



Hybrid Artificial Neural Networks Using Customer Churn Prediction

P. Ramesh¹ · J. Jeba Emily² · V. Vijayakumar³

Accepted: 21 November 2021 / Published online: 1 December 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

The current wave of technologies with increased awareness among customers and retaining customers has a vital role in the growth of the company. A good indicator of service satisfaction of customers and service quality is customer churn. In order to enable the organizations to understand customers for churning, intelligible and accurate models are needed. There have been several techniques of data mining that were applied for the prediction of churn. The extensive research in Artificial Intelligence has made it feasible to study and learn the aspects accounting for such customer churn. The work presents effective solutions to all these challenging problems in Customer Churn Prediction (CCP). The study uses datasets in the telecommunication industry, the Artificial Neural Networks (ANN), and the Random Forests (RF) to determine the factors that influence consumer churn. A hybrid ANN-based work is proposed for predicting CCP. The results of the experiment proved that the proposed method achieves better levels of performance. The classification accuracy of ANN-4 hidden layers improves its result compared to RF and ANN-2 hidden layers. The maximum accuracy attained by ANN-2 hidden layers is 88.14% and by ANN-4 hidden layers is 90.34%.

Keywords Customer churn prediction (CCP) · Data mining · Random Forests (RF) · Artificial Neural Networks (ANN)

✉ P. Ramesh
pramesh.swami@gmail.com

J. Jeba Emily
jleenasamsy@gmail.com

V. Vijayakumar
vijayakumar.varadarajan@gmail.com

¹ Research Scholar, Sona College of Technology, Salem, India

² Present Address: Department of Information Technology, Sona College of Technology, Salem, India

³ School of Computer Science and Engineering, The University of New South Wales, Sydney, Australia

1 Introduction

Information technology has evolved rapidly in the recent two decades, resulting in a revolution for many fields involved. The primary area of the evolution of technology has a major impact on this. Various aspects of the organization are affected in a way in which the business processes had been completed, the training of employees, customer information analysis, and handling customer relationships for the process of decision-making [1]. The building and maintenance of customer relationships may not be new or be tied to internet usage and the web using the websites of the organizations with an improvement in the websites or the organizations. They also helped in improving the interest in customer satisfaction. The role played by the Customer Relationship Management (CRM) systems [2] that have become crucial to improve the lifetime value of customers. Using obtaining customer information, historical process, and profiles, the process of sales and marketing is employed. CRM is a vital component in the strategies of modern marketing. The knowledge of customers can be best built utilizing generating scores employing using predictive models. They extract different attributes and features to make use of them to predict various desired outputs like churn. The generated scores then get inserted into the organization's system to enable a better idea regarding customers.

There are two categories of Churners, which are voluntary and involuntary. The former will occur when the customers initiate service termination and are subdivided into incidental and deliberate churners. On the other hand, the latter will occur when the company initiates a termination from the list of subscribers [2]. The incidental churners will occur due to certain fluctuations in the location or their monetary state. However, deliberate churners occur due to the need for customers to change the price rate or technology. The customer churn is employed in the telecommunication industry and denotes the customer movement from one service provider to that of another. This is mainly due to dissatisfaction with service quality, bad support, wrong service plans, unattractive plans, and high costs. Also, the customer may quit the contract without aiming at switching competitors. The reasons for such changes making it impossible for customers to get service can be the change in the geographical location is resulting in non-availability of service or financial issues that need further service. At times, the company can also stop or further withdraw contracts owing to the company policy [3]. Generally, churn indicates the customers who tend to change their services to a service provider who is a competitor. The churn prediction methods will relate to the customers who might churn in the future, and churn management is helpful as it aims at identifying the churners and initiating certain positive actions to minimize the effects of such churn. Four different data variables are available for attrition of customers, such as customer behavior, perception of customers, customer demographics, and finally, macro-environment [4].

- *Customer behavior* helps in recognizing the components of service and utilize the frequency of employing them. For example, the length and number of calls in telecommunications, the period between the calls, and network usage for data exchange.
- *Customer perceptions* have been recognized as how the customer picks up or stops service and is computed using customer surveys, including the overall contentment of data link. It also includes vendor dependency, customer perception, the company's image, interest given, and problem handling satisfaction.
- *Customer demographics include* geographical data, social status, education, sex, and age to calculate churn.

- *Macro environment*-based variables are the changes seen in the world with a different customer experience affecting the way the service is used. For example, in the telecommunication trade, people that have survived natural disasters relying on mobile phones are likely to use this service.

Prediction of churn provides the company with a notion regarding consumers' tendency to leave their services for other service providers that are competitors. This information can be significant since it can help the company concentrate on such categories of customers utilizing a retention marketing promotion [5]. Retaining a consumer can be favorable to the company for various reasons, like drawing new customers costs more than the maintenance of the current ones. Secondly, the consumers having long-term experience will be less interested in the competitors and their market actions. A constructive Word-of-Mouth can appeal to several other new consumers, and the opposite can result in losing some consumers. They can also give higher levels of profit to the company. Thirdly, even a minor improvement in retaining consumers results in a surge in profit. Loss of customers, at the same time, can result in increased levels of opportunity costs.

Another main objective here is the retention of customers. The importance of such an objective is obvious as the actual cost of customer acquisition can be higher than that of retaining the customer. So, the tools used for developing and applying customer retention (the churn models) are important in the applications of Business Intelligence (BI). In a dynamic market environment, low levels of customer satisfaction, regulations, new products, and competitive strategies may result in churn. The churn models aim to identify the early churn signals and recognize the customers with a high likelihood of voluntarily leaving. In the recent decade, there is an increase in the interest in relevant studies that include gaming, insurance companies, and telecommunication [6–8].

Many types of research have been conducted to check the customer churn of companies and analyze churn prediction, but the accuracy rate is not up to the mark. With extensive research in Artificial Intelligence, it has become possible to dig to the core of the factor responsible for customer churn. There are many popular algorithms of machine learning that were proposed for tackling the problem of churning prediction. These methods include the ANN, rough set approach, linear discriminant analysis, market basket analysis, sequential pattern mining, Naïve Bayes, Support Vector Machines (SVM), logistic analysis and decision tree analysis. In this work, the RF and the ANN classifiers were proposed. The remainder of the paper is organized as related work in literature is presented in Sect. 2. The methods employed are discussed in Sect. 3. The experimental results are discussed in Sect. 4, and the conclusion is made in Sect. 5.

2 Related Works

Hudaib et al. [9] investigated three hybrid models for developing an efficient and accurate churn prediction model. The proposed models are based on two different stages: clustering and prediction. All customer data is filtered and grouped in the clustering stage, and in the latter, customer behavior is predicted. The first method investigates the k-means algorithm used for data filtering and the Multilayer Perceptron Artificial Neural Networks (MLP-ANN) to predict. The second method makes use of hierarchical clustering along with the MLP-ANN. The third method makes use of Self-Organizing Maps (SOM) along with the MLP-ANN. All methods were developed based on real data and accuracy. The Churn rate

values are computed and then compared to existing methods establishing the efficiency of the proposed method.

Farquad et al. [10] had proposed a new hybrid method to obtain rules from the SVM to be used in the CRM. This approach has three different phases. (i) In the first phase, the SVM-Recursive Feature Elimination (SVM-REF) was employed for reducing a feature set. (ii) The reduced features were used in the subsequent stage for obtaining the SVM model and the support vectors. (iii) After this, the Naive Bayes Tree (NBTree) is employed to generate the rules in the final stage. The dataset is analyzed, and the work analysis prediction of Churn in the bank credit card, with about 93.24% loyal and about 6.76% churned customers.

Saghir et al. [11] evaluated the current customers and the ensemble Neural Network (NN) based classifiers to suggest a new ensemble classifier that makes use of the NN to improve the performance of churn prediction. The work employed two datasets from GitHub and had an average accuracy of 81%.

De Caigny et al. [12] investigated the value-added, integrating textual data into the CCP versions. It further extended the previous work utilizing Convolutional Neural Networks (CNNs) versus the currently used best methods in the CCP. Firstly, the results had confirmed earlier research that showed the insertion of various textual data in the CCP model, improving predictive performance. Secondly, the CNNs were able to outperform the currently used best practices used in text mining. Thirdly, textual data was found to be a very important data source for the CCP. There was a computation of added profit from retaining customers through a campaign through the insertion of certain textual information used directly by the experts to help them make knowledgeable judgments regarding investing in text mining.

De Caigny et al. [13] proposed a new hybrid algorithm known as the Logit Leaf Model (LLM) for the classification of data. The primary idea behind this was that various other models were constructed on different data segments instead of the lead of the entire dataset, which could result in better predictive performance. This was by the maintenance of comprehensibility from various forms created into the leaves. There are two stages in the LLM. They are the segmentation and the prediction phase. Segments of the customer are categorized with decision rules in the segmentation phase, and a model is built for each leaf in the tree in the prediction phase. The proposed technique was benchmarked against all decision trees, logistic model trees, RF, and logistic regressions connected to comprehensibility and predictive performance.

Yu et al. [14] had proposed another Particle classification optimization-based Back Propagation (BP) telecommunication network CCP (PBCCP) algorithm that executes the Particle Fitness Calculation (PFC) and the Particle Classification Optimization (PCO). The PCO classifies all particles into three different categories following their fitness values and updated the velocity of various particles with distinctive equations. The PFC computes the fitness value of the particle for every forward training process in the BP-NN. The PBCCP further optimized their initial weights and thresholds in the BP-NN and brought about a remarkable level of improvement to the accuracy of CCP.

Ahmed et al. [15] proposed a new ensemble stacking incorporated with strategies based on uplifting for the telecom churn model of prediction. Some evaluations were performed based on conventional performance or cost heuristic, which was the primary focus. This type of operation had exhibited high correlations among the performance indicators and their business goals, enabling this algorithm to be well-suited for the other cost-sensitive applications. There was a new diverse ensemble designed by employing various algorithms that deliver predictions of a first level. The predictions, along with inconsistencies, were

handled in the second level by making use of a new heuristic based combiner that provides final predictions. The combination heuristics were fine-tuned based on the cost of predicting. Later, customer uplifting was achieved based on their final predictions, where the proposed model was 50% more cost-efficient than the state-of-the-art ensemble models.

Ullah et al. [16] proposed a new churn prediction model, using classification and clustering methods that distinguish churn customers, providing the factors working behind the customer churning for the telecom sector. The choice of features was achieved using a correlation attribute ranking filter and information gain. The proposed model has classified the churn customers' data by using classification algorithms where the RF algorithm has performed well using 88.63% of correctly classified instances. By creating effective retention policies, it becomes an important task for the CRM to manage to prevent churners. Once the classification was complete, the proposed segments churn the customers' data utilizing categorizing churn customers by using a cosine for providing offers of group-based retention. The work further identified various churn factors essential for the determination of the root causes of the Churn.

Wang et al. [17] had investigated the problem of CCP in the industry of Internet funds. The work had designed another novel method of Feature Embedded CNN (FE-CNN), which automatically learned the features from the customer behavioral data and the static customer demographic data utilizing the advantages of CNN. The results proved that the FE-CNN model was able to outperform all other machine learning models using hand-crafted features, like logistic regression, top-decile lift, Area Under the receiver operating characteristics Curve (AUC), NN for accuracy, RF and SVM.

Vijaya et al. [18] further proposed a new Rough Set Theory (RST) to identify all efficient telecommunication CCP features. The chosen features were given to ensemble classification techniques like random subspace, boosting, and bagging. For this work, the Duke University-churn prediction data was used for evaluating the various techniques. The proposed model's performance had been evaluated based on the metrics below, like accuracy, precision, specificity, false Churn, and true Churn. It had been identified that the proposed system using attribute selection with techniques of ensemble classification had an accuracy of classification of 95.13% than the other models.

Table 1 shows the comparison table for the works of literature reviewed in Sect. 2.

Churn prediction is a term widely used to identify the customers who are about to end their subscription or leave the company for another competitive service provider. In literature, machine learning algorithms such as CNN, Back Propagation, and Neural Network are used. Though accuracy is improved, there is a need to propose other different techniques for the best CCP. So this work proposes to predict customer churn using Random Forest (RF) Classifier and Artificial Neural Networks. The architecture of the ANN is critical for achieving improved performance; investigation using a varying number of layers is investigated in this work.

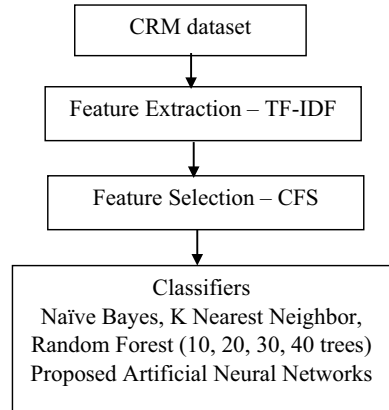
3 Methodology

Here, in this section, the ANN and RF classifiers are discussed. These classifiers are used for classifying the churners and non-churners. The Customer Relationship Management (CRM) dataset is used for evaluating the classifiers. The CRM dataset provides data of the American telecom companies focusing on customer churn prediction. Figure 1 shows the flow diagram for the proposed method.

Table 1 Comparison table for literatures used

| S.No | Author | Techniques/algorithm used | Merits |
|------|----------------------------|---|--|
| 1 | Hudaib et al. [9] | Multilayer perceptron artificial neural networks (MLP-ANN), self-organizing maps (SOM) | An efficient churn rate is obtained |
| 2 | Farquad et al. [10] | Support vector machine-recursive feature elimination (SVM-REF), naive bayes tree | Better performance in prediction of churn in the bank credit card |
| 3 | Saghir et al. [11] | Ensemble neural network (NN) based classifiers | Attain average accuracy of 81% |
| 4 | De Caigny et al. [13] | Convolutional neural networks (CNNs), | Better performance in CCP content to retain customer |
| 5 | De Caigny et al. [12] | Logit leaf model (LLM) | Better predictive performance for churn |
| 6 | Yu et al. [14] | Particle classification optimization-based back propagation (BP) telecommunication network CCP (PBCCP) algorithm, particle fitness calculation (PFC) and the Particle classification optimization (PCO) | Attain better accuracy for CCP |
| 7 | Ahmed and Maheswari [15] | New ensemble stacking for the telecom churn model of prediction | The proposed model was 50% more cost-efficient compared to the ensemble models |
| 8 | Ullah et al. [16] | A new model of churn prediction, correlation attribute ranking filter, and information gain | Better performance of RF algorithm |
| 9 | Wang et al. [17] | A novel method of feature embedded CNN (FE-CNN), SVM | Outperform all other machine learning models |
| 10 | Vijaya and Sivasankar [18] | A new rough set theory (RST) | Accuracy of classification of 95.13% is attained |

Fig. 1 Flow diagram for proposed method



3.1 Term Frequency-Inverse Document Frequency (TF-IDF)

TF-IDF is a commonly used term weighting approach for identifying a set of words in texts. Terms with low occurrence have higher TF-IDF scores and vice-versa. TF-IDF is computed using Eq. (1):

$$TF - IDF(t, d, D) = tf(t, d) \times idf(t, D) \tag{1}$$

wherein $tf(t, d)$ refers to term frequency which is the number of times a word occurs in an instance, while $idf(t, D)$ refers to inverse document frequency computed using Eq. (2):

$$idf(t, D) = \log \frac{N}{|\{d \in D : t \in d\}|}, \quad |\{d \in D : t \in d\}| \neq 0 \tag{2}$$

Here D refers to the complete set of instances, while N refers to the number of instances in the archive.

3.2 Correlation-Based Feature Selection (CFS)

CFS is a heuristic that selects features by measuring the usefulness of individual features in The correlation is computed by (3):

$$F_s = \frac{N\overline{r_{ci}}}{N + N(N - 1)\overline{r_{jj}}} \tag{3}$$

N is the number of features in the subset, r_{ci} is the mean feature correlation with the class, and r_{jj} is the average feature inter-correlation.

In this work, feature selection is used to reduce the dimensionality of the feature set. The features are extracted using TF-IDF and feature selection using CFS. A reduction of 36.4% of features was achieved using CFS.

3.3 Random Forest (RF) Classifier

The RF generates several classification trees. For classifying an object that is new from the input vector, it is put down with each tree in the forest. Every tree provides a classification saying the tree 'votes' for a class. The forest will select the classification with the most number of votes. The RF has the properties mentioned below [19]:

- It is considered to be good in terms of accuracy among other algorithms
- It is efficient on the large databases
- It can manage several input variables without any variable deletion
- It provides estimations on the importance of their classification
- An error of generalization with the progress of forest creating is generated
- Estimate any missing data
- It can handle unbalanced datasets
- It calculates the proximity among pairs used for clustering and locates the outliers (by means of scaling), providing an interesting view of data
- All capabilities of the RF are extended further to unlabelled data resulting in supervised clustering, outlier detection, and data views
- It further offers experimental methods to detect variable interaction

Once each tree is created, the data are run down, and proximities are computed for every pair of these cases. In case two of them fill a similar terminal node, their proximity will increase. At the end of such a run, their proximities will be normalized in connection to the actual number of trees. These proximities will be used for replacing any missing data, producing views that are low-dimensional, locating outliers.

The algorithm to construct the RF is:

1. Let the actual number of training cases be referred to as "n" and the number of variables in the classifier as "m".
2. Input variable used for making node decisions of the tree be "p" (Assumption: p is lower than m).
3. Select a new training set for a decision tree by selecting k times along with a replacement for the "n" available training cases. This is done by using a bootstrap sample. The remaining cases are used for the estimation of errors of a tree. Bootstrapping can also be employed for the estimation of training data properties.
4. For every node in the tree, the variables are chosen randomly for the best split. Now, new data is predicted by taking into consideration the majority votes of the tree.
5. The best split is computed on the basis of the chosen variables within the training set. The decision is based on the node with the best split.
6. Each such tree is grown fully and not pruned. Pruning is used for cutting the leaf nodes to ensure the tree grows further. The tree is retained.
7. The best split has the least error or least deviation from the dataset [20].

In this work, 10, 20, 30, and 40 trees for random forests were considered.

4 Proposed Artificial Neural Networks (ANN)

The ANN is a very complex network consisting of a large set of nodes (termed neural cells) that are simple. The ANN had been proposed on the basis of advanced biology research that is concerned with the tissue of the human brain and its neural system, which may be used for simulating the neural activities information processing. The ANN has a topological structure of the nodes that process information in the human brain. It also has the topological information structure to process nodes and disseminate information in a parallel manner. The input mappings and their projected outputs are got through combinations of their nonlinear functions. They use neural cells, association, memory, and experience for processing noise-containing, non-linear, and fuzzy data without resorting to mathematical models. The algorithms in the NN are the BP computation, Kohonen, Delta, and the Hebb. The BP computation was proposed in 1985 by the PDP group in Rumelhart.

The NNs may be differentiated into a single-layer perception or a Multi-Layer Perception (MLP) network. An MLP contains multiple layers of the two-state, sigmoid transfer function that is simple Sharma et al. [21]. Additionally, the NN also consists of either one or more of the many intermediary hidden layers known as the hidden layer nodes that are embedded. There is also a typical feed-forward MLP NN that contains an input, an output, and a hidden layer. The NN that adopts the error BP training algorithm is known as the BP network with a learning process comprising backward propagation and forward propagation. For the forward processing, sample signals progress through each such layer using the sigmoid function $f(x) = 1/(1 + e^{-x})$. The NN cell, which is the neuron of every layer, will affect the condition of the subsequent node. In case if the projected output for the layer is not achieved in its output layer, weight values for every layer in the neuron will be modified to minimize error. The incorrect output signals tend to be backward from their source. Lastly, a signal error arrives in an area with repeated propagation, and the weights are modified.

This network has been set at n layers and y_j^n , so that y_j^n is able to indicate the output with the n layers and the j nodes. If y_j^0 is found to be equal to x_j , and j denotes their inputs. Let W_{ij}^n refer to the connection weight existing between y_i^{n-1} and y_j^n , then it will get the threshold of θ_j^n for n layers and the j nodes. An NN learning algorithm has the following steps:

1. Initializing randomly the node connection weights.
2. The input signal vector, along with the required output, is read. These signals continue through the network using the formula below (4):

$$y_j^n = F(s_j^n) = F\left(\sum W_{ij}^n y_i^{n-1} + \theta_j^n\right) \tag{4}$$

It begins the calculation of j nodes for every layer with output y_j^n from its first layer to the completion of the processing of calculation. F (s) refers to the sigmoid transfer function.

3. Calculate actual output through computations, working forward.
4. Compute errors. An error value for each node in the output layer will be acquired from various values found between real and expected outputs (D_j^k) that is given by (5):

$$\delta_j^n = y_j^n(1 - y_j^n)(D_j^k - y_j^n) \tag{5}$$

Error values for each node in the previous (hidden) layers are dependent on backward error propagation for every such layer ($n=n, n-1 \dots 1$) as shown in (6):

$$\delta_j^{n-1} = F(s_j^{n-1}) \sum W_{ij}^n \delta_j^n \quad (6)$$

5. Change node connection weights which are by working backward from its output layer through hidden layers by using the formula below (7 and 8):

$$W_{ij}^n(p+1) = W_{ij}^n(p) + \eta \delta_j^n y_i^{n-1} + \alpha [W_{ij}^n(p) - W_{ij}^n(p-1)] \quad (7)$$

$$\theta_j^n(p+1) = \theta_j^n(p) + \eta \delta_j^n + \alpha [\theta_j^n(p) - \theta_j^n(p-1)] \quad (8)$$

Wherein p refers to iterative times of these layers. Constant η refers to the rate of learning, and α is the constant momentum with constant values between 0 and 1.

5 Experimental Setup

The algorithms for the prediction of churn were evaluated using the CRM dataset. The dataset is obtained from American telecom companies, which are widely used for Customer Churn Prediction. It consists of 51,306 subscribers, including 34,761 churners and 16,545 non-churners, from July 2001 to January 2002. In the dataset, subscribers in the same company for six months are considered to be mature customers, and Churn is computed based on the customer leaving the service within 31–60 days after the customer was sampled.

The experiments were conducted using the Matlab platform with Weka. The metrics used for evaluation are accuracy, precision, recall, and f measure. Accuracy is the most innate performance measure and represents the ratio of correctly predicted instances to the total instances. Precision relates to the low false-positive rate of the techniques. The recall is the ratio of correctly predicted positive instances to all instances in the actual class. F measure is the weighted average of Precision and Recall.

The performance metrics are computed as follows:

Accuracy is the proportion of true results (both True Positive (TP) and True Negative (TN)) among the total number of examined instances.

$$\text{Accuracy} = (\text{TP} + \text{TN}) / \text{Total instances} \quad (9)$$

Precision is calculated as the number of correct positive predictions divided by the total number of positive predictions.

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}) \quad (10)$$

where FP is False Positive.

Recall is calculated as the number of positive predictions divided by the total number of positives.

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN}) \quad (11)$$

where FN is False Negative.

F-measure is defined as the weighted harmonic mean of precision and recall.

$$F - \text{measure} = (2 * \text{Recall} * \text{Precision}) / (\text{Recall} + \text{Precision}) \tag{12}$$

6 Results and Discussion

For the experiments, the Naïve Bayes (NB), k Nearest Neighbor (KNN), RF (10, 20, 30 40 trees), ANN-2 hidden layer, and ANN-4 hidden layer methods are evaluated using the CRM dataset. The NB, KNN, and RF are existing algorithms used to compare the results of ANN using metrics like classification accuracy, recall, precision, and f measure. The experiments were conducted using five-fold cross-validation.

The classification accuracy, average recall, average precision, and average f measure are shown in Figs. 2, 3, 4 and 5.

From Fig. 2, it can be observed that the classification accuracy for ANN performs better than NB, KNN, and RF. The ANN-4 hidden layer has higher classification accuracy by 3.61% compared for RF with 40 trees and by 2.46% compared for ANN- 2 hidden layer, respectively. It is also seen that RF with 20 trees has slightly better accuracy than RF with 10, 30, and 40 trees.

From Fig. 3, the ANN-4 hidden layer has higher recall by 4.69% when compared to RF with 20 trees and by 3.6% for ANN-2 hidden layer, respectively. The ANNs are self-adaptive in nature, leading to the higher recall.

From Fig. 4, it can be observed that the precision for ANN performs better than other classifiers. The ANN-4 hidden layer has higher precision by 4.3% for RF and by 2.76% for ANN-2 hidden layer, respectively. The ability of ANNs of learning and to infer relationships, making it generalize and predict data improves the precision.

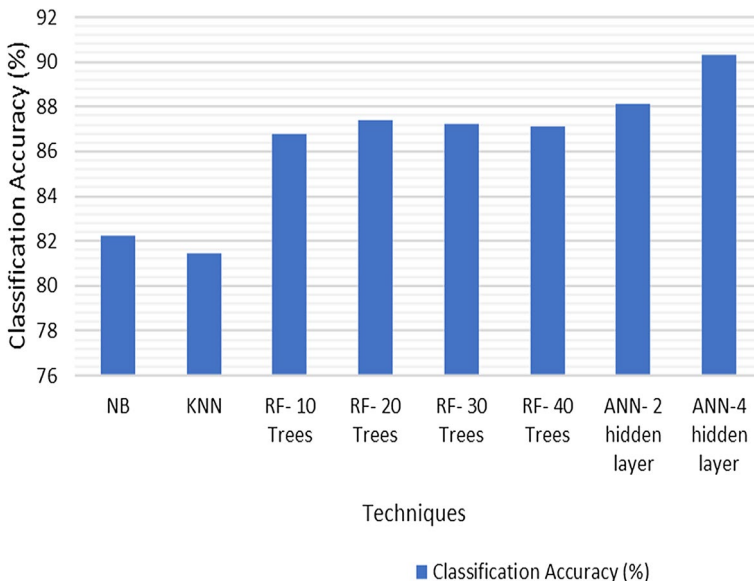


Fig. 2 Classification accuracy for ANN-4 hidden layer

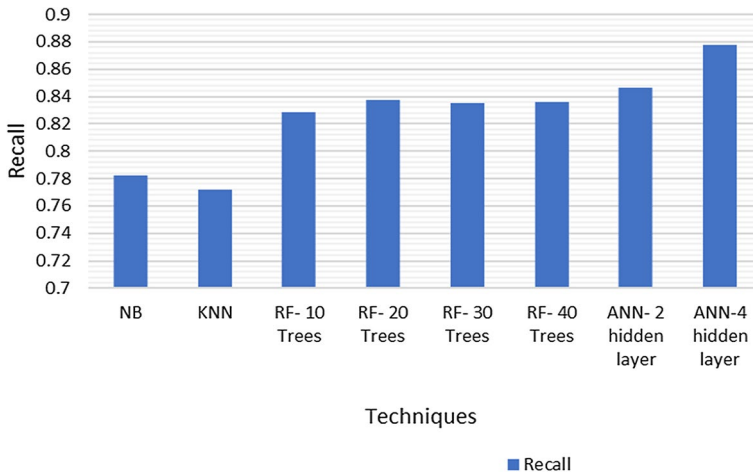


Fig. 3 Recall for ANN-4 hidden layer

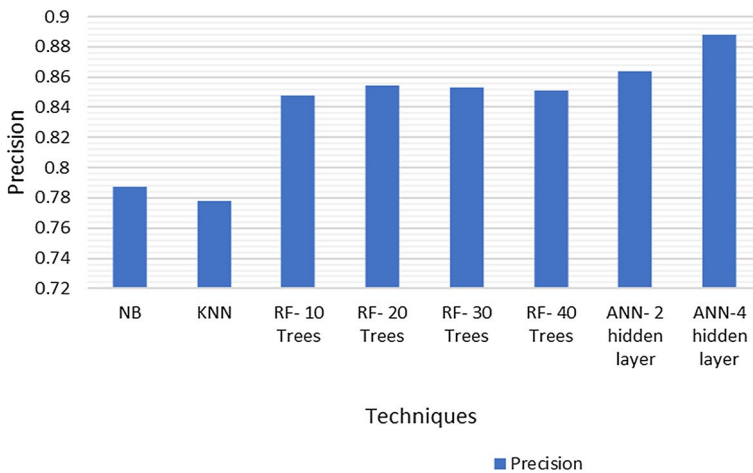


Fig. 4 Precision for ANN-4 hidden layer

From Fig. 5, it can be observed that the ANN-4 hidden layer has a higher f measure by 4.64% for RF with 40 trees and by 3.24% for ANN-2 hidden layer, respectively.

7 Conclusion

Prediction and management of churn are crucial for enterprises in today’s competitive market to predict various churners and take action for retaining customers and profits. Thus, to build a Customer Churn Prediction model that is effective in providing accuracy. In this work, two different techniques of data mining are considered, which are the RF and the ANN. The former is an ensemble one for classification. It includes the construction

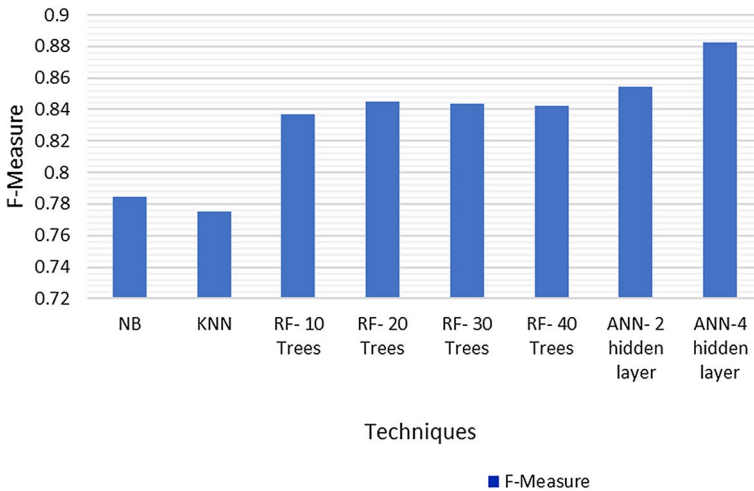


Fig. 5 F measure for ANN-4 hidden layer

of various decision trees and giving training data along with matching test data. It is also used for ranking all important variables in a problem of classification. The ANN further attempts to simulate all the biological neural systems that learn by means of changing the strength of a synaptic connection existing between neurons on repetitive stimulations using a similar impulse. The proposed ANN hidden layer performs in a better way when compared to RF. The results have shown that an ANN-4 hidden layer can give better classification accuracy by about 4.02% compared to the RF and by about 2.46% when compared to the ANN- 2 hidden layer. In the future, the existing classification accuracy can be enhanced by introducing structure optimization for ANN.

Funding Not applicable.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Dewnarain, S., Ramkissoon, H., & Mavondo, F. (2019). Social customer relationship management: An integrated conceptual framework. *Journal of Hospitality Marketing & Management*, 28(2), 172–188.
2. Santharam, A., & Krishnan, S. B. (2018). Survey on customer churn prediction techniques. *International Research Journal of Engineering and Technology*, 5(11), 3.
3. Huang, B., Kechadi, M. T., & Buckley, B. (2012). Customer churn prediction in telecommunications. *Expert Systems with Applications*, 39(1), 1414–1425.
4. Van den Poel, D., & Lariviere, B. (2004). Customer attrition analysis for financial services using proportional hazard models. *European journal of operational research*, 157(1), 196–217.
5. Shan, M. (2017). Building customer churn prediction models in fitness industry with machine learning methods.

6. Yolcu, G., Oztel, I., Kazan, S., Oz, C., & Bunyak, F. (2020). Deep learning-based face analysis system for monitoring customer interest. *Journal of Ambient Intelligence and Humanized Computing*, *11*(1), 237–248.
7. Vafeiadis, T., Diamantaras, K. I., Sarigiannidis, G., & Chatzisavvas, K. C. (2015). A comparison of machine learning techniques for customer churn prediction. *Simulation Modelling Practice and Theory*, *55*, 1–9.
8. Abbasimehr, H., & Shabani, M. (2020). A new framework for predicting customer behavior in terms of RFM by considering the temporal aspect based on time series techniques. *Journal of Ambient Intelligence and Humanized Computing*. <https://doi.org/10.1007/s12652-020-02015-w>
9. Hudaib, A., Dannoun, R., Harfoushi, O., Obiedat, R., & Faris, H. (2015). Hybrid data mining models for predicting customer churn. *International Journal of Communications, Network and System Sciences*, *8*(05), 91–96.
10. Farquad, M. A. H., Ravi, V., & Raju, S. B. (2014). Churn prediction using comprehensible support vector machine: An analytical CRM application. *Applied Soft Computing*, *19*, 31–40.
11. Saghir, M., Bibi, Z., Bashir, S., & Khan, F. H. (2019). Churn prediction using neural network based individual and ensemble models. In *2019 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST)* (pp. 634–639). IEEE.
12. De Caigny, A., Coussement, K., & De Bock, K. W. (2018). A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. *European Journal of Operational Research*, *269*(2), 760–772.
13. De Caigny, A., Coussement, K., De Bock, K. W., & Lessmann, S. (2019). Incorporating textual information in customer churn prediction models based on a convolutional neural network. *International Journal of Forecasting*, *36*, 1563–1578.
14. Yu, R., An, X., Jin, B., Shi, J., Move, O. A., & Liu, Y. (2018). Particle classification optimization-based BP network for telecommunication customer churn prediction. *Neural Computing and Applications*, *29*(3), 707–720.
15. Ahmed, A. A., & Maheswari, D. (2019). An enhanced ensemble classifier for telecom churn prediction using cost based uplift modelling. *International Journal of Information Technology*, *11*(2), 381–391.
16. Ullah, I., Raza, B., Malik, A. K., Imran, M., Islam, S. U., & Kim, S. W. (2019). A churn prediction model using random forest: Analysis of machine learning techniques for churn prediction and factor identification in telecom sector. *IEEE Access*, *7*, 60134–60149.
17. Wang, C., Han, D., Fan, W., & Liu, Q. (2019). Customer churn prediction with feature embedded convolutional neural network: An empirical study in the internet funds industry. *International Journal of Computational Intelligence and Applications*, *18*(01), 1950003.
18. Vijaya, J., & Sivasankar, E. (2018). Computing efficient features using rough set theory combined with ensemble classification techniques to improve the customer churn prediction in telecommunication sector. *Computing*, *100*(8), 839–860.
19. Keramati, A., Ghaneei, H., & Mirmohammadi, S. M. (2016). Developing a prediction model for customer churn from electronic banking services using data mining. *Financial Innovation*, *2*(1), 10.
20. Prajwala, T. R. (2015). A comparative study on decision tree and random forest using R tool. *International journal of advanced research in computer and communication engineering*, *4*(1), 196–199.
21. Sharma, A., Panigrahi, D., & Kumar, P. (2013). A neural network based approach for predicting customer churn in cellular network services. arXiv preprint. [arXiv:1309.3945](https://arxiv.org/abs/1309.3945)



P. Ramesh is a Research scholar in Sona College of Technology, Salem. He did bachelor degree in Computer Science and Engineering from Bharathidasan University. He did Master Degree in Computer Science and Engineering from Anna University. He is pursuing Ph.D. degree in the area of Big data analytics. He is having 10 years of teaching experience and 4 years industry experience. He published one international journal and presented 3 international conferences and 2 National conferences. He handled Guest lecture program and Hands on Training program for CSE students in the topic of Big Data Analytics.



Dr. J. Jeba Emilyn is currently employed as Associate Professor in the Department of Information Technology, Sona College of Technology, Salem. Teaching being her passion, she treasures the 18 years of teaching and research experience. Her area of expertise includes Machine Learning, Data Science, Big data Analytics, Business Intelligence, Information Retrieval and Database Management Systems. She received her Bachelor's degree in Computer Applications and a dual Master's degree – one in Computer Applications and one in Computer Science and Engineering. Her Doctoral Degree is in the faculty of Information and Communication Engineering from Anna University, Chennai. Her quest for research has persuaded her to publish research articles in peer-reviewed national and international journals and conferences, books and patents. She has bagged awards and laurels for her excellence in industry partnership programmes and certification programmes. In addition to teaching and research, she cherishes her expertise in building quality management systems and quality standards for academic institutions.



V. Vijayakumar received the Diploma degree (Hons.) in computer technology from the State Board of Tamil Nadu, the B.E. and M.E. degrees in CSE from Madras University, the Ph.D. degree in CSE from Anna University, and the MBA degree in HRD from Periyar University. Currently, he is working as a adjunct Professor in School of Computer Science and Engineering at The University of New South Wales, Sydney, Australia. He has more than 18 years of experience which includes eight years in teaching and eight years in the industry. He is also a Coordinator of the Cloud Computing Research Group. He is also a Coordinator for internship in India and Worldwide. In VIT, he involved in many Research and Development activities, he has also organized many national/international seminars/workshops/symposiums/conferences/special sessions in the area of cloud computing and big data which includes ISBCC'14, ISBCC'15, ISBCC'16, India, CCCA'14, Vietnam, and CCNC'14, USA. He had published many national/international articles in conferences, journals, and books. His research interests include grid computing, cloud computing, big data,

web semantics, and also involved in the domain such as biomedical application, mammogram, autism, immune system, and as well as key management, security issues in cloud, and grid computing. He is also a Professional Member of IAENG, CSTA, and ISTE.