International Journal of Engineering Trends and Technology
ISSN: 2231 – 5381 / https://doi.org/10.14445/22315381/IJETT-V70I6P202
Volume 70 Issue 6, 13-23, June 2022

Original Article

A Hybrid Machine Learning Model for Bank Customer Churn Prediction

Shawni Dutta[1], Payal Bose[2], Samir Kumar Bandyopadhyay[3], Midhunchakkaravarthy Janarthanan[4]

[1,2,3,4]*Computer Science and Multimedia, Lincoln University College, Petaling Jaya, Malaysia.*

[1]shawnidutta83@gmail.com

Received: 31 March 2022 Revised: 30 May 2022 Accepted: 01 June 2022 Published: 27 June 2022

Abstract - *Customer retention often plays an utmost important role for any organization to ensure better profitability. Recently, business organizations are moving toward an automated information-driven decision-making process. This study facilitates the development of a decision-making system that will gather knowledge from customer databases and segregate the customers who are likely to leave the organization. A hybrid machine learning model has been proposed in this study that will boost predictive performance. Two sets of machine learning-based models, namely; single learners and ensemble learners, are applied to the customer retention database. The models are implemented by adjusting underlying parameters to infer the best predictive performance. Finally, the best model from each category is picked up and assembled to construct the hybrid model. From single learner and ensemble-based models, SVM and Adaboost turn out to be promising models, respectively. Hence, the SVM and Adaboost models are unified under a single platform, which significantly outperforms other pre-existing models. The proposed hybrid model (SVM-Adaboost) can provide informed decisions to the business organizations regarding the customer retention strategy with an efficiency of 87%.*

Keywords - *Support Vector Machines (SVM), AdaBoost, SVM-Adaboost, Churn prediction, Machine learning (ML).*

1. Introduction

In any business policy, 'Churn' is the strongest keyword that needs to be concentrated. The term 'Churn' is centralized on a company's customer retention schemes, which directly impacts the company's profit and future growth [1]. The rapid market expansion in every sector has resulted in service providers having a larger subscriber base. Customer engagement costs are growing regarding new competitors, fresh and unique business strategies, and improved offerings. Companies have grasped the need to secure on-hand consumers in such a quick setup. Service providers must avoid churn, a phenomenon in which clients discontinue using a company's services. The migration of people among different banks is termed customer churn. Unhappiness with customer service, unreasonable fees, incompatible plans, and poor support are the fundamental causes of churn. It is expensive to address in many industries because engaging new clients is five to six times higher than keeping existing customers. The capacity to forecast whether or not a certain client is likely to churn is a substantial extra potential revenue source for any company. Aside from the obvious drop in revenue that comes with a client departing, the costs of acquiring that customer may not even be compensated by the consumer's expenditure yet [2]. One of the key goals of Client Churn Prediction is to assist in the development of customer retention strategies. As markets for supplying services become more competitive, the potential for customer churn rises rapidly. As a result, developing techniques to retain loyal consumers (non-churners) has proven to be essential.

An automated decision-making system construction has been proposed in this study to assist customer retention for a company. Machine learning (ML) is a promising field of computer science that can provide automated solutions to complex problems that applying conventional approaches might unsolve. In some cases, Machine learning methods generate a model based on a dataset and execute that model to predict the labels for new input data. This approach is popularly called the supervised learning paradigm [3,4].
The key points of the investigation are outlined as follows-

- Application of Pre-processing techniques is encouraged, which will yield a better dataset.
- Single learner models are applied to the pre-processed dataset with necessary hyper-parameter optimization.
- Ensemble learner models are also applied to the pre-processed dataset using the parameter tuning process.
- The best model from both the single learner and ensemble learner phases is identified in terms of predictive performances.
- These top two models are assembled to ensure enhanced predictive performance.

2. Related Works

Customer churn is one of the supreme concerns that businesses face. This study aids in identifying crucial moments in the consumer journey where users are losing interest. It allows the particular tactics to be developed to improve their encounters with the product and increase their commitment. Several articles and journals have already been conducted in the context of customer churn prediction. Table 1 provides a complete overview of some existing articles.

Table 1. Complete Overview of Some Existing Articles

Problem Statement	**Method Used**
Customer Churn Prediction [5]	Logistic Regression (LR), DT, K-NN, and the SVM method are implemented as machine learning approaches.
Customer Churn Prediction in the Telecom industry [6]	Using Bayesian Models, LR; Customer Relationship Management is modeled.
Customer Churn Prediction [7]	LR, Random Forest, and lazy learning classification models for performance analysis
Text analytics-based dynamic customer churn prediction technique for business intelligence [8]	To categorize the reduced feature data, the chaotic pigeon-inspired optimization-based feature selection (CPIO-FS) technique and long short-term memory (LSTM) with stacked auto encoder (SAE) model are used.
Customer Churn Prediction in The Telecommunications and E-Commerce Industry [9]	LR, Random Forest, Artificial Neural Network, and Recurrent Neural Network were used for prediction.
Customer Churn Prediction in Banking Sector [10]	This study employs the KNN, SVM, DT, and Random Forest classifiers.
Customer Churn Prediction in Telecom Industry [11]	For prediction, DT, Random Forest, as well as XGBoost classifiers were utilized in this model.
Customer Churn Prediction [12]	The authors construct a dataset using practical surveys and analyze it using Deep Learning, LR, and NB models.
Customer Churn Prediction in Telecom industry in Big data platform [13]	Four algorithms were compared in the model: DT, Random Forest, Gradient Boosted Machine Tree "GBM," and Extreme Gradient Boosting.
A Survey of Customer Churn Prediction [14]	This study examines the most prominent ML models adopted by the researcher for forecasting customer churn across different industries.

3. Backgrounds

Arthur Samuel, an American IBMer and innovator in computer games and machine intelligence, created the term "machine learning" in 1959. The term self-learning computers were also popular in this scenario. Machine learning (ML) analyzes computational models that spontaneously modify themselves by gaining knowledge. ML techniques often accompany Machine Intelligence. ML methods utilize sample data, often called training data, to construct a model for generating forecasts or judgments without being explicitly programmed [3].

3.1. Single learner and Ensemble Learners

Machine learning-based object classification methods demand enormous training and testing data. In this circumstance, single learners are the most effective since they start to learn knowledge about object categories from a small number of training examples.

Ensemble learning refers to systematically constructing and assembling a larger number of models to resolve a particular problem. This technique enhances the performance of the individual learning model or lessens the risk of a weak model being selected at random [15].

3.1.1. SVM

Support Vector Machine (SVM) is an approach widely used for classification analysis, but it can also be used for regression analysis. The model aims to create the optimal decision boundary for classifying N-D data. It can simply place incoming instances in the appropriate class as created by the decision boundary, which is nothing but an optimal hyper-plane. SVM generates the highest data instances to contribute to the formation of the hyper-plane. These data instances are characterized as Support vectors, and the underlying process is called SVM [16].

3.1.2. K-NN

The K-NN approach, also known as the K-Nearest Neighbor algorithm, ensures equivalence between the new data and existing instances and puts the fresh instance in the group that is quite close to the original classes. This algorithm preserves all existing data and categorizes new data points depending on matches. It merely saves the information throughout the learning process, and when a new instance enters, the closest matched class is assigned to it. It is characterized as a lazy learner strategy since it does not immediately acquire knowledge from the training set; rather, the data are stored and saved to apply the suitable classification action later [17].

3.1.3. Naïve Bayes

The Naive Bayes (NB) Method is a well-known and efficient machine learning method that aids in the construction of powerful tools to make promising forecasts. This model forecasts depending on the likelihood of an instance utilizing probability as defined by Bayes' Theorem. [18]. It is termed Naive since it implies the presence of one characteristic that is not impacted by other features. [18].

3.1.4. Decision Tree

The Decision Tree (DT) possesses a tree-based mechanism to address the classification and regression problems. The attribute of the experimental dataset is represented by the non-leaf nodes present in the tree, whereas the edges of each non-leaf node model the possible decisions. Finally, the prediction outcomes are assigned to the tree's leaf nodes. The prediction is calculated starting from the parent node and gradually proceeding through the different branches of the tree until a leaf node is encountered [19].

3.1.5. Adaboost

Adaboost is an ensemble boosting technique. Yoav Freund and Robert Schapire created this approach. AdaBoost, which stands for Adaptive Boosting, is a prominent boosting approach. This approach encompasses numerous weak learners to form a strong predictive model. At first, a model is constructed using the training data, followed by a second model designed to overcome the errors committed by the first model. The procedure, as mentioned earlier, is executed repeatedly until either the whole training data gets exhausted for predictions or the maximum number of models is applied [20].

3.1.6. Gradient boosting

Gradient boosting, a sequential ensemble learning technique, is often used to address classification and regression problems. This technique produces greater efficiencies over iterations. The model is inferred by allowing the development of an absolute differentiable loss function. This ensemble model used weak learners to produce a new model which provides a much more precise evaluation of the predictor variables. It frequently outperforms the random forest model. Using a stage-wise fashion, the gradient-boosted model employs various differentiable loss function generalizations [21].

3.1.7. Random Forest

This approach incorporates different decision trees within a learner to strengthen the forecasting power. Based on the majority of forecasts provided by different base learners, this model generates the final prediction instead of relying on an independent decision tree. Tin Kam Ho invented it in 1995, using the random subspace method. Leo Breiman and Adele Cutler created a modification of the technique and registered "Random Forests" as a brand in 2006 as a Bagging approach for ensemble learners [22].

3.2. Dataset Details

The objective of this paper is fulfilled by utilizing an existing dataset collected from the Kaggle repository [23]. The dataset has 1014 numbers of customer records with 13 attributes. The different attributes present in the dataset include customer id, surname, geographical location, Credit card holder or not, estimated salary, Credit Score of the customer, gender, age, the amount in the account, tenure of investment, and churn tendency. In this dataset, the churn tendency becomes the target or dependent variable, whereas the other parameters belong to the independent variable set. However, some independent attributes are irrelevant while inferring the churn prediction. These irrelevant attributes, such as customer id and surname, are eliminated from the independent variable set. For the further pre-processing procedure, the numerical attributes such as tenure, estimated salary, age, and credit score are scaled into a particular range between 0 and 1. Next, the data set is bifurcated into the training and testing dataset with a ratio of 7:3. The classifier model learns the patterns from the training data and creates the knowledge base. Later, that knowledge base is utilized for evaluation purposes. The training and testing datasets are characterized by the presence of the dependent or target variable. The training dataset contains the independent and dependent variables to construct the knowledge base. In contrast, the testing dataset will have only independent variables to make predictions. The distribution of the dependent variable, i.e., churn tendency in the collected dataset, is represented in Fig. 1.

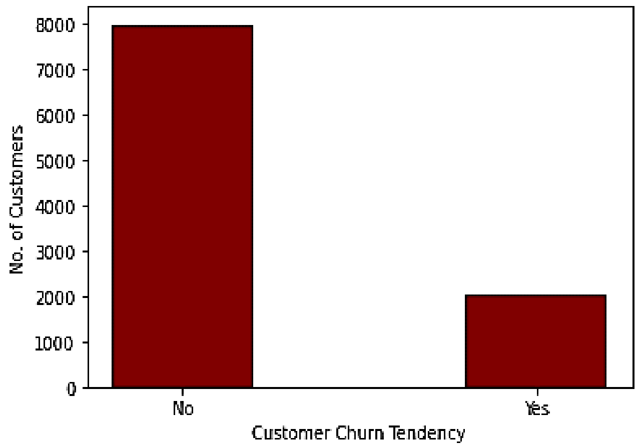


Fig. 1 Customer Churn Tendency Distribution

4. Methodology

The current study proceeds through a multi-step classification process. During the first phase, single learner models such as SVM, DT, NB, and K-NN are implemented, whereas the ensemble learning paradigm is approached in the second phase. Random Forest, Adaboost, and Gradient Boost are popular models which follow ensemble learning policy. The best predictive model is chosen and assembled from each section of the classification process. This assembling process generates a hybrid model, which can even boost the prediction performance of the other employed models. The workflow of this employed methodology is depicted in Fig. 2. The hybrid process is implemented by using the concept of the stacked ensemble model [24], which assembles two best models picked up from single learner and ensemble learners.

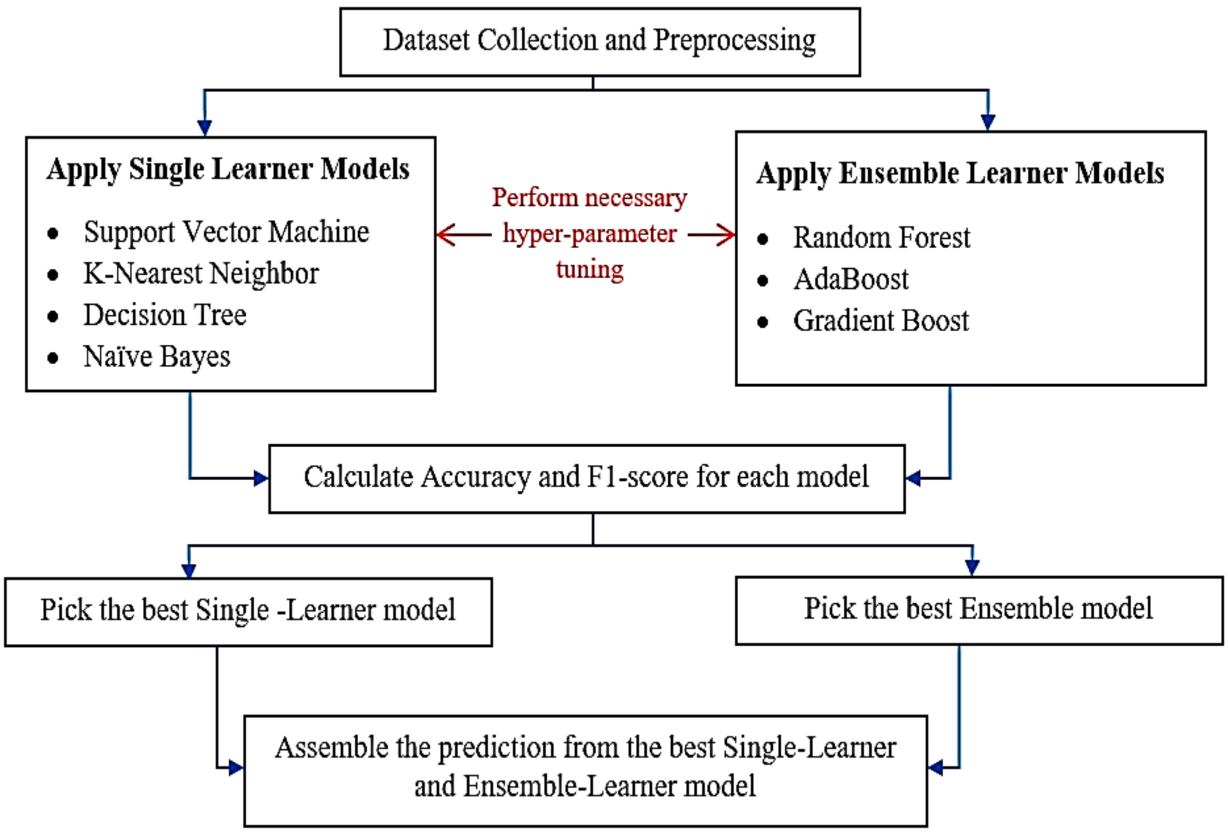


Fig. 2 Workflow of Employed Methodology

5. Experimental Results and Discussions

The workflow of the presented study follows two classification phases to identify the best candidate models for hybridization. In the first phase, single learner models are implemented, and an exhaustive comparison is employed to select the best predictive single learner model. During the second phase, another comparative analysis is provided among the ensemble-based approaches, and finally, the best ensemble model is recognized. These two retrieved models are assembled and stacked into an entity to assess customer churn prediction. The study is motivated to boost the predictive performance so that an automated model could provide enhanced informed decisions.

5.1. Phase 1: Single Learner Models

In this phase, the employed models, SVM, DT, NB, and K-NN, are implemented with necessary hyper-parameter tuning. This section elaborates on the implementation details of the models as mentioned earlier.

- For SVM, proper kernel specification is always favorable for identifying the significant performance. The performance of SVM is shown in Table 2, with the variation of kernel specification. Radial Basis Function (RBF), Polynomial Kernel, Linear Kernel, and Sigmoid Kernel are utilized to measure SVM performance. The results depicted in the figure reveal that Radial Basis Function is the best Kernel that provides enhanced prediction with an accuracy of 85.9% and F1-Score of 0.86.

- The K-Nearest Neighbor model is very sensitive to the proper selection of K-value. This model is a distance-based classifier; hence the distance metric can also impact the model's performance. Different values of K ranging from 1 to 30 are employed with the combinations of different distance metrics such as 'Chebyshev,' 'Euclidean,' and 'Manhattan.' The figure 3, 4, and 5 depict forecasting performance in terms of 'Chebyshev,' 'Euclidean' and 'Manhattan' respectively. The top outcome from each graph is obtained. The k-value of 11 and the 'Chebyshev' distance metric provide the accuracy of 83.6% and an F1-score of 0.8360. Another combination of the K-value of 17 with the distance metric 'Euclidean' has enhanced accuracy of 83.95% and 0.8395. The distance metric 'Manhattan' has reached the best accuracy of 84.95% and an F1-score of 0.8495 for the k-value of 10. The comparative analysis is further tabularized in Table 3. The evidence shows that the distance metric 'Manhattan' and k-value of 10 turn out to be the best hyper-parameter for this model.

- For the decision tree model, different criteria and splitters are incorporated. The model performance has been measured using 'Gini' and 'entropy' criteria along with the 'best,' 'random' splitting technique. The detailed descriptions of predictive outcomes are explained in Table 4; as per the forecasts described in Table 4, combining the 'entropy' criterion and 'random' splitter produces the best predictive solutions with an accuracy of 79.75% and an F1-Score of 0.797.

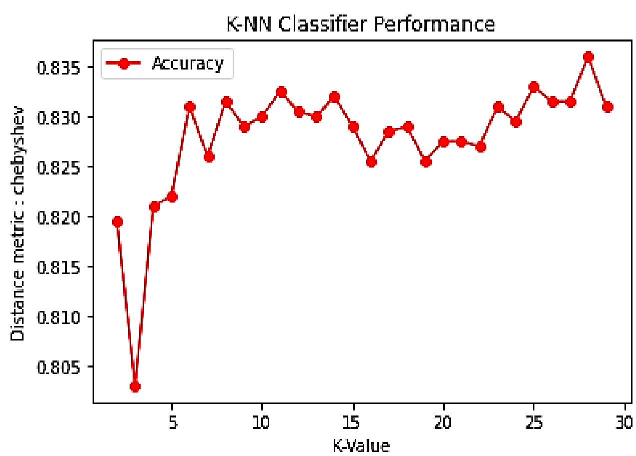
- The NB model is implemented with numerous data distributions such as multinomial, Bernoulli, and gaussian. The retrieved predictions from each of these data distributions are summarized in Table 5. The Gaussian distribution-based naive Bayes model has exhibited an enhanced prediction accuracy of 82.75% with an F1 score of 0.830.

Table 2. Hyper-parameter tuning and performance of SVM

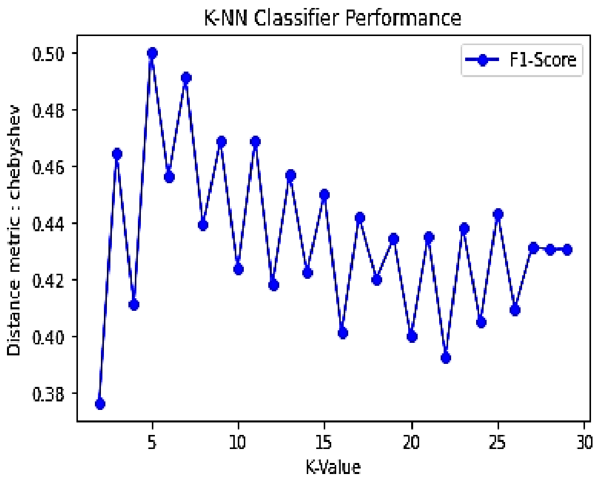
For Single Learner: Support Vector Machine		
Kernel Specification	**Accuracy (%)**	**F1-Score**
Sigmoid	67.80	0.678
Linear	79.75	0.797
Polynomial	84.55	0.845
Radial Basis Function (RBF)	85.90	0.859

Table 3. Hyper-parameter tuning and performance of KNN

For Single Learner: K Nearest Neighbour			
Distance Matric	**Best K value**	**Accuracy (%)**	**F1-Score**
Chebyshev	11	83.60	0.8360
Euclidean	17	83.95	0.8395
Manhattan	10	84.95	0.8495

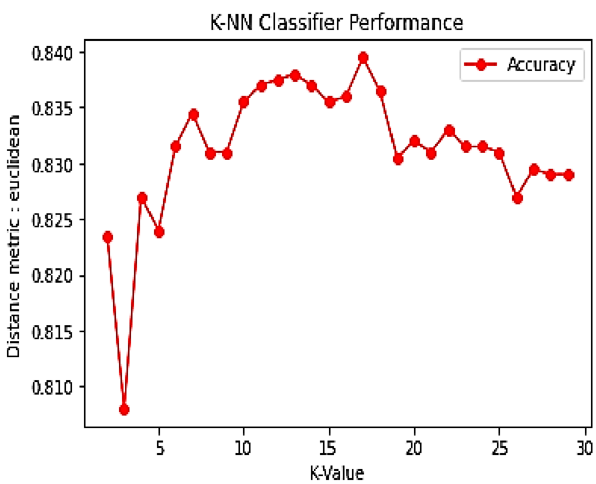


(a)

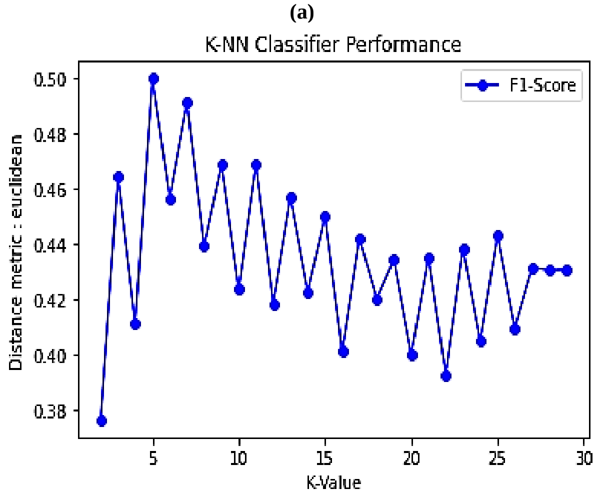


(b)

Fig. 3 K-NN classifier performance by using the distance metric 'Chebyshev' a) Accuracy, b) F1-Score

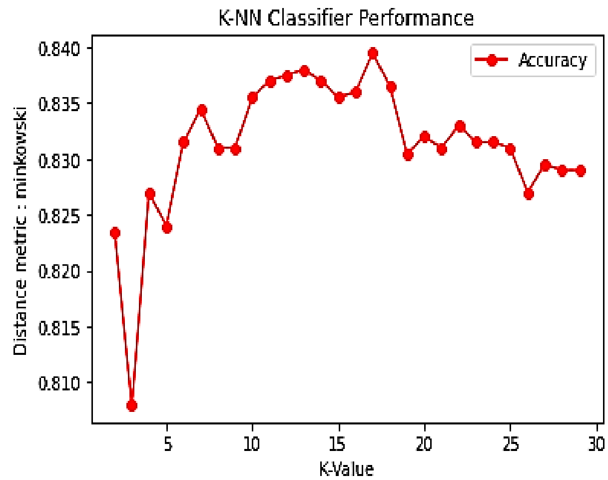


(a)

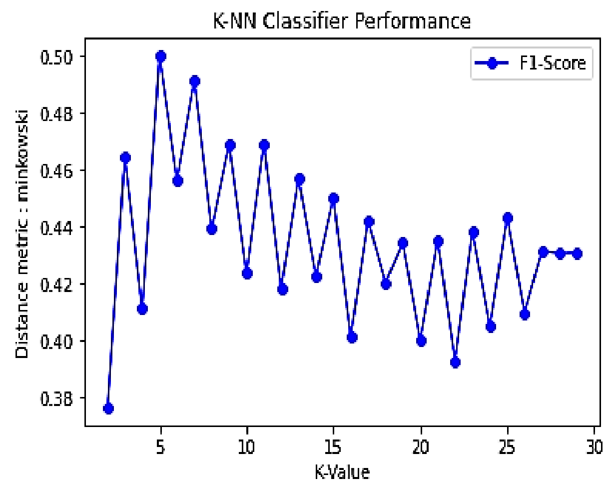


(b)

Fig. 4 K-NN classifier performance by using the distance metric 'Euclidean' a) Accuracy, b) F1-Score



(a)



(b)

Fig. 5 K-NN classifier performance by using the distance metric 'Minkowski' a) Accuracy, b) F1-Score

Table 4. Hyper-parameter tuning and performance of Decision Tree

For Single Learner: Decision Tree			
Criterion	**Splitter**	**Accuracy (%)**	**F1-Score**
Gini	Best	79.50	0.795
Entropy	Best	78.85	0.790
Gini	Random	78.10	0.781
Entropy	Random	79.75	0.797

Table 5. Different Types of Naive Bayes Classifiers and their Performances

For Single Learner: Naïve Bayes		
Model Name	**Accuracy (%)**	**F1-Score**
Multinomial	79.75	0.797
Bernoulli	79.75	0.797
Gaussian	82.75	0.830

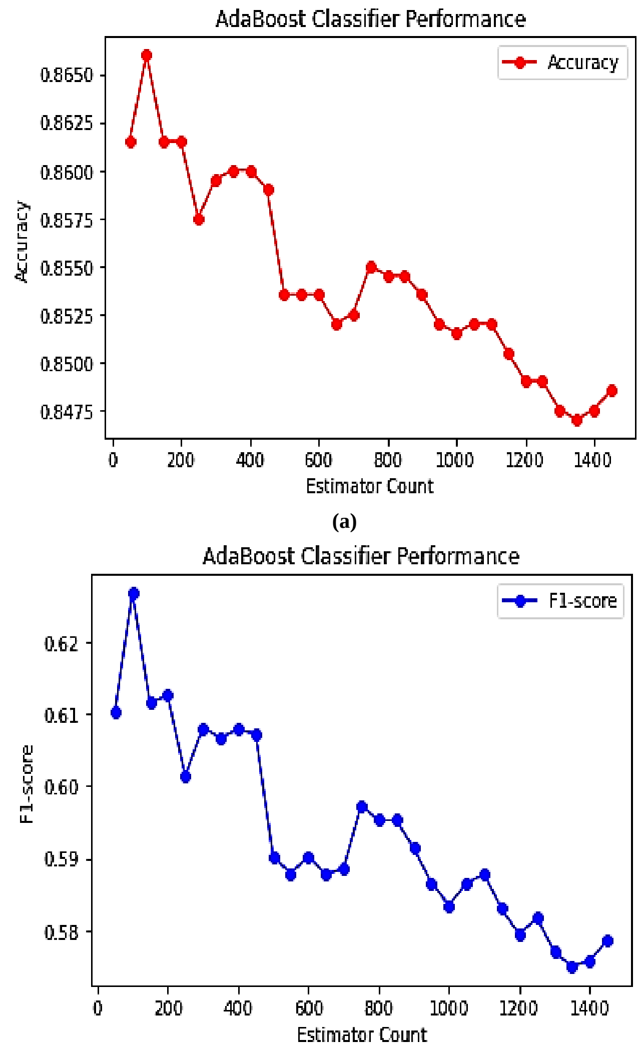
After conducting the hyper-parameter tuning process, each single learner model is recognized in terms of its best accuracy and f1-score. Table 6 has been utilized for inferring a better comparative study. This comparison declares that the SVM model wins the race over the other employed single learner models.

Table 6. Comparative Analysis Among All Single Learner models

Model Name	Best Hyper-parameter	Accuracy	F1-Score
SVM	Kernel = Radial Basis Function	85.9%	0.86
Decision Tree	'entropy' criterion and 'random' splitter	79.75%	0.797
Naive Bayes model	Gaussian	82.75%	0.830
K-NN	Distance metric 'Manhattan'; k-value =10	84.95%	0.8495

5.2. Phase 2: Ensemble Learners

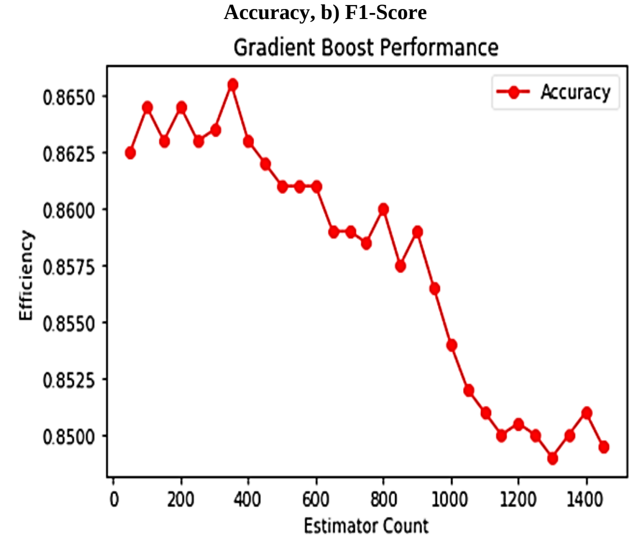
Random Forest, Adaboost, and Gradient Boost during this classification process are employed to infer the prediction. Each model undergoes a series of hyper-parameter selection procedures to favor improved decision-making. The models were implemented using the base estimator size between 50 and 1400 with a step size of 50. Figures 6, 7, and 8 display how the model performance varies when different estimator sizes are applied in AdaBoost, Random Forest, and Gradient Boost models. Table 7 summarizes the best possible predictive outcome for each employed ensemble model. The Random Forest, Adaboost, and Gradient Boost stick the promising result with the estimator count of 700, 350, and 100, respectively. The summarized result, as shown in Table 7, can comprehend that the AdaBoost model wins the race among the ensemble-based methods through accuracy of 86.6% and the F1 score of 0.630.



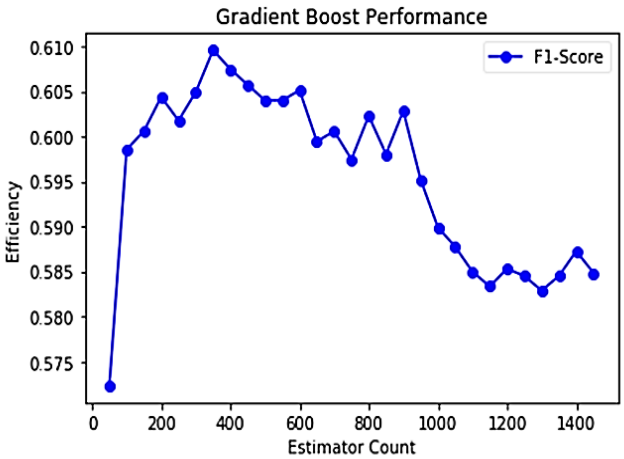
(a)

(b)

Fig. 6 Hyperparameter Tuning Process of AdaBoost Classifier a) Accuracy, b) F1-Score

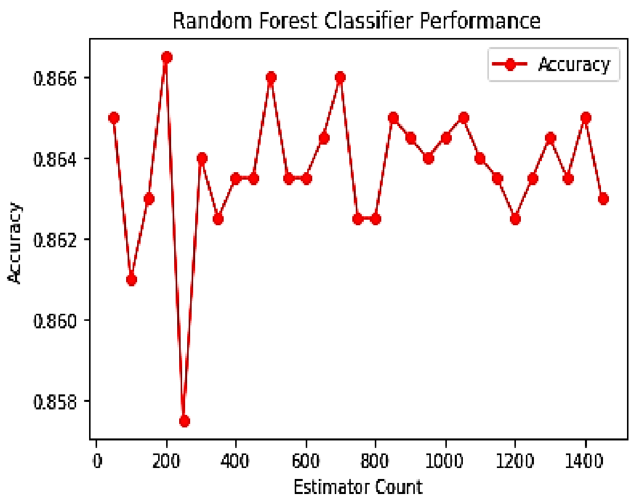


(a)

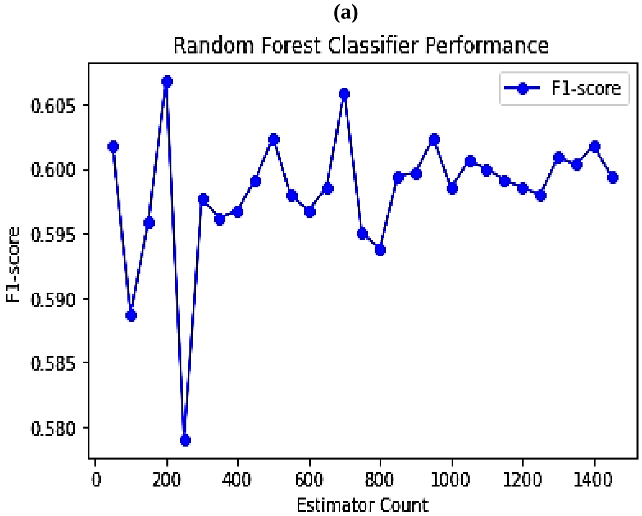


(b)

Fig. 7 Hyperparameter Tuning Process of Gradient Boost Classifier a) Accuracy, b) F1-Score



(a)



(b)

Fig. 8 Hyperparameter Tuning Process of Random Forest Classifier a) Accuracy, b) F1-Score

Table 7. Comparative Analysis among All Ensemble Learner models

Ensemble Learner			
Model Name	**Best Estimator Count**	**Accuracy (%)**	**F1-Score**
AdaBoost	100	86.60	0.630
Random Forest	700	85.70	0.610
Gradient Boost	350	85.55	0.610

5.3. Proposed Hybrid model

From phase 1, the SVM turns out to be the best single learner model, whereas the AdaBoost model is the best in the ensemble model's category. The proposed model combines the SVM and AdaBoost models to infer the prediction. A stacked ensemble model is constructed to perform the hybridization. The stacked ensemble requires two levels of models as its constituents. The level-0 models, also known as base models, receive the training data as input, and their predictions are assembled. The level-1 models, alternatively recognized as meta-models, receive the training data and the base models' predictions. To develop the training dataset required for the meta-model, a process of k-fold cross-validation is applied across the base models. Further, the out-of-fold forecasts accompany the meta-model's training dataset [24, 25].

In this study, the level-0 models are implemented using SVM and AdaBoost along with their enhanced predictions and optimized hyper-parameters. The meta-model is modelled by using the AdaBoost model. This model also receives a 15-fold cross-validation technique to model the learning dataset preparation for the meta learner. The detailed description of the model is outlined in Table 8. The hybrid model, i.e., the stacked SVM-AdaBoost model, can reach a predictive performance by exhibiting an accuracy of 87% and an f1-score of 0.87. Table 9 further summarizes the performance of the hybrid stacked SVM-AdaBoost model.

Table 8. Configuration of the Hybrid model

Level-0	1. SVM with Kernel =' Radial Basis Function.' 2. AdaBoost with base learner count=100
Level-1	AdaBoost with base learner count=100
Method Used	15-fold cross-validation

Table 9. Performance Analysis of Proposed Hybrid Model

Model Name	Accuracy	F1-score
Stacked SVM+AdaBoost	87%	0.87
SVM	85.9%	0.86
AdaBoost	86.60%	0.63

5.4. Discussion

An emphasis has been provided on applying the Hybrid model as a predictive tool for fulfilling the objective of this study. The fundamental motive of this study is to boost the efficiency of the predictive model. The hybridization of a single learner model, i.e., SVM, and an ensemble learner, i.e., Adaboost, has worked on the improvisation of both accuracy and f1-score. The reason for encouraging the hybridization of the top two models obtained from two dissimilar phases are outlined as follows-

- As described in Table 9, the SVM and Adaboost have reached 85.9% and 86.60%, respectively. However, the hybridization of these two models has increased the accuracy to 87%. Hence, a significant improvement has been discovered while comparing the efficiency of the hybrid approach with its constituent models.

- It is quite evident that accuracy does not consider the false positive and false negative parts during calculation. On the other hand, precision and recall are the metrics that take care of the false positive and false negative, respectively. Hence, using the f1-score can be highly recommended as it considers both precision and recall during its calculation.

- While comparing the f1-score of the constituent parts of the hybrid model, it can be observed that AdaBoost has given a relatively lower f1-score than the SVM model. Despite providing enhanced accuracy, the f1-score is 0.63, shown by AdaBoost, which cannot be recommended as a promising predictive model. The SVM model, however, reached a good f1-score of 0.86. From this scenario, it is necessary to exhibit a promising f1-score over SVM and AdaBoost. The hybridization has provided a substantially improved f1-score of 0.87 than that of AdaBoost (>0.63) and SVM (>0.86).

- The hybridization incorporates two dissimilar models, namely SVM and AdaBoost. SVM is a single learner which employs the best fit hyper-plane to segregate n-dimensional feature space into different classes. On the other hand, AdaBoost is an ensemble-based approach that uses multiple tree-based learners as its constituent. The approach employed by these two models is quite different from each other. SVM has provided enhanced efficiency than other employed single learner models in phase-1.

- Similarly, in phase-2, the AdaBoost is the top model among the other ensemble models. Both of these models outperform the other peer models in their classification process. It is why these two learners are stacked into a single entity (SVM-AdaBoost) to infer the best predictive analytics in terms of accuracy and f1-score.

- The hybrid approach is constituted by applying the k-fold cross-validation technique to overcome over-fitting. Here, necessary k-values should be tried to identify the best possible predictive modelling. The SVM-AdaBoost model is implemented using 5-fold, 10-fold, 15-fold, and 20-fold stratified cross-validation. The comparative investigation provided in Table 10 on different k-values indicates that k=15 reaches the highest accuracy and f1-score. Hence, the 15-fold cross-validation has proven to be the best-possible predictive model.

Table 10. K-fold Performance Analysis of Proposed Hybrid Model

K-fold Value	Accuracy	F1-score
5	86.71%	0.867
10	86.79%	0.868
15	87%	0.87
20	86.91%	0.869

The importance of different features responsible for retrieving the best predictive outcome should be analyzed. This analysis is depicted in Fig 9 as revealed by the hybrid stacked SVM-AdaBoost model. The relevant factors, such as the estimated salary, possession of credit cards, etc., become the evident parameters that interfere with the customer retention schemes. Hence, these features should be analyzed with a concentration on attracting customers.

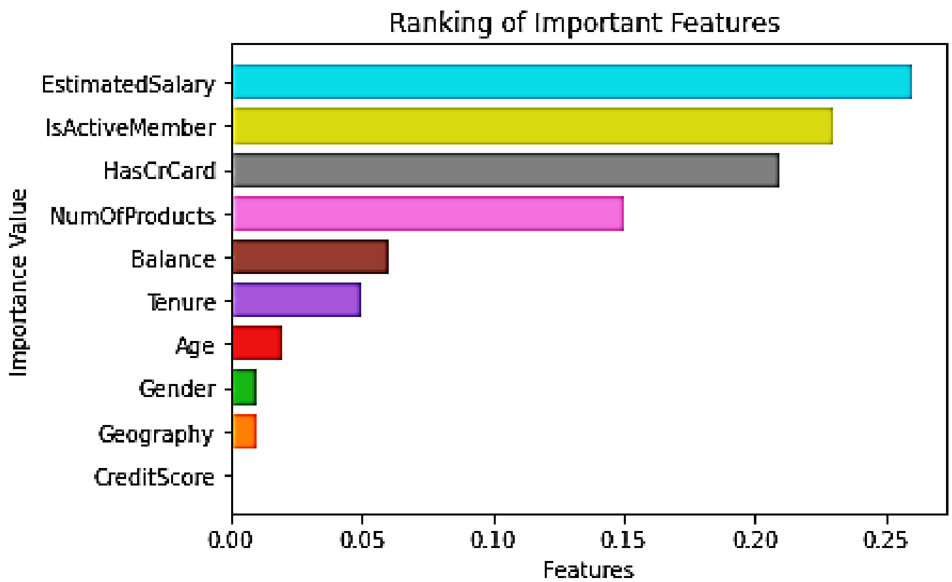


Fig. 9 Feature Importance Revealed by hybrid stacked SVM-AdaBoost model

6. Conclusion

The presented study is intended to assist companies in achieving enhanced profit. To provide this assistance, churn prediction seems to be an evident focus area. To improve customer retention, the company imposes mechanisms to recognize subscribers who are at risk. These at-risk consumers should be focused on, and extra effort should be encouraged. This situation can be handled by analyzing the company's past customer database using an automated process directed by Machine Learning based techniques. Both single and ensemble-based techniques are utilized in the customer database with proper hyper-parameter adjustment techniques. The top two models are selected from both types of techniques. These top two models are stacked to form the hybrid SVM-Adaboost model. The presented hybrid model also undergoes cross-validation to ensure benchmark prediction results. The presented automated system can segregate the customers having to leave with an efficiency of 87%. The stacked model provides informed decisions regarding the customer retention problem, which will direct the company to encompass suitable strategies to reform the customer base.

References

[1] Ahn J. H, Han S. P & Lee Y. S, Customer Churn Analysis: Churn Determinants and Mediation Effects of Partial Defection in the Korean Mobile Telecommunications Service Industry, Telecommunications Policy. 30(10-11) (2006) 552-568.

[2] Lee J, Lee J & Feick L, The Impact of Switching Costs on the Customer Satisfaction, Loyalty Link: Mobile Phone Service in France, Journal of Services Marketing. 15(1) (2001) 35-48.

[3] Badillo S, Bánfai B, Birzele F, Davydov I.I, Hutchinson L, Kam-Thong T, Siebourg-Polster J, Steiert B, Zhang J.D, An Introduction to Machine Learning, Clinical Pharmacology and Therapeutics. 107 (2020) 871 - 885.

[4] Jayakumar Sadhasivam, Arpit Rathore, Indrajit Bose, Soumya Bhattacharjee, Senthil Jayavel, A Survey of Machine Learning Algorithms International Journal of Engineering Trends and Technology. 68(4) (2020) 64-71.

[5] Rani K. S, Shaik T, Prasanna N.G.L, Vindhya R, Srilakshmi P, Analysis of Customer Churn Prediction in Telecom Industry Using Logistic Regression. International Journal of Innovative Research in Computer Science & Technology IJIRCST. 9(4) (2021). https://doi.org/10.21276/ijircst.2021.9.4.6.

[6] Lalwani P, Mishra M. K, Chadha J. S, Sethi P, Customer Churn Prediction System: A Machine Learning Approach Computing. 104(2) (2022) 271-294.

[7] Naidu G, Zuva T, Sibanda E.M, Systematic Review on Churn Prediction Systems in Telecommunications, In Bindhu V, Tavares J.M.R.S, Du KL, eds., Proceedings of Third International Conference on Communication, Computing and Electronics Systems, Lecture Notes in Electrical Engineering Springer, Singapore. 844 (2022). https://doi.org/10.1007/978-981-16-8862-1_64.

[8] Pustokhina I. V, Pustokhin D. A, Aswathy R. H, Jayasankar T, Jeyalakshmi C, Díaz V. G Shankar K, Dynamic Customer Churn Prediction Strategy for Business Intelligence Using Text Analytics with Evolutionary Optimization Algorithms, Information Processing & Management. 58(6) (2021) 102706.

[9] Kavita M, Sharma N & Aggarwal G, Churn Prediction of Customer in Telecommunication and E-Commerce Industry Using Machine Learning, PalArch's Journal of Archaeology of Egypt/Egyptology. 17(9) (2020) 136-145.

[10] Rahman M & Kumar V, Machine Learning Based Customer Churn Prediction in Banking, In 2020 4th International Conference on Electronics, Communication and Aerospace Technology, IEEE ICECA. (2020) 1196-1201.
[11] Kavitha V, Kumar G. H, Kumar S. M & Harish M, Churn Prediction of Customer in Telecom Industry Using Machine Learning Algorithms, Int. J. Eng. Res. Technol, IJERT. 9(5) (2020) 181-184.
[12] Abou el Kassem E, Hussein S. A, Abdelrahman A. M & Alsheref F. K, Customer Churn Prediction Model and Identifying Features to Increase Customer Retention based on User Generated Content, International Journal of Advanced Computer Science and Applications. 11(5) (2020) 522-527.
[13] Ahmad A. K, Jafar A & Aljoumaa K, Customer Churn Prediction in Telecom Using Machine Learning in Big Data Platform, Journal of Big Data. 6(1) (2019) 1-24.
[14] Saran Kumar A & Chandrakala D, A Survey on Customer Churn Prediction Using Machine Learning Techniques, International Journal of Computer Applications. 975 (2016) 8887.
[15] Hang C & Ma Y, Eds., Ensemble Machine Learning: Methods and Applications, Springer Science & Business Media. (2012) 1-34.
[16] Suthaharan S, Support Vector Machine. Machine Learning Models and Algorithms for Big Data Classification, Springer, Boston, MA. (2016) 207-235.
[17] Zhang Z, Introduction to Machine Learning: k-Nearest Neighbors, Annals of Translational Medicine. 4(11) (2016) 218.
[18] Berrar D, Bayes' Theorem and Naive Bayes Classifier, Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics. 403 (2018) 1-18.
[19] Priyam A, Abhijeeta G. R, Rathee A & Srivastava S, Comparatanalive Ysis of Decision Tree Classification Algorithms, International Journal of Current Engineering and Technology. 3(2) (2013) 334-337.
[20] Schapire R. E, Explaining Adaboost, Empirical Inference, Springer, Berlin, Heidelberg. (2013) 37-52.
[21] Natekin A & Knoll A, Gradient Boosting Machines, A Tutorial, Frontiers in Neurorobotics. 7 (2013) 21.
[22] Livingston F, Implementation of Breiman's Random Forest Machine Learning Algorithm, ECE591Q Machine Learning Journal Paper. (2005) 1-13.
[23] (2022). The Kaggle Website [Online] Available: https://www.kaggle.com/code/kmalit/bank-customer-churn-prediction/data
[24] Naimi A. I & Balzer L. B, Stacked Generalization: An Introduction to Super Learning, European journal of Epidemiology. 33(5) (2018) 459-464.
[25] Jyoti Goyal, Bal Kishan, TLHEL: Two Layer Heterogeneous Ensemble Learning for Prediction of Software Faults International Journal of Engineering Trends and Technology. 69(4) 2021 16-20.