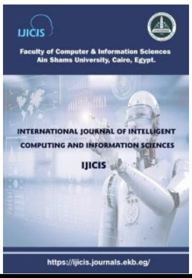
IJICIS, Vol.24, No.2, 28-40
DOI:10.21608/ijicis.2024.283191.1334

International Journal of Intelligent Computing and Information Sciences

https://ijicis.journals.ekb.eg/



ENHANCING MACHINE LEARNING ENGINEERING FOR PREDICTING YOUTH LOYALTY IN DIGITAL BANKING USING A HYBRID META-LEARNERS

Mohamed Galal*
Faculty of Computer and Information sciences,
Ain Shams University,
Cairo, Egypt
mhdgalal@yahoo.com

Sherine Rady
Faculty of Computer and Information sciences,
Ain Shams University,
Cairo, Egypt
srady@cis.asu.edu.eg

Mostafa Aref
Faculty of Computer and Information sciences,
Ain Shams University,
Cairo, Egypt
mostafa.aref@cis.asu.edu.eg

Received 2024-04-16; Revised 2024-04-16; Accepted 2024-06-25

Abstract: *Customer retention is a top priority for organizations due to its significant impact on corporate profitability. There is a lot of competition between banks to acquire and retain customers. The youth customer segment is the future of digital banks, and hence, this study was conducted to forecast the youth segment loyalty. This will help banks identify the degree of customer loyalty and the factors that affect their satisfaction. Customer churn may lead to a financial loss of revenue and market share. Therefore, forecasting customer loyalty has become essential to maintaining profitability and the customer base. Using Fintech (financial technology) and digital transformation techniques in digital banking works on enhancing the youth customers experience and increasing their lifetime value using machine learning techniques. This research presents a new model of stacking ensemble learning, which combines optimized base learner algorithms after applying hyperparameter tuning and the voting model to the stacking meta-learner algorithm. The research compares various base machine learning models, such as KNN (K-Nearest Neighbors), LR (Logistic Regression), RF (Random Forest), Adaboost, and GB (Gradient Boosting), for customer loyalty prediction. The experiment was generated using 10,000 banking customers, which contains 6,420 youth customers. The model assessment proved that using base learners combined with a voting mechanism as an input to stacking modeling received an accuracy of 88.9%. This research discusses challenges related to existing classification models, including mitigating biases and errors, preventing overfitting, addressing imbalanced data, enhancing model stability, improving interpretability, and automating model selection by using hybrid models for tuning.*

Keywords: *Ensemble modeling, Machine learning, Customer Loyalty, Digital banking, Stacking.*

1. Introduction

***Corresponding Author**: Mohamed Galal

Faculty of Computer and Information Science, Ain Shams University, Cairo, Egypt

Email address: mhdgalal@yahoo.com

Customer loyalty is the behavior of the customer, which indicates whether the customer will stop or continue to use the services provided by the service provider during a specific period of time. In recent years, the competition between financial institutions has increased due to the variety of digital services and financial products. So, customer churn has become a noticeable issue that needs to be handled. Any increase in the churn rate can decrease the profits [1] of the institution. The cost of acquiring new customers is more expensive than retaining existing ones. Therefore, to develop a robust machine learning model, service providers aim to define loyal and churn customers and develop proactive campaigns for retaining their customers. The complexity in the customers' behavior of digital banks increased the competition to satisfy customers' needs as the organizations aimed to increase the lifetime value of their customers. This could be done by enhancing the customer's loyalty through leveraging their experience and providing more tailored products and services that meet customers' expectations at each churn stage [2]. The main problem is that the youth segment can change their banks and switch to competitors who provide more adjustable conditions easily. Customer relationship management (CRM) aims to increase the lifetime value of customers using machine learning techniques [3]. Customers' needs in the digital ban/king domain differ based on customer segment. Thus, the CRM platform needs to predict the next activity of the customers to cope with the changes in customer behavior and enhance their loyalty by delivering the right action [4, 5].

This research is based on the fact that customers' maturity rates about digital experiences differ across age segments. The youth segment adopts the digital platform much faster than the older segment. The youth segment doesn't have face-to-face interactions with the bank, as they prefer to interact online and consume services by themselves. Bank experience with digital services can enhance the maturity of youth customers, resulting in their loyalty. In the coming years, winning customers will become very difficult. So it will be an achievement for the organizations that have a better relationship with the youth segment and are able to tailor services that provide better user satisfaction [6].

This research talks about the problems with the classification models that are currently used. These problems include fixing biases or errors in individual models, lowering overfitting and variance, dealing with imbalanced data, lowering bias and variation, making the final predictor easier to understand, and automating hyperparameter tuning. This research discusses enhancing prediction performance using a hybrid meta-learner machine learning model to develop customer loyalty and churn prediction. The new proposed framework combined supervised machine learning techniques as base learners with 3-step optimization steps. The first step is using hyperparameter tuning to enhance the model's accuracy; the second step is to use automated voting ensemble modeling; and the third step is to merge the base learners and the ensemble voting into the stacking modeling as a final estimator to get better accuracy.

This paper is structured as follows: Section II introduces the background and related works; Section III presents the proposed methodology; Section IV explains the results and discussion; and Section V summarizes the conclusion of the work.

2. Literature Review

Customer loyalty and the probability of his churn have become critical financial institution issues, such as in the banking industry [7]. The bank consultants are working on setting proactive models to forecast the propensity of customer loyalty before the decision to churn. The loyalty prediction models should be integrated with a smart retention engine to recommend some personalized services based on customer preferences [8]. Banking sectors invest in keeping the loyalty of their customers, as focusing on increasing the interactions from the current customers is cheaper than the cost of new customers' acquisition. Banks can proactively provide more tailored products and services to their customers to retain them and reduce the loss of funds [9].

Machine learning techniques are being used in ensemble methods to more accurately tackle classification or regression issues. Instead of utilizing only one model, ensemble learning mixes several models to increase the efficiency of the outcomes. Bagging, boosting, and stacking procedures are ensemble modeling techniques. Bagging is an ensemble approach that seeks a diverse group of ensemble members by varying the training data and creating weak learners using a subset of the dataset. The bagging approach works on improving the model's accuracy and supports enhancing the variance error as it has less variance than single models. The boosting approach fits a random data sample into the model and trains it sequentially. The boosting approach changes the weight for the wrong classification until it reaches a better accuracy. The stacking mechanism supports providing an improved prediction compared to using a single model, as it combines predictions from base models with meta-models to generate a final model with better predictions [10].

This section discusses the past research and presents some technical findings. The research [11] worked on optimizing the accuracy of predicting customer churn in banking by using the voting classifier ensemble technique. The accuracy of using the voting ensemble approach in this research is 85%. The research [12] applied ensemble Bagging, RandomForest, GradientBoosting, and ExtraTree models to 602 young bank customers to predict the youth customers' defects. The results of this study proved that the ExtraTree Classifier gets the best AUC results. The research [13] used a meta-classifier algorithm to predict customer churn using Randomforest (RF), Logistic Regression (LR), DecisionTree (DT), and XGboost (XGB). The results proved that merging different machine learning models in a meta-classifier leads to better classification results. The research [14] improved churn prediction accuracy in banking using the stacking ensemble method. The stacking got 83% prediction accuracy, but it took 20.25 seconds in the training time.

ABC Multistate Bank uses machine learning techniques to forecast customer attrition. The study applied six models: K-Neighbors, Support Vector Machine, Naïve Bayes, Decision Tree, Random Forest, and Extreme Gradient Boosting (XGBoost). XGBoost achieved the highest performance with an accuracy rate of 84.76%, F1 score of 56.95%, and ROC curve graph of 71.64%. Future studies should examine various machine learning techniques [15]. The research [16] discussed a churn prediction system in telecommunication operators to identify prospective churn customers. Powell's optimization technique is used to enhance the ensemble learning classification model. The model achieved an accuracy score of 84% and an F1 score of 83.42%.

The research [17] conducted a comparison between four machine learning techniques to forecast customer turnover in e-commerce, including neural networks, support vector machine, Naïve Bayes and random forest, and Adam deep learning model. The random forest technique achieved the highest prediction accuracy. Future enhancements include incorporating more features and evaluation metrics and using various datasets in model training. The research [18] compared traditional and ensemble machine learning algorithms on a telecom company dataset. The results showed that the Stacking and Voting algorithm achieves the best prediction of churn customers. Ensemble learners outperform single-base learners, and a balanced training dataset is expected to improve classifier performance. Future research could focus on developing models that can effectively incorporate these data sources. The research [19] examined multiple machine learning techniques, including Logistic Regression, Decision Tree, K-Nearest Neighbors, Gaussian Naive Bayes, Support Vector Machine, AdaBoost, Random Forest, Extra Trees, Gradient Boosting, and deep neural networks, to forecast individuals who are likely to default on bank loans. The ensemble voting model achieved an accuracy of 87.26%, hence improving the efficiency of bank loan approval procedures.

3. Methodology

The content of this section is to discuss the design of customer loyalty forecasting model using the set of Mixed ensemble modeling Metal-Learners as shown in Figure. 1. The dataset contains 6,420 youth customer's data which extracted from 10,000 digital banking consumers from open-source data on Kaggle website, and the target label detects whether the consumer is churner or still loyal to the bank [20]. The dataset structured from 13 fields (features). The data of the experiment were loaded from banking database that contains transactions and customer data models. The description of these fields is shown in Table 1. The loyalty proposed framework in Figure. 1 will be discussed below:

A. Load churn data

- Using pandas in Jupyter notebook to read the youth segment data from Banking data mart.

B. Data manipulation

In this step the data is handled to match the modeling scope as the followings:

- Dropping the unrelated columns ("customer_id", "row_number", "surname"),
- Dividing the data scope into Loyal dataset and churn dataset.

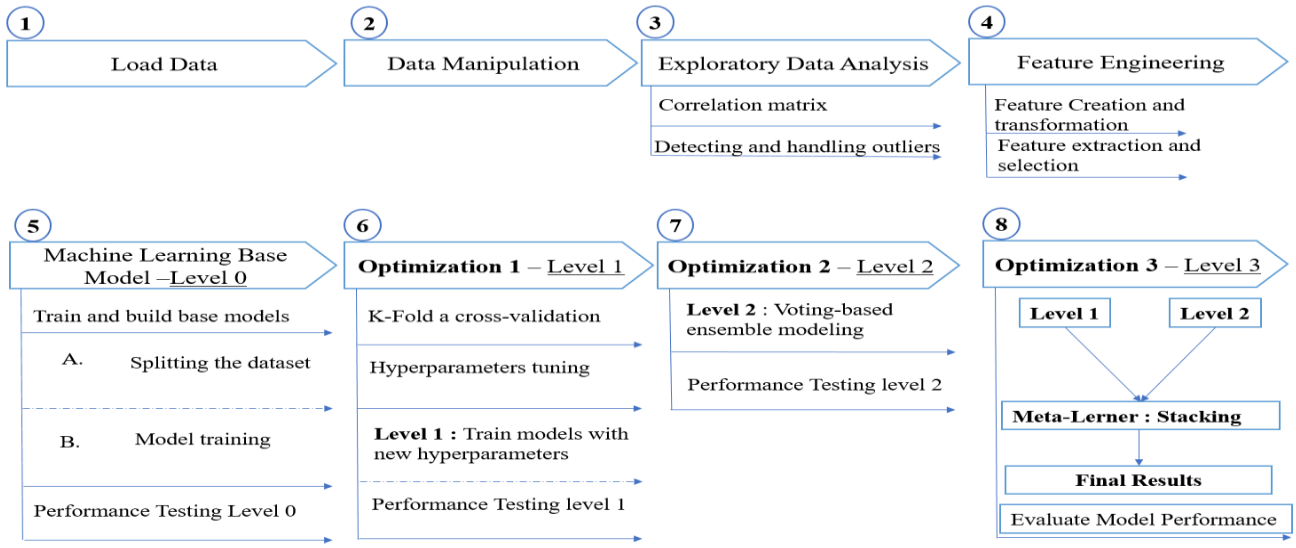


Figure. 1: Loyalty prediction using Mixed Meta-Learner framework

Table 1 Features description

Feature Name	Feature Description
Row number	Unique identifier
Customer Id	Unique identity for each customer.
Surname	Customer’s Last name.
Credit Score	Customer’s credit scoring.
Geography	Defines the residence country.
Gender	Defines The gender of the customer.
Age	Defines customer’s Age.
Tenure	The period in years when the customer start his first account with the bank
Balance	All balances related to customer accounts.
Num of Products	The exiting bank products belongs to the customer
Has Cr Card	A flag that defines if the customer has a credit card or not
Is Active Member	A flag that defines the activity of the customer from a perspective of making a transactions.
Estimated Salary	Defines the customer’s estimated salary.
Exited	A flag that defines if the customer is loyal or churned.

C. Exploratory Data Analysis (EDA)

The EDA step is done to understand the data using visualization functions. The EDA facilitates the process of discovering more patterns and getting more insights as the following:

1) Perform statistical analysis: In this step, there are some ratios will be performed to study the dataset and to get a deeper understanding of the scope as shown in Table 2.

Table 2 Features Statistics

Feature	Mean	Std	Min	Max
creditscore	651.33	97.00	350.00	850.00
age	32.81	5.03	18.00	40.00
tenure	5.03	2.87	-	10.00
balance	74,852.27	62,852.80	-	250,898.09
numofproducts	1.54	0.54	1.00	4.00
hascrcard	0.71	0.45	-	1.00
isactivemember	0.50	0.50	-	1.00
estimatedsalary	99,889.92	57,587.53	90.07	199,953.33

2) Visual dataset analysis: This step provides a graphical representation of the data analysis for different features. As shown in Figure. 2, the loyal and churned customer’s data distribution suffers from class-imbalance which can lead to bias in model performance. The over sampling technique like SMOTE (Synthetic Minority Oversampling Technique) is used to overcome the issue of class-imbalance. One of the useful insights shows that churned customers who are living in France have a high churn rate with 39.5%, followed by Germany with 39.4% and Spain with 21.1%. For loyal customers France comes first with 53.7%, Spain with 25.3% and Germany with 21% as shown in

Figure. 3. Females have 57.1% of churners and 42.7% of loyal customers. Number of products owned by the customers can detect customer loyalty as the churn rate of customers who own one only one product is high as they represent 67.9%, followed by those who own two products with 19%, three products with 10.8%, and four products with 2.3%. For loyal customers, customers who own two products represent 53.7%, one product is 45.7%, and three products are 0.59%.

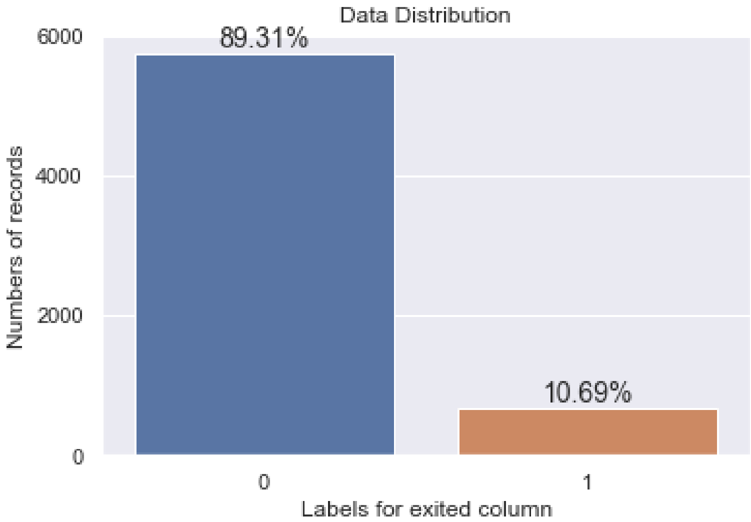


Figure. 2: Exited feature distribution

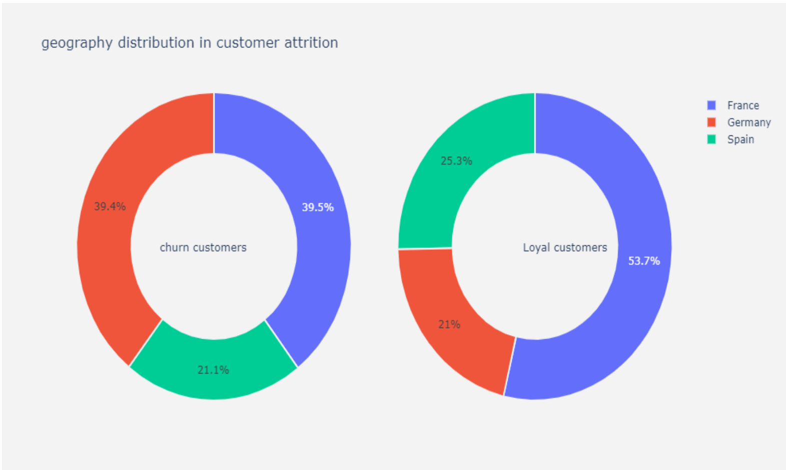


Figure. 3: Loyal customers and Churners per geography

3) Explore the correlation matrix:

The heatmap graph show the correlation degree between each feature as shown in Figure. 4. The attributes are very high correlated with itself as it contains the value 1.0 in their intersected cells. Number of products attribute are negatively correlated with Balance attribute as it contains -0.4 correlation degree.

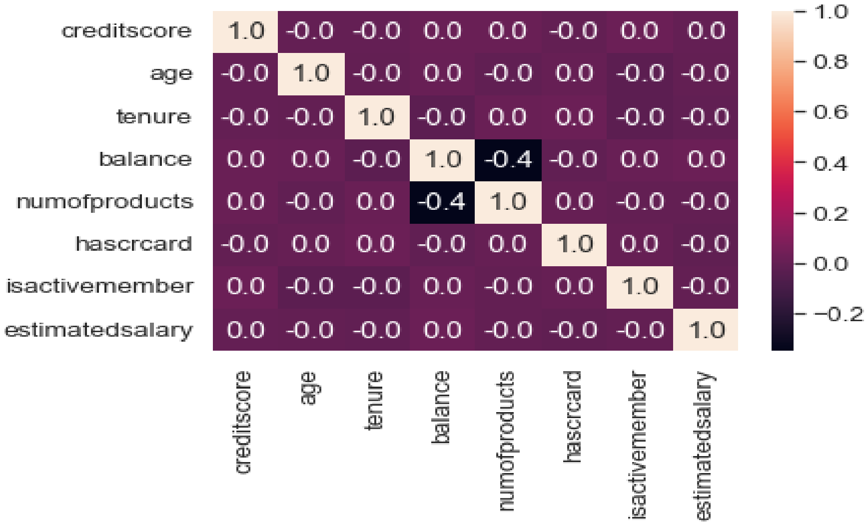


Figure. 4: Heatmap for the correlation matrix

4) Detecting and handling outliers

Outliers are values that are exceeds the normal maximum or minimum values. They lead to getting a bad performance of the model. Therefore, the output can be skewed based on the outlier values. In our data there is no outliers detect as the example shown in Figure. 5. In case of finding an outlier values, a mini algorithm could be used to handle outlier findings by replacing the outlier value by a median value based on a predefined threshold.

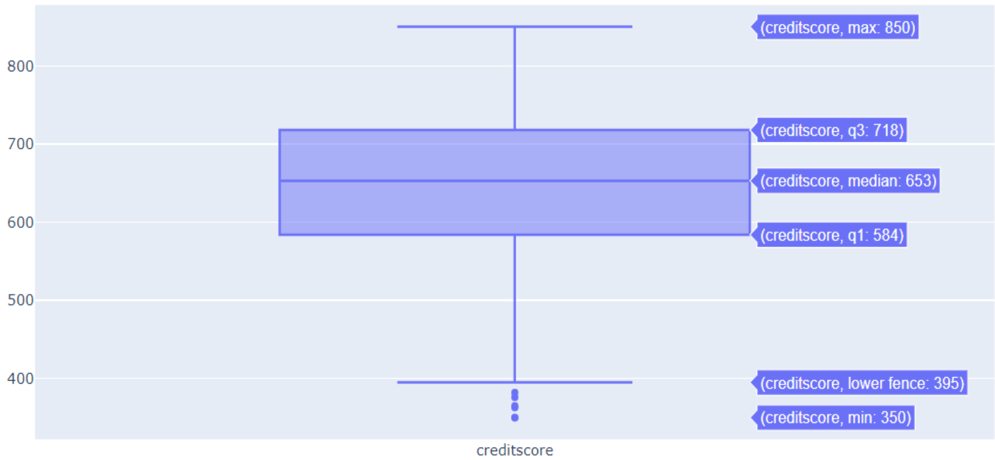


Figure. 5: boxplot for age and tenure to check the outliers

D. Feature engineering

Feature engineering consists of creation, transformation, extraction, and selection of features. The raw data is prepared in this the step of so that it is suitable for further processing and analysis as the following:

1) Feature Creation and transformation

In this step the one hot encoding technique is used to transform categorical variables into a new derived numerical features to be fitted to the algorithm. In our model the “geography” feature will be transformed into the following features (geography_Spain, geography_France, geography_Germany), and “gender” feature will derive the following features (gender_Male and gender_Female).

2) Feature extraction and selection

In this step the feature selection is done by selecting the suitable features for the model and preparing the features in a way that is suitable for the machine learning model. The important features will be extracted and selected to develop the predictive model. To find the importance of each feature, the Random Forest classifier is used to train the model as it indicates the importance of each feature based on the tree decision in easy representation as shown in Table 3. Feature importance technique generates a score for each feature and indicates how they are useful to the prediction model as the higher score of the importance means that this feature will have a more effect on the predictive model.

Table 3 Features importance

Rank	Feature	Importance
1	creditscore	0.192078
2	age	0.186287
3	tenure	0.176259
4	balance	0.118872
5	numofproducts	0.117698
6	hascrcard	0.104934
7	isactivemember	0.023177
8	estimatedsalary	0.022668
9	geography_France	0.017151
10	geography_Germany	0.011083
11	geography_Spain	0.010246
12	gender_Female	0.009776
13	gender_Male	0.009772

E. Selecting the machine learning algorithms

This section implemented more than a machine learning classification model to evaluate and choose the most successful model with the best accuracy:

1) Build base models:

1.1 Data splitting: Data splitting is when data is divided into two or more subsets. With a two-subset split, one subset is used to train the model and the other allows the process of testing the unseen data. The training data used for the model training and it contains 75% from all dataset. The model testing is used to test the accuracy of the trained model using the unseen 25% of the dataset.

1.2 Model Training: In this work there are more than classification algorithm used to reach better predicting accuracy. The selected machine learning models in this work are Random Forest, Logistic Regression, KNN (k-Nearest Neighbors), AdaBoost and GradientBoosting model. This models are base learners that contains a set of parameters need to be tuned to reach the best accuracy.

2) Testing the performance base models

2.1 Predictive model accuracy is the use of some metrics that measures the true output from the predictive model in relation to the whole predictions. In this exercise the testing data is used to validate the model performance. The final findings identify the healthiness of the models used in the training phase and how they perform in the validation phase as shown in Table 4.

2.2 Model performance evaluation in this research is done by using the mean accuracy metric and ROC_AUC (Area Under the Receiver Operating Characteristic Curve). The ROC-AUC will assess the base classifiers on the testing dataset that represents 25% from all dataset as shown in Table 4 and Figure. 6. The ROC-AUC considers the score of all possible values used in generating the final scoring. The ROC_AUC is more significant as it doesn't consider only a single threshold as the mean accuracy scoring. The GradientBoosting model gets 86.63% ROC_AUC and 90.09% accuracy which is proves the best prediction accuracy. The AdaBoost model is considered the second ranked in scoring model as it gets 84.62% ROC-AUC score and 89.71% accuracy.

Table 4 Model accuracy

Model	Accuracy	ROC_AUC
KNN	0.879751	0.540706
Logistic Regression	0.890343	0.679673
AdaBoost	0.897196	0.846256
GradientBoosting	0.900935	0.866295
Random Forest	0.890343	0.821835

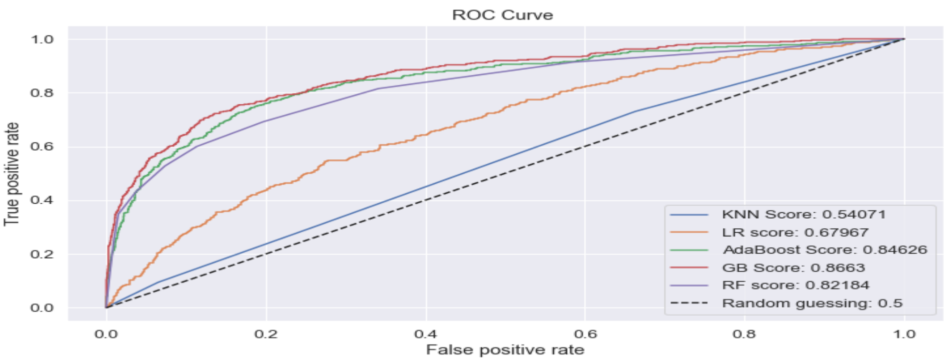


Figure. 6: ROC curve for base models

F. Model optimization

In this research, there are 3 levels of enhancement to enrich the model to boost the prediction accuracy as the following:

Optimizer 1 : Apply K-fold Cross-validation and tuning

This enhancement will be implemented through 2 steps as the following:

1) Apply K-fold cross-validation method

Cross-validation is a statistical method that is used to enhance the performance evaluation of machine learning models on the unseen data. It is used to solve the issue of overfitting in predictive models, especially in the limited dataset use cases. The cross-validation approach partitions the data into subsets (folds) and validates the data using each fold and train the model with the remaining folds. The enhancement done by using K-folds cross validation approach and considering the k variable = 5. The output show that AdaBoost algorithm gets 84.9% ROC_AUC score and Gradient Boosting algorithm gets 86.8%. From the results, we can proof that there is a minor modification done by applying the cross-validation mechanism.

2) Apply hyperparameters tuning

Hyperparameter tuning allows tweaking model performance for getting more accurate results by choosing suitable hyperparameter values. RandomizedSearchCV is a tuning method that used to boost the model performance by randomly passing the group of hyperparameters and produces the best set of hyperparameters which generates the best score in a faster process than grid search. After applying RandomizedSearchCV on GradientBoosting it gets 87.2% ROC-AUC score and the AdaBoost gets 85.7% ROC-AUC score. Applying K-fold cross-validation and RandomizedSearchCV is considered a first level of models optimization.

Optimization 2 : Apply Voting Mechanism

A Voting Classifier is a machine learning model that trains on an ensemble of multiple models and predicts an output based on their highest probability of chosen class as the output. The voting mechanism uses soft mode in this research to consider the weighted and averaged probabilities resulted from each model as shown in Figure. 7. The class with the highest weighted and averaged probability is considered the winning class. In this research voting classifier implemented on the test dataset to enhance the model classification performance by applying two steps as the following:

2.1 Feature transformation is implemented on the training dataset to transform the skewed data to normality using standard scaler/log transformation method.

2.2 The voting-based ensemble meta-classifier used to combine different machine learning classifiers for classification through an automated majority voting. In this research, the aggregation for voting is applied on AdaBoost and GradientBoosting to achieve better accuracy. The voting ensemble modeling achieved 87.72% ROC-AUC score and 91.2% accuracy. The optimized voting meta-classifier is considered as a second level of model optimization.

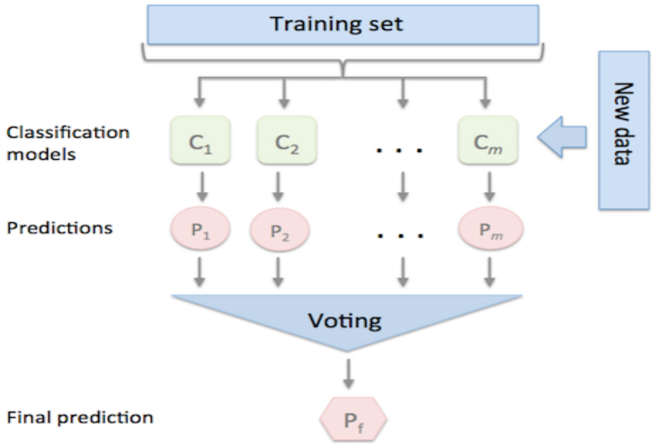


Figure. 7: Voting Classifier [21]

Optimization 3 : Stacking Meta-learner

Stacking is an ensemble learning method that combines multiple machine learning models through meta learning to build a new model and improve model performance. The Meta-learners are trying to reduce the weakness and maximize the power of individual base models. In this research, the stacking based on logistic regression classifier has considered the level-1 optimization (The modified AdaBoost and GradientBoosting) and the level-2 optimization (Voting mechanism) as base learners. The final generated classification model (final estimator) reached 88.9% ROC_AUC score and 91.5% accuracy, which proofs the highest prediction scoring as shown in Table 5.

Table 5 Optimized Models ROC_AUC score

Model	Accuracy	ROC_AUC
AdaBoost	90.2%	85.7 %
GradientBoosting	90.3%	87.2 %
Voting Classifier	91.2%	87.72%
Stacking	91.5%	88.9 %

4. Results and Discussion

This research applied hybrid meta-learner ensemble techniques (bagging, boosting, and stacking) to enhance the accuracy of predicting youth segment loyalty and churn in digital banking. The first step considered applying five machine learning classification models as base learners: Logistic Regression, KNN, Random Forest, AdaBoost, and Gradient Boosting classifiers. The second step applied three levels of optimization. The first optimization (Level 1) used K-fold cross-validation to handle the overfitting by increasing the training data, and it also used RandomizedSearchCV to improve the tuning of hyperparameters. The second optimization (Level 2) used automated voting classifier ensemble learning, which satisfied a better prediction using accuracy and the ROC_AUC score. The third optimization (Level 3) considered levels 1 and 2 as inputs to the meta-learner stacking mechanism.

The results show that using hybrid ensemble modeling approaches gets better performance than using base learners only. The contribution of using three optimization levels reached better performance in prediction and reduced the bias and variance in the final forecasting. The automation capability of the voting classifier and the optimized meta-learner classifiers used as inputs for the stacking modeling. The ROC_AUC score of the new hybrid framework is 88.9% and the accuracy score is 91.5%, which is better than Logistic Regression, Random Forest, KNN, AdaBoost, Gradient Boosting classifiers, and voting ensemble modeling.

5. Conclusion

Retaining customers is one of the top concerns of any institution, as they add a lot of value to the business. The youth customers segment are the long-standing customers of banks. So, this research was developed to predict their loyalty. The complex behavior of the youth segment required a new, enhanced machine learning model to help decision-makers retain their customers. The market share of any bank could be decreased by the increase in customer churn rate, which leads to a decrease in profitability and the extra cost of getting new customers. To increase the loyalty and satisfaction of customers, banks need to proactively forecast the propensity for customer churn. This study focused on building a loyalty prediction framework for the youth segment in digital banking to minimize customer churn. This research applied a new framework that contains a sequence of hybrid meta-learners and ensemble machine learning algorithms (Bagging, boosting, and stacking). The framework applied boosting to enhance the accuracy of churn prediction and to solve the issue of low performance resulting from using weak learners. The study proved that applying the mixed approach using three levels of optimization increased accuracy, enhanced the overfitting compared to individual models, reduced the bias and variance, and reached a robust framework. The combination of the ensemble voting classifier and the enhanced boosting classifiers as inputs for the ensemble stacking model (final estimator) achieved higher prediction accuracy results than using weak learners. The hybrid stacking meta-learners achieved an 88.9% ROC_AUC score and 91.5% accuracy, which demonstrated better prediction performance.

References

1. Rouhani S, Mohammadi A. A Novel Hybrid Forecasting Approach for Customers Churn in Banking Industry. Journal of Information & Knowledge Management. 2023 Oct 19;22(05):2250089.
2. A. Chorianopoulos, Effective CRM using predictive analytics. Chichester, West Sussex: John Wiley and Sons Ltd, 2015.
3. Kimura T. CUSTOMER CHURN PREDICTION WITH HYBRID RESAMPLING AND ENSEMBLE LEARNING. Journal of Management Information & Decision Sciences. 2022 Feb 1;25(1).
4. Guliyev H, Tatoğlu FY. Customer churn analysis in banking sector: Evidence from explainable machine learning models. Journal Of Applied Microeconometrics. 2021 Dec 29;1(2):85-99.
5. Singh PP, Anik FI, Senapati R, Sinha A, Sakib N, Hossain E. Investigating customer churn in banking: A machine learning approach and visualization app for data science and management. Data Science and Management. 2024 Mar 1;7(1):7-16.
6. Gabbi, G.; Giammarino, M.; Matthias, M.; Monferrà, S.; Sampagnaro, G. Does face-to-face contact matter? Evidence on loan pricing. Eur. J. Financ. 2019, 26, 820–836.
7. Muneer A, Ali RF, Alghamdi A, Taib SM, Almaghthawi A, Ghaleb EA. Predicting customers churning in banking industry: A machine learning approach. Indonesian Journal of Electrical Engineering and Computer Science. 2022 Apr;26(1):539-49.

8. Brownlee J., Ensemble learning algorithms with Python: Make better predictions with bagging, boosting, and stacking. Machine Learning Mastery, 2021.
9. Patil K, Patil S, Danve R, Patil R. Machine Learning and Neural Network Models for Customer Churn Prediction in Banking and Telecom Sectors. InProceedings of Second International Conference on Advances in Computer Engineering and Communication Systems: ICACECS 2021 2022 Feb 22 (pp. 241-253). Singapore: Springer Nature Singapore.
10. Brownlee J., Ensemble learning algorithms with Python: Make better predictions with bagging, boosting, and stacking. Machine Learning Mastery, 2021.
11. Latheef, J. and Vineetha, S., Exploring Data Visualization to Analyze and Predict Customer Loyalty in Banking Sector with Ensemble Learning, International Journal of Innovative Research in Applied Sciences and Engineering (IJIRASE), Volume 4, Issue 9, 2021.
12. V. Bharathi S, D. Pramod, and R. Raman, An Ensemble Model for Predicting Retail Banking Churn in the Youth Segment of Customers, Data, vol. 7, no. 5, p. 61, 2022.
13. F. Ehsani, Customer churn prediction from Internet banking transactions data using an ensemble meta-classifier algorithm, 2022.
14. O. M. Oladimeji, A. R. Ajiboye, and F. E. Usman-Hamza, An optimized stacking ensemble technique for creating prediction model of customer retention pattern in the banking sector, Gadau Journal of Pure and Allied Sciences, vol. 2, no. 1, pp. 22–29, 2023.
15. Hui SH, Khai WK, XinYing C, Wai PW. Prediction of customer churn for ABC Multistate Bank using machine learning algorithms/Hui Shan Hon [et al.]. Malaysian Journal of Computing (MJoC). 2023;8(2):1602-19.
16. Khoh WH, Pang YH, Ooi SY, Wang LY, Poh QW. Predictive churn modeling for sustainable business in the telecommunication industry: optimized weighted ensemble machine learning. Sustainability. 2023 May 25;15(11):8631.
17. Baghla S, Gupta G. Performance evaluation of various classification techniques for customer churn prediction in e-commerce. Microprocessors and Microsystems. 2022 Oct 1;94:104680.
18. Kumar S, Logofatu D. Comparative Study on Customer Churn Prediction by Using Machine Learning Techniques. InAsian Conference on Intelligent Information and Database Systems 2023 Jul 24 (pp. 339-351). Cham: Springer Nature Switzerland.
19. Uddin N, Ahamed MK, Uddin MA, Islam MM, Talukder MA, Aryal S. An ensemble machine learning based bank loan approval predictions system with a smart application. International Journal of Cognitive Computing in Engineering. 2023 Jun 1;4:327-39.
20. Bank Customer Churn Dataset. https://www.kaggle.com/datasets/gauravtopre/bank-customer-churn-dataset (accessed Mar. 15, 2024).
21. EnsembleVoteClassifier: A majority voting classifier - mlxtend, rasbt.github.io. https://rasbt.github.io/mlxtend/user_guide/classifier/EnsembleVoteClassifier (accessed Mar. 15, 2024).