CART Algorithm," 2020, pp. 457–466.

[25] Francisco, "Churn in The Telecom Industry Dataset," 2017. https://bigml.com/user/cesareconti89/gallery/dataset/58cfbada49c4a13341003cba.

[26] T. T. Le *et al.*, “Differential privacy-based evaporative cooling feature selection and classification with relief-F and random forests,” *Bioinformatics*, vol. 33, no. 18, pp. 2906–2913, Sep. 2017, doi: 10.1093/bioinformatics/btx298.
[27] Z. M. Hira and D. F. Gillies, “A Review of Feature Selection and Feature Extraction Methods Applied on Microarray Data,” *Adv. Bioinformatics*, vol. 2015, pp. 1–13, Jun. 2015, doi: 10.1155/2015/198363.
[28] C.-L. Lin and C.-L. Fan, “Evaluation of CART, CHAID, and QUEST algorithms: a case study of construction defects in Taiwan,” *J. Asian Archit. Build. Eng.*, vol. 18, no. 6, pp. 539–553, Nov. 2019, doi: 10.1080/13467581.2019.1696203.
[29] A. SIMION-CONSTANTINESCU, A. I. DAMIAN, N. TAPUS, L.-G. PICIU, A. PURDILA, and B. DUMITRESCU, “Deep Neural Pipeline for Churn Prediction,” in *2018 17th RoEduNet Conference: Networking in Education and Research (RoEduNet)*, Sep. 2018, pp. 1–7, doi: 10.1109/ROEDUNET.2018.8514153.
[30] S. A. Qureshi, A. S. Rehman, A. M. Qamar, A. Kamal, and A. Rehman, “Telecommunication subscribers’ churn prediction model using machine learning,” in *Eighth International Conference on Digital Information Management (ICDIM 2013)*, Sep. 2013, pp. 131–136, doi: 10.1109/ICDIM.2013.6693977.
[31] X. Jiang *et al.*, “Genome analysis of a major urban malaria vector mosquito, Anopheles stephensi,” *Genome Biol.*, vol. 15, no. 9, p. 459, Sep. 2014, doi: 10.1186/s13059-014-0459-2.
[32] M. O. Arowolo, M. O. Adebiyi, A. A. Adebiyi, and O. J. Okesola, “Predicting RNA-seq data using genetic algorithm and ensemble classification algorithms,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 21, no. 2, p. 1073, Feb. 2021, doi: 10.11591/ijeecs.v21.i2.pp1073-1081.
[33] M. O. Arowolo, M. O. Adebiyi, A. A. Ariyo, and O. J. Okesola, “A genetic algorithm approach for predicting ribonucleic acid sequencing data classification using KNN and decision tree,” *TELKOMNIKA (Telecommunication Comput. Electron. Control.*, vol. 19, no. 1, p. 310, Feb. 2021, doi: 10.12928/telkomnika.v19i1.16381.
[34] M. O. Arowolo, M. Adebiyi, A. Adebiyi, and O. Okesola, “PCA Model For RNA-Seq Malaria Vector Data Classification Using KNN And Decision Tree Algorithm,” in *2020 International Conference in Mathematics, Computer Engineering and Computer Science (ICMCECS)*, Mar. 2020, pp. 1–8, doi: 10.1109/ICMCECS47690.2020.240881.
[35] L.F. Khalid, A.M. Abdulazeez, Y.H. Falah, D. Zeebaree, D.A. Zebari, " Customer Churn Prediction in Telecommunications Industry Based on Data Mining" *IEEE Symposium on Industrial Electronics and Applications*, 2021.
[36] N. Saini, Monika, K. Garg, " Churn Prediction in TelecommunicationIndustry using Decision Tree", *International Journal of Engineering Research and Technology*, Vol. 6, nno. 4, 2017. DOI: 10.17577/IJERTV6IS040379.
[37] A.K. Ahmad, A. Jafar, K. Aljoumaa, " Customer churn prediction in telecom using machine learning in big data platform", *Journal of Big Data*, Vol. 6, no. 28, 2019. doi.org/10.1186/s40537-019-0191-6.