EAI Endorsed Transactions
on Mobile Communications and Applications
Research Article EAI.EU

A Churn Prediction System for Telecommunication Company Using Random Forest and Convolution Neural Network Algorithms

Sulaiman Olaniyi Abdulsalam[1,*], Jumoke Falilat Ajao[1], Bukola Fatimah Balogun[1], and Micheal Olaolu Arowolo[4,5]

[1]Department of Computer Science, Kwara State University, Malete, Nigeria.
[2]Department of Computer Science, Landmark University, Omu-Aran, Nigeria.
[3]Department of Electrical Engineering and Computer Science, University of Missouri Columbia, Missouri, United State of America.

Abstract

INTRODUCTION: Customer churn is a severe problem of migrating from one service provider to another. Due to the direct influence on the company's sales, companies are attempting to promote strategies to identify the churn of prospective consumers. Hence it is necessary to examine issues that influence customer churn to yield effective solutions to minimize churn.
OBJECTIVES: The major purpose of this work is to create a model of churn prediction that assists telecom operatives to envisage clients that are more probably to be prone to churn.
METHODS: The experimental strategy for this study leverages the machine learning techniques on the telecom churn dataset, employing an improved Relief-F feature selection algorithm to extract related features from the enormous dataset.
RESULTS: The result demonstrates that CNN has a high prediction capability of 94 percent compared to the 91 percent Random Forest classifier.
CONCLUSION: The results are of enormous relevance to the telecommunication business in improving churners and loyal clients.

Keywords: Telecoms, Churn, Relief-F, CNN, Random Forest..

Received on 02 May 2022, accepted on 21 July 2022, published on 27 July 2022

Copyright © 2022 Sulaiman Olaniyi Abdulsalam *et al.*, licensed to EAI. This is an open access article distributed under the terms of the Creative Commons Attribution license, which permits unlimited use, distribution and reproduction in any medium so long as the original work is properly cited.

doi: 10.4108/eetmca.v6i21.2181

1. Introduction

Data-driven sectors have been able to carry out analysis of data and pull out significant knowledge through technology improvements. Methods of data mining have helped in obtaining the prediction of specific future customer behaviors [1]. Customer churn is categorized as customer attrition, it is among the most crucial factors that affect a company's earnings. The techniques of business intelligence for discovering customers who wish to change from a company to another competitor can be termed customer churn [2]. The telecoms sector is a highly technological industry that has expanded greatly in the past, as a consequence of the advent and commercial success of mobile telecommunications, over two decades [3-4].

For many telecoms organizations, client churn or customer attrition is a serious concern, it occurs when a customer quits his subscription and moves to another rival. Various

*Corresponding author. Email: sulaiman.abdulsalam@kwasu.edu.ng ; arowolo.olaolu@gmail.com

aspects affect the decision of the client to turn to another rival. In general, these characteristics were associated with the high cost, lousy jobs, fraud, and privacy issues related to customer service [5]. Customer churn causes a considerable loss of earnings when those thresholds are crossed. Companies know that obtaining fresh clients can be more expensive than retaining old ones [6].

In numerous areas, such as telecom providers, Internet service providers, credit cards, e-commerce, banking sectors, and newspaper publication organizations, among others, Consumer Churn Prediction (CCP) has been mentioned as a critical problem 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, and previous behaviors and offering some services to convince them to live[8]. Long-term clients, on the other hand, are more lucrative for service providers since they are more focused on purchasing additional goods and spreading the satisfaction of the consumer throughout their radius, so drawing more and more customers indirectly [9]. Businesses must have a deep awareness of why churn emerges to keep their clientele. There are various variables to be explored, such as organization discontent, specific businesses' competitive costs, customer migration, and the demand for better services for clients that might motivate users to switch to their present service provider and migrate to a different one [4]. Companies nonetheless, understand that gaining new clients is a great deal More pricey than current ones being retained [10].

In general, churn prediction obtained data are imbalanced, instances in non-churner customers may surpass churners class instances. Typical classification techniques seem to obtain relevant accuracy results for enormous classes and miss smaller ones, this is deemed one of the toughest and crucial issues. Different strategies have been presented to solve the issue of uneven churn prediction data. These techniques comprise sufficiently common evaluation matrices, usage of cost-sensitive learning, adjustment of training set distributions via method sampling, and the use of approaches to minimize dimensionality, among others [11][12].

Reducing dimensionality is of the essence in data mining, it is prompted by developing feature dimensionality in specified concerns and increasing interest in innovative yet costly computational approaches capable of modeling complicated relationships. Feature Selection is one of the ways of preprocessing to classify the data sub-set from large-dimensional data. In particular, feature selection algorithms such as Relief-F, and Genetic Algorithm, among others, are computationally fast, but responsive to complicated association patterns, such as associations, so that before downstream modeling, useful features are not incorrectly omitted [13]. Relief-F-based algorithms are a distinct group of filter-based feature selection algorithms that have garnered attention by establishing an efficient balance among these goals and efficiently adjusting to different input features [14] [15][16].

In an attempt to fine-tune established models, new investigations for churn analysis have proposed and suggested that methodologies such as SVM, ANN, and CART, among other classification methods are the most widely applied. Numerous optimization techniques and strategies have been examined and proposed to have been identified to run the best and make research in several domains such as telecommunications businesses, banking, business, and insurance, among others improved in productivity of these sectors [17][18].

In this study, an updated Relief-F feature selection algorithm is constructed with a novel learning approach by employing the subsets of the relevant churn prediction method data based on Random Forest and CNN classifiers using optimal predictors that increase the predictive output of variables. Using varied evaluation measures, and connected to standard prediction approaches as well as other relevant processes presented in the literature, the presented methodology will be evaluated.

The rest of this paper is structured as follows: Literature review of related works, research techniques of algorithms and dataset used, discussion of results, and conclusions.

2. Related Works

The loss of customers to competitors is defined as a major challenge for telecom carriers when it comes to classifying churners and non-churners. Pre-classifying customer turnover provides telecom companies with valuable knowledge regarding client retention. Some of the most popular churn categorization algorithms have been tested recently. Customer churn can be traced back to the service quality, customer satisfaction/dissatisfaction, and economic worth of a product or service, according to the most creative models. Using Random Forests with discriminating features for prediction research. Based on a developmental search Random Forest, they used discriminant feature research as innovation postponement of the classic Random Forest to learn slanted Developmental Detection tree to predict churners and non-churners in the telecoms business. PPtree construction uses two ways of discriminant investigation to measure the project catalog, and the proposed strategy controls the advantages of both. they produced oblique PP-tree classifiers that are more robust than typical Random Forest ones by combining Support Vector Machines with Linear Discriminant Analysis to achieve linear variable division. In terms of Accuracy, the detection methods are demonstrated to be superior. The LDA-based prediction model, PPForest, generates highly effective assessors. [4] A comparative analysis of customer attrition prediction utilizing Negative Correlation Learning has been offered [4]. Negative correlation learning is used to learn how to forecast customer attrition in the telecom business. Testing showed that churn analysis can be improved by using the NCL-MLP-ensemble rather than the non-NCL MLP-

ensemble, as well as other classic data mining methods. A detailed investigation of churn-based machine learning prediction in the telecoms business for eight years prompted the suggestion of investigating synthetic and ensemble techniques in the telecommunications industry [19]. It was discovered that telecommunication churn concerns and problems could not be predicted and that proposals and solutions to these problems were presented. Data specialists in telecommunications can use the study and summary to find the best methodologies and design strategies to build new models for predicting future churn. Many scholars have used a churn prediction support framework to study consumer purchase decision-making [20], and their findings demonstrate that a wide range of parameters is employed to construct a consumer churn model. It outlines the most recent churn prediction approaches. Modeling techniques, including neural networks, support vector machines, decision trees, logistic regression, and random forests, are used to uncover churn. It has been found that using predictive analytics, the customer churn projection can be more accurate than using conventional methods. Analytical analytics can be used to forecast customer churn and retain them. A Decision Tree-based churn study for the telecoms industry has been proposed [21]. The churn rate of a large sample of customers was factored into the decision tree categorization approach. The Exhaustive CHAID technique was found to be even more reliable and consistent in predicting customer turnover when SPSS implemented all potential decision tree versions. With machine learning, they proposed an approach to developing and selecting features for churn prediction in vast amounts of data [22]. The model's efficiency is assessed solely by looking at its Area Under Curve (AUC), and the AUC value achieved is 93%. This study's usage of social network analysis features to mine customer communal networks into the prediction model is an additional significant contribution. The model's output compared to the AUC benchmark rose from 84 to 93 percent while using the system network study. Using a massive dataset generated by transforming enormous raw data collected from Syriatel telecommunications firm, the model was systematized and evaluated in Spark settings. Syriatel used the database to train, test, and evaluate the categorization system over nine months. Random Forest, Decision Tree, Gradient Boost, and Extreme Gradient Boost were all tested as part of the process. Using the xgboost algorithm, on the other hand, leads to superior results. This method was used to classify data in this churn-predictive framework. Innovative attribute selection and framework telecommunications churn evaluation correlation have been proposed [23]. The material utilized for the analysis includes actual customer phone metadata from a big Turkish telecom business for the years 2013 and 2014. Five datasets were created using four different feature selection techniques: R-correlation coefficient, -correlation coefficient, Relief-F, and Gain Ratio feature selection procedures. Random Forest, Naive Bayes, Decision Tree, and AdaBoost were the next four classifier methods implemented. Accuracy, Specificity, Sensitivity, F-score, and run-time were utilized to analyze the results of the experiment. According to the correlations, the predicted feature selection algorithm outperforms the current best practices in terms of predicting user attrition. The CART algorithm has been proposed for predicting consumer churn in the telecom market [24]. It has been a hot topic in the telecoms business in recent years to study customer attrition. Because it can identify customers who are on the verge of canceling or changing their service subscription. It is possible to identify the reasons for customer turnover by analyzing data gathered from telecom providers. As a result, forecasting client attrition is critical for telecommunications businesses to keep them happy. Researchers in this study created a call tree classification model, assessed its output indicators, and compared it to a logistic regression model to see which performed better.

Several models have been proposed by several authors, however, CNN and Random forest are efficient models that are proposed to further enhance the classification for predicting bank credit risks.

3. Materials and Methods

An attempt is being made in the proposed learning to build a classification model that can predict whether a consumer in Telecom datasets is likely to switch providers. Customer relationship management can be improved by establishing important retention strategies that are likely to keep and attract customers who have the greatest inclination to leave. Customer turnover data from previous conversations, as well as personal and corporate records stored by telecom providers, was used to help develop a prediction model for customer churn. The test dataset and model must be able to predict churners after being fully trained with the training dataset after the prediction model. As depicted in Figure 1, the proposed steps for predicting churners are shown.

Machine learning is a technique for deriving usable information from large amounts of data. Analytics, math, artificial intelligence, and data science are used to extract useful information from a wide range of large datasets, and it provides advanced, valuable knowledge and data. According to data learning theory, classification, regression, clustering, and correlation can all be solved by machine learning if the research goal is clear. The data is arranged in a specific way. Using this strategy, the information is presented in a way that is both descriptive and intelligent.

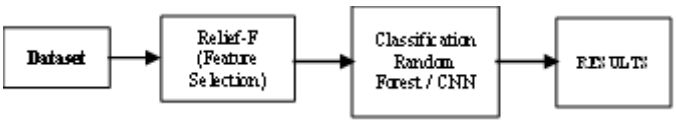

Figure 1. Customer churn prediction Approach using Relief-F with Random Forest and CNN models

3.1. Datasets

The realistic component of this research is based on telecom statistics gathered from the Francisco gallery of bigml.com and includes 20 attributes and 3333 cases. Customer churn data relative to the functionality and use of telephony account features [25]. A few of the dataset's primary attributes include name, account length, zone code, global plan, voicemail messages, and number v-mail, as well as all-day minutes, all-day calls, all-day charge, and all-day eve minutes and churn. [25].

3.2. Feature Selection based on enhanced Relief-F

The realistic component of this research is based on telecom statistics gathered from the Francisco gallery of bigml.com and includes 20 attributes and 3333 cases. Customer churn data relative to the functionality and use of telephony account features [25]. A few of the dataset's primary attributes include name, account length, zone code, global plan, voicemail messages, and number v-mail, as well as all-day minutes, all-day calls, all-day charge, all-day eve minutes, and churn. [25].

The most crucial stage in preprocessing data is to identify features that apply to the target variable. There are a few features that don't make much of an impact on the classifier learner model, however. Because of the large datasets used by telecom providers, the feature selection method has become critical for improving efficiency and making the customer churn prediction model easier to interpret, reducing overfitting, and eliminating variables that are redundant and add no value to the model's production. The prediction problem is also reduced and classification algorithms can generate answers as rapidly as feasible because of this.

The first Relief algorithm was influenced by instance-based learning. As a different assessment filtering feature selection procedure, Relief computes a deputation statistic for each feature that may be used to estimate the feature's prominence or significance to the target definition. Information on these features can be represented in the form of weights (e.g., W[A] = W[A]) or as "scores" (e.g., -1 (worst) to +1 (best)) (finest). Concerning binary classification and misplaced data, in particular, the Relief algorithm was unable to deal with these issues. Multi-class or constant endpoints necessitate relief approaches [13]. [13] The Relief-F Algorithm has always used pseudocode to emphasize the training of only a small subset of examples, rather than the entire dataset. Since it depends on the number of neighbors to improve weight estimation reliability and deal with noisy situations, Relief-F has emerged as the most often used form. It can handle missing data values and 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 $D[X] = 0.0$

for i: = 1 to m do

arbitrarily pick a 'target' instance R_i

discover an adjoining hit 'H' and adjoining miss 'M' (instances)

for X: = 1 to a do

$$D[X] = D[X]\, d_i ff\,(X, R_i, H)/d + d_i ff\,(X, R_i, D)/d$$

$$D[X] = D[X]\sum_{j=1}^{k} di\,\frac{ff(X,ri,hj)}{m.c} + O\,(n^2.a)\,(d+c_3{}^2c.n.\,a)$$

end for

end

return vector D of feature scores that estimate the feature values

The Relief-F uses neighboring instance changes in feature and class data to compute feature ratings. Relief-F reduces a feature's score if a group of neighbor instances has positive variations for that feature but the same class value. The function's score is also raised if neighboring instances exhibit positive fluctuations in a feature's value across distinct classes, according to ReliefF. Repeat this for a set of tested examples and their nearest neighbors to get an average score for each attribute [16][27]. An improved version of the Relief-F algorithm is proposed in this paper to retrieve important data from the churning telecom dataset. To classify the data, the findings of relief-f are used as a reduced preprocessed dataset.

3.3. Classification based on Random Forest

Random forest classifiers are a type of ensemble-based learning method, which is a big group. They are easy to set up, work quickly, and be very successful in many different areas. The key idea behind the random forest approach is that in the training stage, many "simple" decision trees are made, and in the classification stage, the majority vote (mode) across all of them is used. Among other things, this voting strategy fixes the fact that decision trees tend to overfit training data, which is not a good thing. During the training stage, random forests use a general method called "bagging" on each tree in the group. Bagging takes random samples from the training set and fits trees to these samples over and over again. Each tree grows without being cut back. The number of trees in the ensemble is a free parameter that can be easily learned automatically using a technique called "out-of-bag error."

Random forests are popular in part because, like naive Bayes and k-nearest neighbor-based algorithms, they are easy to understand and work well most of the time. But unlike the first two methods, random forests make it hard to predict how the structure of the final trained model will look. This is a natural result of the fact that building a tree is a random process. As we will talk about in more detail soon, one of the main reasons why this feature of random forests can be a problem for regulatory reasons is that

clinical adoption often requires a high degree of repeatability, not just in terms of how well an algorithm works in the end, but also in terms of how a specific decision is made [21], [28].

3.4. Classification based on CNN

A traditional neural network (ConvNet/CNN) is a deep learning algorithm that can take an image as input, figure out the importance (weights and bias) of different parts or objects in the image, and tell them apart from each other. A CNN needs a lot less pre-processing than other classification algorithms. In primitive methods, filters are made by hand, but convection grids can learn these filters and properties with the right training. It was made to classify images. it also has a lot of potentials. In sequential data analysis, such as NLP, convolution and pooling are two very important operations. the process of convolution used to pull out details from images (dataset). the pooling operation is used to reduce the number of dimensions of the features taken from the convolution operation. Operations like maximum pooling and average pooling are often used in CNNs. Backpropagation was used to transfer the gradient. The activation function, called relu, was used to do this. It is a type of deep neural network (DNN) used in studies of computer vision and natural language processing. It's like the way neurons connect in our brains, and it's the regular version of the multilayer perceptron found in networks that are fully connected. CNN's are made up of exactly one input layer, several hidden layers, and an output layer. The Convolution layer is in the hidden layer, relu (activation function). There are fully connected layers, pooling layers, and normalization layers. Compared to other algorithms used to classify images, this one is the best because it needs much less preprocessing and gets better as the number of training runs goes up. In natural language processing (NLP), CNN is used to find predictive features in large structures and make vectors that represent this structure. The CNN-based method uses a 1-D convolution operation to find the information needed for local word order search [29-31].

3.5. Evaluation Criteria and Experimental Setup

In this study, the MATLAB data mining tool has been used to change analytical data [32, 33]. To find a clear answer, it's important to have a good functional setup, use study variables, and have good performance metrics. The telecom industry uses a variety of performance measures [34] to figure out how well the churn prediction model works.

Accuracy: Figure out which predictions the prediction model got right out of all the different kinds of predictions it made. How often does the classifier model get it right?

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

Precision: The number of samples that have been correctly identified after being checked. Precision = TP/(TP+FP)

Sensitivity is the number of true positives that have been picked out correctly.

Sensitivity = (TP/TP+TN+FP+FN)/(TP/TP)

Specificity: The number of real negative situations that were dealt with in the right way. Specificity = FP/FN + FP

F-Score: Accuracy is important for figuring out how well data mining classifiers work, but this doesn't include enough details and will be too hard to use for that purpose. The Recall is part of the true optimistic predictions in the dataset for overall positive observations. Find out how much of the churn rate is correctly labeled as churn and how much is not churn. The low-recall prediction models show that a lot of positive cases are wrongly labeled as negative.

F-Score = 2 X precision

* sensitivity/ precision and sensitivity.

4. Results and Discussions

This study was done and built using Matlab programming (MATLAB 2016A) and connected parts of the MATLAB graphical user interface framework. This was done to make the user experience more pleasant. The built systems used different component environments in Matlab to develop the output of the data mining task, which included data filtering, feature selection using Relief-F, classification using Random Forest and CNN, and performance evaluation. Figure 2 shows the user interface and the loaded telecom dataset used in this study. 3333 samples with 21 attributes were loaded. Relief-F ranking was used as a feature selection technique to pick out relevant information from the data, which was then sent to Random Forest and CNN classifiers separately.

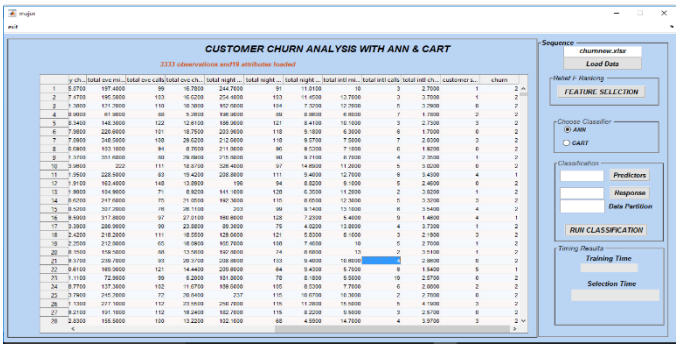

Figure 2. Loaded Telecom Customer Churn Data

For the input data matrix and response vector, relief-f calculates the ranks and weights of attributes (predictors). The Releif-f filter selection method was able to find the predicting variables based on their weight scores concerning the class mark. The characteristics were chosen from the scale of the positive response variable. There were a total of fourteen characteristics. Figure 3 shows the features that were chosen using the Relief-F algorithm. A

subset dataset of 14 features was chosen from the data that was given.

The chosen data were sent to the training set and the testing set. The data were then split into the training set and the testing set. For both the Random Forest algorithm and the CNN algorithm, 75% of the data was used to train the system. The loaded class mark shows that the split rate is set to 0.25, which means that 25 percent of the data for each algorithm is left out.

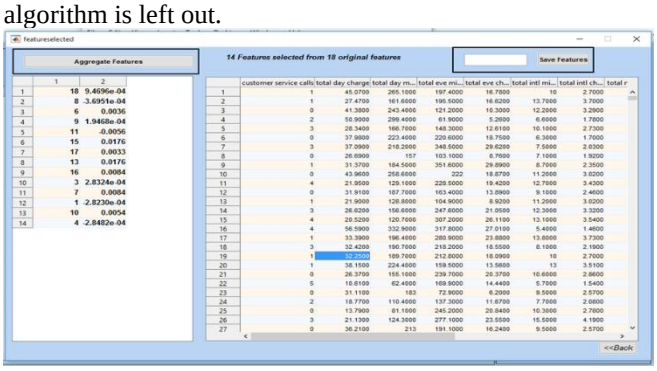

Figure 3. Selected Features Using Relief-F Algorithm

4.1. CNN Training Approach

The convolutional neural network architecture used 14 inputs, 10 neurons in the hidden layer, 1 neuron in the output layer with an activation role, and 1 output. The 14 inputs are the churn dataset input data that was given to the CNN with adjustable weight and bias (W, b). The hidden layer was processed with 10 neurons, and the output layer was processed with one neuron so that a single churner or non-churners outcome could be predicted. The total number of seconds used to run the training process was 42.2313Sec. This is the amount of time needed to process the CNN to train it on the dataset.

The results of the experiments are given for each classification algorithm, as well as a comparison of the two algorithms. The results of the classifier combinations are shown in the evaluation parameter. The research (probing) assessment was done with the TP, FP, TN, and FN, accuracy. With the help of False Acceptance Rate (FAR), False Rejection Rate (FRR), Accuracy (Recognition Rate), and Error Rate, the classification rate assessment parameters were met.

The confusion matrix is utilized as a description of the prediction outcomes of this investigation on a classification challenge. The quantity of precise and unfitting predictions is succinct and divided 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, whereas the class of non-churners described by mark 2 gives a total of 712 out of the test comment set, a total of 694 were correctly classified and 18 were misclassified. Figure 4 illustrates the Confusion matric used in CNN, with 88 =TP, 694=TN, FP= 18, and FN=33.

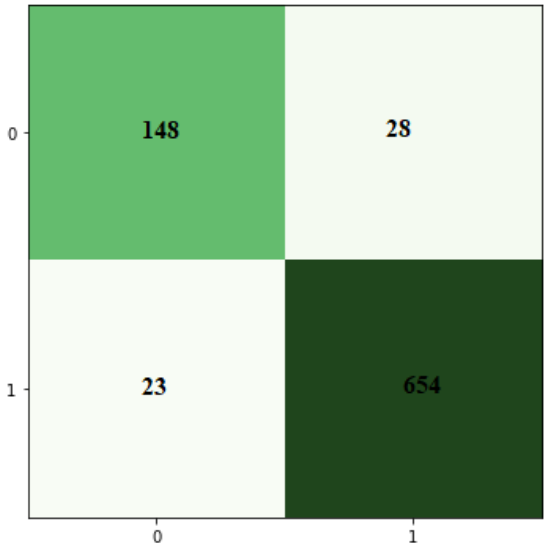

Figure 4. CNN Confusion Matrix

The Confusion matrix is used, to sum up, the predictions for a classification problem. By counting values, the number of right and wrong predictions is added up and broken down by class. Class 1 is true, which is the group of customers who are likely to churn. Class 2 is false, which is the group of customers who don't churn. Class 1 gives a total of 121 observations from the test set, of which 90 were correctly categorized and 31 were not. Mark 2 describes a class of non-churners, which gives a total of 712 observations from the test set, of which 673 were correctly categorized and 39 were not. Figure 4 shows the Random Forest confusion matrix when TP = 148, TN = 654, FP = 28, and FN = 23. The real amount of computer time used to process the Random Forest dataset for training is 7,811 seconds. This is calculated based on the total amount of time used for the training phase to be completed.

4.2. Random Forest Training Approach

The Random Forest analysis per each class based on the churners and non-churners class are shown in figure 5.

The Confusion matrix is used, to sum up the predictions for a classification problem. By counting values, the number of right and wrong predictions is added up and broken down by class. Class 1 is true, which is the group of customers who are likely to churn. Class 2 is false, which is the group of customers who don't churn. Class 1 gives a total of 121 observations from the test set, of which 90 were correctly categorized and 31 were not. Mark 2 describes a class of non-churners, which gives a total of 712 observations from the test set, of which 673 were correctly categorized and 39 were not. Figure 4 shows the Random Forest confusion matrix when TP = 148, TN = 654, FP = 28, and FN = 23. The real amount of computer time used to process the Random Forest dataset for training is 7,811

seconds. This is calculated based on the total amount of time used for the training phase to be completed.

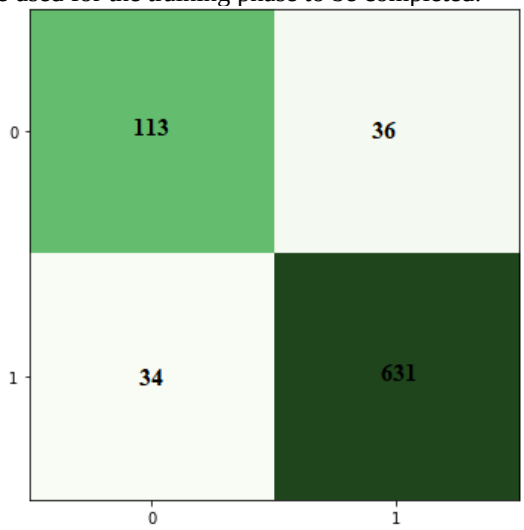

Figure 5. Random Forest Confusion Matrix

Figure 5 shows how the evaluation performance metrics for classifying telecom churn prediction using Random Forest and CNN classifier compare to each other. Table 1 compares the results of the Convolutional Neural Network and Random Forest. It shows that the CNN classification algorithm is better than the Random Forest classification algorithm for the telecom churn dataset, as it has a higher classification accuracy of 94 percent compared to 91 percent for the Random Forest.

Table 1. Performance Metrics for CNN and Random Forest

Performance Metrics (%)	CNN (%)	Random Forest (%)
Accuracy	94.02	91.40
Sensitivity	86.55	76.87
Specificity	95.89	94.60
Precision	84.09	75.84
F-Score	85.30	76.35
Matthews Correlation Coefficient	81.56	71.10

In this study, Relief-F was used to pick out relevant features from a large churn telecom dataset. The relevant features were then classified using Random Forest and CNN, but the results of the classification showed that CNN did better than Random Forest, suggesting that this is the best method for this study.

Using uncertainty matrix research, a comparison was made between Relief-F-CNN and Relief-F- Random Forest. To make sure the goal was met, the assessment focused on how accurate About R-F-CNN was. RF-CNN prediction procedures were then used to protect the device architecture that used MATLAB execution. The R-F-ANN prediction method was made for data mining to give a better overview of how decisions are made in telecommunications.

Comparing the study with existing works in literature, the study obtains relevant efficiency and can be adopted for credit risk evaluation.

Conclusion

This study used the selection algorithm of a Relief-F function with Random Forest and CNN classifiers to predict the number of customers who will leave a telecom company. Predicting how many customers will leave is important and hard at the same time. Telecom companies are putting more money into making accurate churn prediction models so that they can come up with better ways to keep customers. In this study, Relief-F, Random Forest, and CNN were tested and trained to predict customer churn in the telecommunications industry. The results of experiments show that, compared to Relief-F-Random Forest machine learning models, Relief-F-CNN does a better job of generalization when it comes to predicting churn rate with a high level of accuracy. This study is limited to credit risk data and can be adopted by larger datasets, it is recommended that the model can be adopted by authors to further enhance the model, and also adopt other machine learning models such as optimizers and classifiers.

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.