www.czasopisma.pan.pl PAN www.journals.pan.pl

INTL JOURNAL OF ELECTRONICS AND TELECOMMUNICATIONS, 2023, VOL. 69, NO. 1, PP. 11–18
Manuscript received April 30, 2022; revised February, 2023. DOI: 10.24425/ijet.2023.144325

An Efficient Hybrid Classifier Model for Customer Churn Prediction

Anitha M A, and Sherly K K

***Abstract*—Customer churn prediction is used to retain customers at the highest risk of churn by proactively engaging with them. Many machine learning-based data mining approaches have been previously used to predict client churn. Although, single model classifiers increase the scattering of prediction with a low model performance which degrades reliability of the model. Hence, Bag of learners based Classification is used in which learners with high performance are selected to estimate wrongly and correctly classified instances thereby increasing the robustness of model performance. Furthermore, loss of interpretability in the model during prediction leads to insufficient prediction accuracy. Hence, an Associative classifier with Apriori Algorithm is introduced as a booster that integrates classification and association rule mining to build a strong classification model in which frequent items are obtained using Apriori Algorithm. Also, accurate prediction is provided by testing wrongly classified instances from the bagging phase using generated rules in an associative classifier. The proposed models are then simulated in Python platform and the results achieved high accuracy, ROC score, precision, specificity, F-measure, and recall.**

***Keywords*—customer churn prediction; bag of learners; ANN; SVM; regression; associative classifier; Apriori Algorithm**

I. Introduction

GLOBALIZATION dand advancements in the telecommunications industry result in an exponential expansion in the number of operators in the market, increasing competition. It is vital to optimize earnings on a regular basis in this competitive period, which has led to the creation of several approaches such as recruiting new customers [1], up-selling existing customers [2], and increasing client retention time [3], [4].

Marketing's focus has transitioned from products to consumers, as the marketing method has shifted from product to customer. Poor service quality, unhappiness with customer service, and high expenses are the primary causes of customer churn [5]. Customer acquisition and retention are two of the most crucial parts of any company's success. The process of recruiting new customers or persuading individuals to buy something is known as customer acquisition [6], [7]. It is a strategy for moving customers from brand awareness to purchase choices through the marketing funnel. Customer retention refers to a business's ability to turn one-time clients into repeat customers and keep them from switching to a rival [8]–[10].

The organization may provide better services or take appropriate steps against the unpleasant circumstance that causes the consumer to churn by forecasting clients who are ready to churn. Customer Relationship Management System (CRM) [11]–[13] is used to study consumer behavior to maximize profits. Lost consumers result in opportunity costs [14]–[16] as a result of lower revenues. As a result, a little increase in client retention might result in a substantial gain in profit [17]–[19]. To identify customers who are going to churn and their motivations, reliable and understandable churn prediction models are required [20], [21]. Customer Churn Prediction models have been developed to effectively predict customers intended to shift from one service to another. Data mining (DM) techniques have emerged as a viable option for analyzing client behavior. Decision trees, logistic regression, support vector machines, artificial neural networks, inductive rule learning, and more DM approaches can be found. They've been used to anticipate client behavior. However, the accuracy of some of these single model-based classifier approaches are insufficient, or there is a need for improvement in the prediction accuracy of others. The prediction accuracy can be improved by using ensemble methods that comprises more than one machine learning approach. Ensemble learners categorized into homogeneous and heterogeneous ensembles which include bagging(Bootstrap Aggregation) methods and boosting methods [22]–[24]. Here, the training set is selected at every iteration and it is based on the accuracy of previous training. It improves the prediction accuracy more than the single model-based classifier approach. However, for further improvement in performance, a hybrid model is a suitable choice. Hence, this study proposes an efficient hybrid classifier model that combines two multiple learning approaches; bagging and boosting for better churn prediction. The main contribution of this paper is given by,

- Preprocessed data is given to Bag of learners that determine wrongly classified instances and correctly classified instances separately.
- Associative classifier with Apriori Algorithm is proposed which acts as a booster to perform classification with high accuracy.

The proposed hybrid classification model enhances the accuracy in prediction performance of customer chum prediction. The content of the paper is organized as follows: section

Anitha M A is with Faculty of Computer Science and Engineering, College of Engineering Cherthala, Alappuzha, Kerala, India (e-mail: anithadevdas@gmail.com)

Dr Sherly K K is with faculty of Information Technology Department, Rajagiri School of Engineering & Technology Kochi-682039, Kerala, India (e-mail: sherlykk@rajagiritech.edu.in)

1 represents the introduction; section 2 presents the related work; the novel solutions are presented in section 3; the implementation results and its comparison are provided in section 4; finally, section 5 concludes the paper.

II. LITERATURE SURVEY

Ahmad et al [25] designed a churn prediction model that helps telecom carriers estimate which customers are most likely to churn. The model created in this paper employs machine learning techniques on a large data platform to create a novel approach to feature engineering and selection. The Area Under Curve (AUC) standard metric is used to assess the model's performance, and the AUC value achieved is 93.3 percent. Another significant addition is the extraction of Social Network Analysis (SNA) elements from the consumer social network in the prediction model. The use of SNA improved the model's performance from 84 to 93.3 percent when compared to the AUC benchmark. Working on a huge dataset obtained by converting enormous raw data, the model is constructed and evaluated in the Spark environment. However, the decline in outcome is attributed to the non-stationary data model phenomena which necessitate model training at regular intervals.

Kaur et al [26] utilized prediction models to predict clients who are likely to churn in the future. Because serving long-term consumers is less expensive than losing a customer, which results in a financial loss for the bank. In addition, returning clients generate larger benefits and new referrals. This research different machine learning models, such as logistic regression (LR), decision tree (DT), K-nearest neighbor (KNN), random forest (RF), and others, are used to estimate the likelihood of a customer churning.

Rahman et al [27] presented a strategy for predicting client attrition in a bank using machine learning techniques. This study encourages the investigation of the possibility of turnover by customer behavior. In this investigation, the KNN, SVM, Decision Tree, and Random Forest classifiers were employed. In addition, several feature selection algorithms have been included to identify the most important characteristics and system performance. Feature selection lowers the tree classifiers' prediction score. Unlike other classifiers, oversampling in SVM lowers the score. Hence, SVM is unable to handle the data effectively. However, machine learning algorithms along with advanced ensembling techniques such as boosting, bagging are required to improve the system performance.

Hoppner et al [28] created lucrative and hopefully interpretable churn prediction models. To choose the most lucrative churn model, the recently established anticipated maximum profit measure for customer churn (EMPC) has been presented. This work introduced a classifier that directly incorporates the EMPC measure into model development. ProfTree is a strategy for developing profit-driven decision trees that employ an evolutionary algorithm. ProfTree generates considerable profit increases over the standard accuracy-driven tree-based strategy in a benchmark analysis using real-life datasets from multiple communication service providers. However, to improve the profit, there is a need to maximize property by building a large collection of profit-induced trees and then aggregating them.

De Caigny et al [29] investigated the value contributed by introducing textual data into customer churn prediction (CCP) algorithms. It builds on previous research by comparing convolutional neural networks (CNNs) to current best practices for analyzing textual data in CCP and validating a framework that explains how textual data is incorporated into a predictive model using real-world data from a European financial services provider. First, the findings back with prior research shows that including textual input in a CCP model increases its predictive ability. Second, CNNs surpass existing text mining best practices in CCP. Third, while textual data is a significant source of information for CCP, unstructured textual data alone cannot produce churn prediction models that are competitive with those that employ structured data. However, this model does not reveal what kind of information the textual data truly contains and lowers churn rates in the long run.

Bayrak et al [30] discussed that churn prediction is critical to achieving. It's not easy to predict churning clients. It's much more difficult in the fast-food sector since customers might quit ordering for a variety of reasons. This study uses an alternative windowing technique in which the customer data structure is built sequentially with customers' distinct churn periods. To forecast customers' churn phases, a long short-term memory model is created with sequential data and compared to other conventional categorization approaches. However, this approach requires different windowing techniques and boosting strategies to calculate the features more accurately.

The model[[25] requires model training at regular intervals and in [26], oversampling in SVM lowers the score and is unable to handle the data effectively. The work [27] needs improvement in the system performance and in [[28], there is a need to improve the profit by building a large collection of profit-induced trees, and then aggregating them. Also [29] does not reveal what kind of information the textual data truly contains and lower churn rates in the long run. The model [30] requires different windowing techniques and boosting strategies to calculate the features more accurately. Hence to tackle the aforementioned issues a novel solution has to be developed.

III. AN EFFICIENT HYBRID CLASSIFIER MODEL FOR CUSTOMER CHURN PREDICTION

Predicting customer churn is critical in customer relationship management which involves analyzing past data to anticipate whether or not a client will leave the company in advance. A variety of methods for churn prediction have been presented previously, however, existing single model-based classification techniques increase the dispersion of prediction with overfitting problems and cause model performance degradation. Whereas in ensemble models, wrong model selection can lead to lower predictive accuracy. Hence to tackle the issues in single and ensemble classifier approaches, this work proposes a new hybrid classification model which combines the two categories of multiple learning approaches; bagging and boosting. First step is to test the individual classifier performance using telecom and bank datasets and identify the best performing classifiers. Here, the associative classifier with

Apriori algorithm is selected as a booster since it outperforms in both datasets. Then, a bag of learners based classification is employed in which the next three high-performing learners are chosen as base learners among SVM, ANN, Decision Tree, and Regression to estimate incorrectly and correctly classified instances. The base learners for the telecom dataset are ANN, SVM, and regression, while the base learners for the bank dataset are ANN, Decision tree, and regression. Based on their unique performance, suitable weight values for each classifier are assigned.

During testing, the test dataset is fed to the bag of base learners for classification, then wrongly classified instances generated from the base learners are fed to the booster phase using the rules created by the associative classifier. Hence, the proposed hybrid model provides a robust and accurate prediction of customer churn.

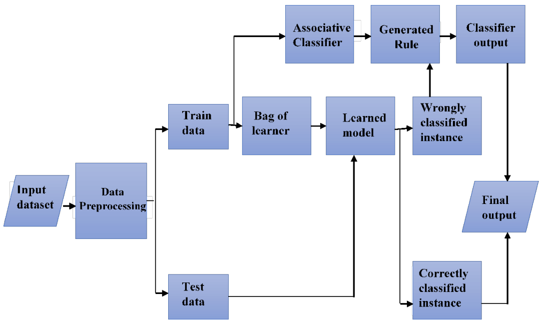

Fig. 1. Explain the significance of the figure in the caption

Fig. 1 depicts the architecture of the proposed customer churn prediction model. Initially, pre-processing is performed which includes the removal of tuples with missing data and the conversion of alphanumeric characteristics to numerical values. Then, using the training set create several classifier models and these models are then evaluated for their performance using test data and selects the best performing classifier models. The classifier models are given weights based on their performance. The associative classification is performed on frequent items generated from the original dataset using Apriori algorithm, which acts as a booster in this model. Pass the pre-processed dataset to the bag of learners which identifies correctly and incorrectly classified instances using weighted average. Then, send the wrongly classified instances from the bagging phase to the booster. Finally, performance is calculated by merging the output of the Associative classifier with the bagging result.

A. *Bag of learners based Classification*

Customer churn is a major concern in organizations when services are highly competitive. Hence prediction of customer churn is essential in business organizations. To enhance the prediction level of the model with the elimination of overfitting problems and model performance degradation, Bag of learners based Classification is presented in which the best models are selected to classify the customers into the churn and non-churn customers. Initially, pre-processing is done to remove tuples with missing data and convert alphanumeric characteristics to numerical values. Then, split the dataset into a training set and testing set. 70% of the pre-processed data is utilized as a training set, while the remaining 30% is used as a testing set. In Bag of learners based Classification, several classifier models are constructed using the training set, performance of each classifier model is evaluated using test data, and choose the best performing classifier model as a booster. The remaining classifier models are given weights based on their performance. The performance of a model is defined by equation (1),

$$Per_i = (A_i + R_i + S_i)/3 + F_i + RS_i \tag{1}$$

where Ai denotes accuracy, Ri denotes recall, Si denotes specificity, Fi denotes f-measure, and RSi denotes ROC score of the model Mi.

$$M^3_{i=1} = max^4_{i=1} Per_i \tag{2}$$

In equation (2), Mi represents the models selected for the bagging phase. In the bagging phase, the best of three models from Regression, SVM, Decision Tree, and ANN are used to categorize consumers into churn and non-churn. The test data is then used to make a prediction. The test data is originally supplied to the individual classifiers. Only three of the models are chosen as base learners in the bag of learners based on their performance.

Training sample with N observation is denoted by =x1,y1,x2,y2,. . . ,xN,yN , where yi has the value 1 or 0 indicates the response that the customer is churn or not. The classes churn and non-churn is indicated by C+ and C- respectively. The three models with the highest performance value according to equation (2) are selected as base learners in the bag of learners. Classification into one of the two classes is accomplished by,

$$y_i = \sum_{j=1}^{L} w_j d_j(x^t) \tag{3}$$

In equation (3), L is the number of base learners used in the bagging phase hence L = 3

d_j is the learner j
i is the number of classes, here, i = 2
y_i is the response of the respective model
w_j is the weight associated with j^{th} learner

For each model, the class is predicted by multiplying the predicted class probability of each model by the weight assigned and finding the label with the greatest weighted average probability. The weighted average probability is determined using the equation (4):

$$prob_{avg-weight} = \frac{1}{L}\sum_{j=1}^{L} w_j * I(d_j(x^t) = y_i) \tag{4}$$

In equation(4), $I(d_j(x^t) = y_i)$ is the indicator function used to count the tuple with the given response y_i for the learner d_j

The predicted class with maximum average weighted probability is given in equation (5) as,

$$C_i = max(prob_{avg-weight}) \tag{5}$$

The base learners for the telecom dataset are determined as ANN, SVM, and regression, while the base learners for the bank dataset are considered as ANN, Decision tree, and regression. Suitable weight values are assigned to each classifier based on their unique performance. Each classifier's weight is determined by

$$\omega_j = -\frac{1}{L}\log\frac{(1 - Per_j)}{Per_j} \tag{6}$$

In equation(6), $0 \leq \omega_j \leq 1$ and Per_j is the performance of j^{th} classifier by using the equation (1).The class with the largest weighted sum is the class label generated from the bag of learners. The algorithm for Bag of learners based Classification (Bagging phase) is illustrated in Algorithm 1.

Algorithm 1 *Algorithm for Bag of learners based Classification (Bagging phase)*

Training:
Input:X:$\{x^t, r^t\}_{t=1}^{N}$
for *j= 1 to L* **do**
  Train d_j using X
end
Testing:
for *each* $(x^t, r^t) in X$ **do**
  for *j= 1 to L* **do**
    $y_i = w_j d_j(x^t)$
  end
  if $y_i = 1$ **then**
    $Classoutput = C_+$
    $\theta_+ = \theta_+ + 1$
  else
    $Classoutput = C_-$
    $\theta_- = \theta_- + 1$
  end
end
if $\| \theta_+ - \theta_- \| > \tau$ **then**
  $Classoutput = C_i$ *which has highest* θ *value*
  Stop
else
  go to Phase2
end

Despite the fact that weights are allocated based on the performance of each individual model, there are two scenarios that might cause issues for a new instance.

Case 1: When the weighted sum of both classes(C+ and C-) are same: In the bagging phase, if the weighted sum is same for both classes, it is impossible to correctly forecast the instances. In such a circumstance, the boosting phase is required to properly forecast the outcome.

Case 2: If the difference between the weighted sums of C+ and C- is negligible. The difference between the weighted sums of the two classes is negligibly small, the prediction may be incorrect. This case also advances to the boosting phase, which predicts correctly. Let +represent the weighted sum of C+ and - represent the weighted sum of C-. The weighted sum difference threshold is then expressed in equation (7).

$$\| \theta^+ - \theta^- \| > \tau \tag{7}$$

However, the classified output from the bagging phase contains some wrongly classified instances due to false-positive rates. Hence, an associative classifier in the boosting phase is required to classify these outputs from the bagging phase to predict correct values.

B. Associative classifier with Apriori Algorithm

The accuracy of predicting customer churn is increased by using a booster. Associative classifier with Apriori Algorithm which functions as a booster. A strong model is utilized in this step to ensure that incorrectly classified instances are predicted properly. Associative classification is employed as a strong model. The apriori method is used to produce frequent itemsets and to construct rules in associative classification. The wrongly classified instances from the bagging phase are retested with the rules created by the associative classifier which works as a booster in order to obtain an accurate prediction.

$$P(d_j(x^t) \neq y_i) = \sum_{j=1}^{L}\sum_{t=1}^{N} w_j * I(d_j(x^t) \neq y_i) \tag{8}$$

Equation (8) denotes the wrongly classified instances from the bagging phase which is given to an associative classifier with apriori algorithm to predict customer churn with high accuracy.

Fig. 2 depicts the block diagram of the associative classifier with Apriori algorithm. The training data from the dataset is given to the Apriori algorithm to generate frequent item sets in which strong association rules are generated from frequent itemsets using apriori algorithm. Then ranking and pruning are performed to select a subset of rules. After the rules have been organized, pruning is done to remove the rules that are redundant or irrelevant. The classifier is made up of the remaining subset of ranking rules. Once the classifier subset of ranking rules has been discovered, this model is used to predict the class label for testing wrongly classified instances from the bagging phase with unknown class labels. The classifier is then assessed based on the accuracy of the test data item prediction.

The algorithm for Associative classifier with Apriori Algorithm (Boosting phase) is illustrated in Algorithm 2. All frequent itemsets in the dataset are mined using the Apriori algorithm. The algorithm does several database searches to locate common itemsets, with k itemsets being utilized to build k+1 itemsets. Each k-itemset must be higher than or equal to the minimal support criterion that is candidate itemsets. Then, the Apriori algorithm generates association rules by using frequent itemsets. It is based on the notion that a frequent itemset subset must also be a frequent itemset. A frequent itemset is one whose support value exceeds a certain threshold.

Rule pruning reduces the size of the classifier which improves performance and eliminates problematic rules which improve classification accuracy. As a result, the classification size has a significant impact on accuracy and efficiency. After

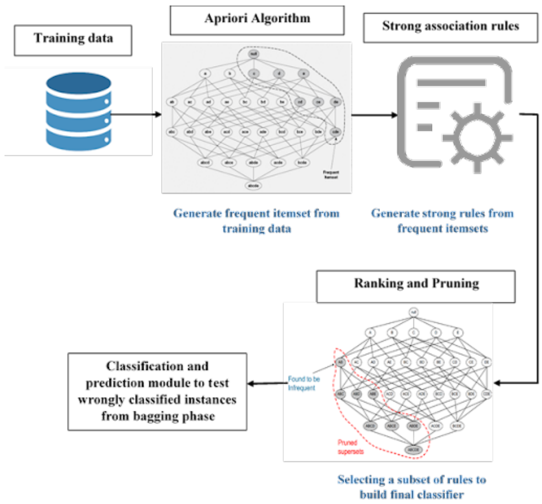

Fig. 2. Block diagram of Associative classifier with Apriori Algorithm

generating ranking rules, the classifier is built to test the wrongly classified output from the bagging phase. In the last step, performance is calculated by combining the output of the Associative classifier and the correct result obtained from the bagging phase.

Algorithm 2 Algorithm of Phase2(Boosting phase)

Training:
Input:$\{x^t, r^t\}_{t=1}^N$ **for** *each class* **do**
  Generate I_k is the frequent itemset of size k, using Apriori algorithm
  Generate strong association rules using confidence
end
Testing:
Wrongly classified instances from Phase1 is tested using the generated association rules and determine class output y_i
Analyze the performance

Overall an Efficient Hybrid Classifier Model for Customer Churn Prediction provides an accurate prediction of customer churn without overfitting and performance degradation using a Bag of learners based Classification and Associative classifier with Apriori Algorithm. The next section explains the result obtained from the Efficient Hybrid Classifier Model for Customer Churn Prediction and discusses it in detail.

IV. RESULT AND DISCUSSION

This section provides a detailed description of the implementation and result of the proposed prediction model simulated in Python platform. Also, provides the performance of the proposed model and a comparison section to ensure that the proposed model performs well.

A. Dataset Description

Telecom and bank data from Kaggle are taken as the dataset. The telecommunication data from Kaggle consists of 7043 instances with 21 features. Each row represents a customer and the column contains customer features. The bank dataset consists of 10000 instances and 15 features. Both datasets have an imbalanced problem as the target variable distribution has much more non-churned compared to churn.

The telecom dataset has two classes where 73.5% of data belongs to non-churn customers and 26.5% of data belongs to churn customers. This dataset includes four numerical features and 16 categorical features. Most of the categorical features have four or fewer unique values. Relevant features need to be selected to predict churn more effectively and accurately. In the bank dataset, 20.4% of the total data are churn customers and 74.6% are non-churn customers. This dataset includes 6 numerical features and 5 categorical features. Table I describes the telecom and bank datasets.

TABLE I
DATASET DESCRIPTION

Dataset	Instances	Attributes	Target class
Telecom	7043	21	Two classes, yes represent churn customers and no represents non-churn customers.
Bank	10000	15	Two classes, 1 represents the churn customer and 0 represents the non-churn customer.

B. Performance Analysis of the proposed method

The performance of the proposed model and the obtained results for ROC, accuracy, specificity, recall, and F-measure are presented in this section.

1) Area Under Curve Receiver Operating Characteristics curve (AUC-ROC curve): AUC-ROC curve is an approach to assess the performance of a classification problem. The probability curve is referred to as the ROC, and AUC indicates the degree of measure of class separability. It demonstrates how well the model discriminates between classes. If the AUC value is extremely close to 1, the model is efficient.

Fig. 3 depicts the AUC curve of various classification models in which receiver operating characteristic is determined. Both true positive and false positive rates values in the associative classifier are close to 1 since the value of area under the curve is close to 1, it is evident that the associative classifier model is efficient. Receiver Operating Characteristics score of classification models in both telecom and bank datasets are shown in Figure 4 and values are provided in Table II. Results given in Table II and Fig. 4 illustrate that in both datasets ROC score of associative classification is maximum. Apart from associative classification, ANN has a good ROC score compared to all other classification models for both datasets.

2) Accuracy, precision, recall, specificity, and F-measure: Since the dataset is imbalanced, the ROC curve alone is not a practical approach for assessing performance. As a result, it is preferable to calculate precision, recall, accuracy, and F-measure as well. The performance measure of classification models for the Telecom dataset is shown in Table III.

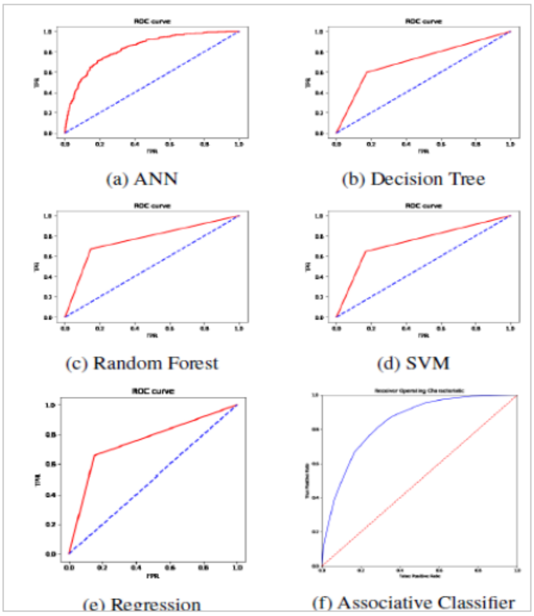

Fig. 3. AUC curve of various classification models

TABLE II
ROC SCORE OF CLASSIFICATION MODELS

Classification model	ROC Score for Telecom	ROC score of Bank
Regression	0.752	0.578
Decision Tree	0.716	0.798
Random Forest	0.726	0.830
SVM	0.739	0.556
ANN	0.751	0.811
Adaboost	0.740	0.815
Associative Classification	0.861	0.873

From Fig. 5, it is clear that the accuracy, precision, recall, specificity, and F-measure of associative classification for telecom dataset is high since associative classifier with Apriori algorithm which is considered as a strong model and acts as a booster that enhances the efficiency of prediction level. The

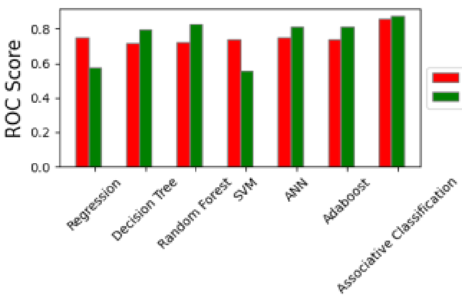

Fig. 4. ROC score of classification models for telecom and bank dataset

TABLE III
PERFORMANCE MEASURE OF CLASSIFICATION MODELS FOR TELECOM DATASET

Model	Accuracy	Precision	Recall	Specificity	F-measure
Regression	0.813	0.666	0.523	0.911	0.586
Decision Tree	0.787	0.589	0.518	0.878	0.551
SVM	0.799	0.645	0.456	0.915	0.534
ANN	0.809	0.654	0.518	0.907	0.578
Associative classification	0.844	0.728	0.618	0.921	0.669

performance measure of classification models for the bank dataset is shown in table IV.

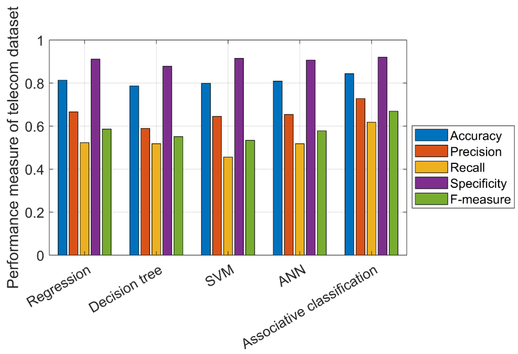

Fig. 5. Performance measure of telecom dataset

Fig. 6 shows that the accuracy, precision, recall, specificity, and F-measure of associative classification for bank dataset is also high. Hence, the associative classifier with Apriori algorithm is selected as a booster. Apart from associative classification, ANN has high-performance in all the measures for the bank dataset. Performance comparison of two ensemble classifier models (Random Forest and Adaboost) and proposed hybrid approach for telecom dataset are shown in table V.

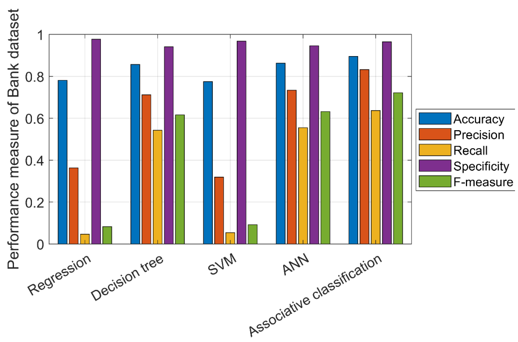

Fig. 6. Performance measure of bank dataset

Performance comparison of two ensemble classifier models, Random Forest and Adaboost and proposed hybrid approach for telecom dataset and bank are shown in Fig. 7 and Fig. 8 respectively. The proposed hybrid approach outperforms

TABLE IV
PERFORMANCE MEASURE OF CLASSIFICATION MODELS FOR BANK DATASET

Model	Accuracy	Precision	Recall	Specificity	F-measure
Regression	0.781	0.363	0.047	0.977	0.083
Decision Tree	0.857	0.712	0.543	0.941	0.616
SVM	0.775	0.319	0.054	0.968	0.092
ANN	0.863	0.734	0.555	0.946	0.632
Associative classification	0.895	0.832	0.637	0.965	0.722

TABLE V
PERFORMANCE MEASURE OF CLASSIFICATION MODELS FOR TELECOM DATASET

Model	Accuracy	Precision	Recall	Specificity	F-measure
Random Forest	0.793	0.615	0.481	0.898	0.540
AdaBoost	0.802	0.636	0.507	0.902	0.564
Hybrid Approach	0.871	0.773	0.644	0.932	0.702

the ensemble approaches since it combines both bagging and boosting approaches.

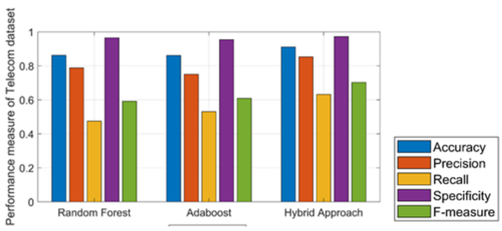

Fig. 7. Performance comparison of hybrid classifier for telecom dataset

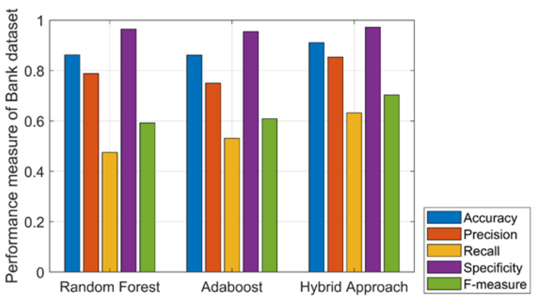

Fig. 8. Performance comparison of hybrid classifier for bank dataset

The test results illustrate that the proposed hybrid classification model outperforms the single and ensemble learners for both bank and telecom customer churn datasets.

TABLE VI
PERFORMANCE MEASURE OF CLASSIFICATION MODELS FOR BANK DATASET

Model	Accuracy	Precision	Recall	Specificity	F-measure
Random Forest	0.862	0.788	0.475	0.965	0.592
AdaBoost	0.861	0.75	0.531	0.954	0.609
Hybrid Approach	0.911	0.853	0.632	0.972	0.730

V. CONCLUSION

In this research, an Efficient Hybrid Classifier Model for Customer Churn Prediction has been proposed to eliminate overfitting and dispersion problems in the prediction of customer churn using Bag of learner based classification in the bagging phase and also removes issues in interpretability using Association classifier with Apriori algorithm in boosting phase to test wrongly classified instances from bagging phase. This model combines both bagging and boosting phases to generate accurate classification output. The proposed hybrid classifier model for customer churn prediction with bagging and boosting phase outperforms various existing classification models with high ROC, accuracy, specificity, precision, F-measure, and recall for both bank and telecom customer churn datasets. Hence, the hybrid classification model can be considered as the best suitable model for customer churn prediction.

REFERENCES

[1] Vo, N.Y. Nhi, et al. "Leveraging unstructured call log data for customer churn prediction," Knowledge-Based Systems, vol. 212, 2021, pp. 106586. https://doi.org/10.1016/j.knosys.2020.106586
[2] Amin, Adnan, et al. "Just-in-time customer churn prediction in the telecommunication sector," The Journal of Supercomputing, vol. 76, no. 6, 2020, pp. 3924-3948. https://doi.org/10.1007/s11227-017-2149-9
[3] Pustokhina, V. Irina, et al., "Multi-objective rain optimization algorithm with WELM model for customer churn prediction in telecommunication sector," Complex & Intelligent Systems, 2021, pp. 1-13. https://doi.org/10.1007/s40747-021-00353-6
[4] Ullah, Irfan, et al. "A churn prediction model using random forest: analysis of machine learning techniques for churn prediction and factor identification in the telecom sector," IEEE access, vol. 7, 2019, pp. 60134-60149. https://doi.org/10.1109/ACCESS.2019.2914999
[5] Ascarza, Eva, et al. "In pursuit of enhanced customer retention management: Review, key issues, and future directions", Customer Needs and Solutions, vol. 5, no. 1, 2018, pp. 65-81. https://doi.org/10.1007/s40547-017-0080-0
[6] Sharma, Tanu, et al. "Customer Churn Prediction in Telecommunications Using Gradient Boosted Trees." International Conference on Innovative Computing and Communications: Proceedings of ICICC 2019, Volume 2. Springer Singapore, 2020.
[7] Al-Mashraie, Mohammed, Sung Hoon Chung, and Hyun Woo Jeon, "Customer switching behavior analysis in the telecommunication industry via push-pull-mooring framework: A machine learning approach," Computers & Industrial Engineering, vol. 144, 2020, pp. 106476. https://doi.org/10.1016/j.cie.2020.106476
[8] Slof, Dorenda, Flavius Frasincar, and Vladyslav Matsiiako, "A competing risks model based on latent Dirichlet Allocation for predicting churn reasons," Decision Support Systems, vol. 146, 2021, pp. 113541. https://doi.org/10.1016/j.dss.2021.113541
[9] Lalwani, Praveen, et al. "Customer churn prediction system: a machine learning approach," Computing, 2021, pp. 1-24. https://doi.org/10.1007/s00607-021-00908-y

[10] Li, Yixin, et al. "Giant fight: Customer churn prediction in traditional broadcast industry," Journal of Business Research, vol. 131, 2021, pp. 630-639. https://doi.org/10.1016/j.jbusres.2021.01.022
[11] Pustokhina, V. Irina, et al. "Dynamic customer churn prediction strategy for business intelligence using text analytics with evolutionary optimization algorithms," Information Processing & Management, vol. 58, no. 6, 2021, pp. 102706. https://doi.org/10.1016/j.ipm.2021.102706
[12] Keramati, Abbas, et al. "Improved churn prediction in telecommunication industry using data mining techniques," Applied Soft Computing, vol. 24, 2014, pp. 994-1012. https://doi.org/10.1016/j.asoc.2014.08.041
[13] Ahmed, Ammara, and D. Maheswari Linen, "A review and analysis of churn prediction methods for customer retention in telecom industries," 2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS). IEEE, 2017. https://doi.org/10.1109/ICACCS.2017.8014605
[14] Shirazi, Farid, and Mahbobeh Mohammadi, "A big data analytics model for customer churn prediction in the retiree segment," International Journal of Information Management, vol. 48, 2019, pp. 238-253. https://doi.org/10.1016/j.ijinfomgt.2018.10.005
[15] Ullah, Irfan, et al. "A churn prediction model using random forest: analysis of machine learning techniques for churn prediction and factor identification in telecom sector," IEEE access, vol. 7, 2019, pp. 60134-60149. https://doi.org/10.1109/ACCESS.2019.2914999
[16] Amin, Adnan, et al. "Customer churn prediction in telecommunication industry using data certainty", Journal of Business Research, vol. 94, 2019, pp. 290-301. https://doi.org/10.1016/j.jbusres.2018.03.00
[17] Maldonado, Sebastián, et al. "Profit-driven churn prediction for the mutual fund industry: A multisegment approach," Omega, vol. 100, 2021, pp. 102380. https://doi.org/10.1016/j.omega.2020.102380
[18] De Caigny, Arno, et al. "Uplift modeling and its implications for B2B customer churn prediction: A segmentation-based modeling approach," Industrial Marketing Management, vol. 99, 2021, pp. 28-39. https://doi.org/10.1016/j.indmarman.2021.10.001
[19] Karimi, Nooria, et al. "Customer Profiling and Retention Using Recommendation System and Factor Identification to Predict Customer Churn in Telecom Industry." Machine Learning: Theoretical Foundations and Practical Applications (2021): 155-172.
[20] Slof, Dorenda, Flavius Frasincar, and Vladyslav Matsiiako, "A competing risks model based on latent Dirichlet Allocation for predicting churn reasons," Decision Support Systems, vol. 146, 2021, pp. 113541. https://doi.org/10.1016/j.dss.2021.113541
[21] Othman, Bestoon, et al. "The effects on service value and customer retention by integrating after sale service into the traditional marketing mix model of clothing store brands in China," Environmental Technology & Innovation, vol. 23, 2021, pp. 101784. https://doi.org/10.1016/j.eti.2021.101784
[22] Tianyuan, Zhang, and Sérgio Moro. "Research trends in customer churn prediction: a data mining approach." Trends and Applications in Information Systems and Technologies: Volume 1 (2021): 227-237.
[23] Khalid, Lawchak Fadhil, et al. "Customer churn prediction in telecommunications industry based on data mining." 2021 IEEE Symposium on Industrial Electronics & Applications (ISIEA). IEEE, 2021.
[24] De, Soumi, P. Prabu, and Joy Paulose, "Effective ML Techniques to Predict Customer Churn," 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA). IEEE, 2021. https://doi.org/10.1109/ICIRCA51532.2021.9544785
[25] Ahmad, Abdelrahim Kasem, Assef Jafar, and Kadan Aljoumaa, "Customer churn prediction in telecom using machine learning in big data platform," Journal of Big Data, vol. 6, no. 1, 2019, pp. 1-24. https://doi.org/10.1186/s40537-019-0191-6
[26] Kaur, Ishpreet, and Jasleen Kaur, "Customer Churn Analysis and Prediction in Banking Industry using Machine Learning," 2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC). IEEE, 2020. https://doi.org/10.1109/PDGC50313.2020.9315761
[27] Rahman, Manas, and V. Kumar, "Machine learning based customer churn prediction in banking," 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA). IEEE, 2020. https://doi.org/10.1109/ICECA49313.2020.9297529
[28] Höppner, Sebastiaan, et al. "Profit driven decision trees for churn prediction," European journal of operational research, vol. 284, no. 3, 2020, pp. 920-933. https://doi.org/10.1016/j.ejor.2018.11.072
[29] De Caigny, Arno, et al. "Incorporating textual information in customer churn prediction models based on a convolutional neural network," International Journal of Forecasting, vol. 36, no. 4, 2020, pp. 1563-1578. https://doi.org/10.1016/j.ijforecast.2019.03.029
[30] Bayrak, Ahmet Tuğrul, et al. "Personalized customer churn analysis with long short-term memory." 2021 IEEE International Conference on Big Data and Smart Computing (BigComp). IEEE, 2021.