International Journal of Advanced and Applied Sciences, 9(7) 2022, Pages: 139-149

Contents lists available at Science-Gate

International Journal of Advanced and Applied Sciences

Journal homepage: http://www.science-gate.com/IJAAS.html

Building a proper churn prediction model for Vietnam's mobile banking service

Nguyen Thi Ha Thanh [1], Nguyen Thao Vy [2,*]

[1]*Faculty of Banking and Finance, Foreign Trade University, Hanoi, Vietnam*
[2]*MB Shinsei Finance Limited Liability Company, Hanoi, Vietnam*

ARTICLE INFO

Article history:
Received 14 January 2022
Received in revised form
15 April 2022
Accepted 21 April 2022

Keywords:
Customer churn
Prediction model
Mobile banking
Data mining
Machine learning

ABSTRACT

This study aims to build a model predicting the churn rate of customers using mobile banking services in Vietnam by applying data mining techniques. Customer churn is an issue that any service provider must pay attention to because it is decisive to the development of the business. The competition between banks is getting tougher, hence customer churn prediction has become of great concern to banking service companies. It is necessary for banks to collect colossal data and establish a valued model for classifying types of customers. In this study, three supervised statistical learning methods which are KNN, Random Forest, and Gradient Boosting are applied to the churn prediction model using the data source of VIB's customers. In addition to selecting models belonging to the group of weak single learners such as Neural Networks, Naïve Bayes Classifier, and K-nearest Neighbor..., this paper utilizes Random Forest and Gradient Boosting which are assessed as better models because they can combine weak learners for improving model efficiency and capable of classification. The results exhibited that Gradient Boosting is the best performance in the three above classifiers with a 79.71% of accuracy rate, and 86.23% of ROC (Receiver Operating Characteristic) curve graph. Moreover, the decision tree algorithm generates readable rules for churner and non-churner classification which are potentially helpful to managers. Finally, this study suggests a proper model that can be used to forecast churners of mobile banking services in Vietnam.

© 2022 The Authors. Published by IASE. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Big data is proven to be a really beneficial resource for organizations. Big data is about not only just gathering enormous volumes of data but also understanding and analyzing the data in order to extract useful information from it. It is difficult for a firm to comprehend its consumers if it does not monitor their behavior over time, or if it does not assess their requirements and preferences while developing new goods and the target consumer group's behavior, for instance. Big data solutions that use machine learning techniques (Machine Learning) have resulted in significant advances in business data analysis and management and are used in a variety of sectors. Organizations do not have to wait long to evaluate historical data, since most Big data initiatives are implemented in order to make rapid choices. Many areas have seen success using data analysis and data mining approaches. A well-known example of an early cancer detection model in medical, construction, credit scoring in the financial sector, customer categorization, recommendation systems based on consumer demands, and so on.

From there, the bank may view the big picture, such as the overall development of the bank, or it can assess particular and granular features, such as the behavior of each client. Banks must employ analytical approaches or machine learning models to understand consumer elements before implementing the most suitable and timely policy.

Customer retention is one of the bank's most essential tactics for long-term viability. This is even more necessary for the Vietnamese banking system which is undergoing a strong transformation toward digital services (Phan et al., 2019). Businesses must understand how loyalty-building processes are used to forecast consumers' desire to depart and categorize customers in order to proactively implement the fairest and most successful rules

* Corresponding Author.
Email Address: vynt.ho@mcredit.com.vn (N. T. Vy)
https://doi.org/10.21833/ijaas.2022.07.014
Corresponding author's ORCID profile:
https://orcid.org/0000-0002-5007-8821
2313-626X/© 2022 The Authors. Published by IASE.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

(Mahmoud et al., 2019). If a company's primary goal is to grow by acquiring new consumers, this technique is less expensive. Because the cost of acquiring new consumers is 5 to 6 times that of maintaining existing customers (Athanassopoulos, 2000; Bhattacharya, 1998; Colgate and Danaher, 2000; Rasmusson, 1999), loyal customers are more profitable and less susceptible to competitive and manipulative activities from competitors; additionally, they may refer friends and relatives to the company's products and services, whereas new customers require time to experience and feel the values that the company brings (Colgate et al., 1996; Ganesh et al., 2000; Mizerski, 1982; Paulin et al., 1998; Reichheld, 1996; Stum and Thiry, 1991; Zeithaml et al., 1996). As a result, effectively encouraging them to stay will result in large revenues for enterprises, especially banks (Poel and Lariviere, 2004).

Increasing the number of consumers who stay implies decreasing the number of customers who leave. Customer churn refers to the migration of clients from one bank to another, and it occurs when a customer shuts all of his accounts and ceases doing business with the bank (Kaur et al., 2013). Banks evaluate customer relationships based on research on customer turnover rates. Accurately forecasting the likelihood of a client leaving and taking steps to persuade that customer to stay may greatly enhance a company's income. In reality, every firm must deal with the problem of consumers leaving, thus the study of departing customers is a critical issue for banks.

2. Literature review

Some previous studies have determined the churn rate of customers, including in the banking sector, using different models and techniques. Mutanen et al. (2010) applied Logistic Regression to customer churn in Portland bank. Dataset was separated into many samples based on periods of time and the researcher assessed each sample's performance lift curve and accuracy rate. In general, the accuracy rate of all models reached over 60%. Forhad et al. (2014) proposed the use of a rule-based classifier. According to their paper, some facts that could be used to create the rules are that if a customer doesn't pay his bill on time, he could become a churner and if there is a rapid decrease in the bill amount, the customer could churn. With these defined rules, customer churn prediction became easier than other methods. Amin et al. (2014) compared the performance of four different rule generation algorithms (i.e. exhaustive, genetic, covering, and LEM2) in a one-class and multi-class classification approach. The result showed that the genetic algorithm is the best method in both one-class classifiers and multi-class classifiers with the accuracy rate are 97.3% and 98.1%, respectively.

Kaur et al. (2013) considered Naïve Bayes, decision tree, and support vector machine (SVM) to build a model for forecasting the churner and non-churner. Dataset was collected from an Indian bank and had already a churn variable for churned customers and loyal active ones. The decision tree provided the most effective measurement for customer churn prediction based on the success of customer classification. Kumar and Ravi (2008) applied several models such as Multilayer Perceptron (MLP), Logistic Regression (LR), Decision Tree (J48), Random Forest, Radial Basis Function (RBF) network, and SVM and then found a the effective model. Dataset was from an American bank has had faced a rapid increase of customers leaving and decided to implement the promotion to improve the customer retention rate. As an unbalanced dataset, the authors used under-sampling, oversampling, a combination of under-sampling and oversampling, and SMOTE to balance it. Their finding is that all models recorded the best performance in the dataset was balanced by SMOTE and a combination of under-sampling and oversampling methods. In addition, Random Forest is considered the most successful model when compared with others. Besides, authors create rules for customer classification which act as an early warning system.

Dahiya and Bhatia (2015) worked on decision trees and logistic regression in three samples extracted from one dataset in the Weka environment. The first sample included 50 records and 10 features, the second one had 200 records and 50 features, and the last one covered 608 records and 1000 features. This study determined that the most suitable model in this situation was the decision tree because the accuracy rate of the decision tree was far much higher than the Logistic Regression technique. To study the percentage of a customer leaving, In another paper by Jinbo et al. (2007), AdaBoost was assessed better than CWC-SVM and the decision tree. Dataset was extracted from customers having outstanding in a China bank. Further review of the AdaBoost algorithm, Modest AdaBoost's effectiveness as well as learning ability were not as good as Real AdaBoost and Gentle AdaBoost.

Makruf et al. (2021) measured the churn rate in telecommunication services by 5 models such as Artificial Neural Network (ANN), Decision tree, K-Nearest Neighbor (KNN), G. Naive Bayes, and Support Vector Machine (SVM). The dataset was extracted from IBM Cloud Pak which consists of 20 features and 7043 records. This study applied all 5 models and determined the most efficient based on using a confusion matrix and evaluating the Accuracy rate, Precision rate, Recall rate, and F-Measure ratio. According to this paper, ANN had the highest accuracy with 79% and the highest precision with 67% while the decision tree had the lowest accuracy and the lowest precision with 70% and 67% respectively. In terms of recall, the model with the highest recall was Gaussian Naïve Bayes with 80% and the model with the lowest one was the decision tree with 49%. In conclusion, this paper recommended ANN and Gaussian Naïve Bayes for churn prediction in the telecommunication industry because of the outperformed models compared with others.

To improve the performance of machine learning measurements, Hudaib et al. (2015) proposed three hybrid models for predicting customer churn in a telecommunication company. Authors believed that the hybrid model could predict more accurately than an individual model. The aim of the paper was a combination of Multilayer Perceptron Artificial Neutral Networks (MLP-ANN) with K-means, Hierarchical clustering, and Self-organizing maps (SOM). By evaluating the accuracy rate and churn rate, the result shows that the hybrid model of MLP-ANN and K-means is the top of the exact prediction method and strengthens the above perception that the hybrid model could be more effective when compared with the individual model.

3. Theoretical framework

Basically, machine learning is a type of artificial intelligence that allows computers to learn on their own. The computer can modify its own actions to improve during the lessons and eventually achieve more model accuracy. The more times you learn, the more accurate the model will be. Machine learning was first defined by Samuel (1967). The author stated that machine learning is a field of study that provides learning capabilities for computers without being clearly programmed.

3.1. K-nearest neighbor

The first model presented in the following paragraph is K-Nearest neighbor (KNN). According to KNN, the same data will exist close together in a space, then deciding the label for undefined observation depends on the conditional probability of nearby observations. To distinguish the undefined observation, KNN uses a set of K coefficients (K>0) which are the numbers of nearby observations. One of the methods to determine the distance between neighborhoods to classify the above-undefined point and the point that needs to be defined is calculating Euclidean (Makruf et al., 2021) which is as follows:

$$Euclidean\ distance_{(x,x_k)} = \sqrt{\sum_{i=1}^{k}\left(x_{ij} - x_{kj}\right)^2} \qquad (1)$$

where, x_{ij} is the value of variable-j of observation–i (nearby observations), x_{kj} is the value of variable-j of observation–k (undefined observations)

K coefficient has a great influence on the classification results of the KNN model. Because the K coefficient represents the number of neighbors used to classify unspecified observations. Moreover, the flexibility of the model also depends on the K coefficient. If K is 1, it is not difficult to process, the result is low bias but is high variance and finally, the model is overfitted. If K increases, the flexibility of the model goes down, so variance also decreases but biasedness is higher (Gareth et al., 2013). To sum up, a model with low K has high flexibility but it tends to be overfitted, by contrast, the model with higher K is less flexible and could miss some information from the dataset. Therefore, the selection of the K coefficient is considered very important in the KNN model.

3.2. Random forest

A decision tree is ordinarily utilized for inductive inference. It is also useful for measuring discrete-valued functions for noisy data and being able to learn inconsistent phrases (Bardab et al., 2021). The decision tree consists of the root node–the first branch of the decision tree, the internal node–the next branches of the decision tree, and the leaf node–the final branches of the decision tree. Every node will interface with each other by small branches. To build a tree based on data, the data is divided into specific groups then a clustering-based model can be modeled in a tree structure. It is important to focus on capable of classifying data of nodes. It is dynamic between the classes of the node to evaluate the fit of the group in the model to the actual data set. If a node does not correctly classify the data, it is called an impurity node. Applying the concept that one decision tree might have high variance but an average of many decisions tree will have lower variance, Random forest measure decision trees and after that build a new final synthetic decision tree. Due to the process of creating all decision trees having random factors which are random samples and random sets of features, the result of each decision tree in a random forest absolutely different. The random forest contains many decision trees and each of them applies a decision tree algorithm with different data set as well as a set of features. Finally, the output of the random forest model is a result of all the above decision trees (Nohuddin et al., 2021).

Because one decision tree algorithm does not measure based on all of the features of the dataset, the model can be either overfitting or underfitting. However, this issue can be solved by a random forest algorithm when there is information aggregating between all decision trees and the random forest model has low bias and low variance, as well as the performance of the random forest, is better.

The theory of random forest is similar to the practical example in that before consumers decide to buy a certain type of product, they often read multiple reviews of previous shoppers instead of reading only one person's review and then making a final decision for themselves.

3.3. Gradient boosting

Boosting was developed based on the expectation to improve the limitations of previous weak hypotheses or weak models. Basically, Boosting could create several models and they learn to complement each other, in other words, the following model learns to avoid mistakes from previous models in boosting process. Additionally, models have their own weighted coefficient, these numbers can be changed up to different optimization methods. Therefore, Boosting is a sequence process, which cannot be handled in parallel, so it takes time relatively to train the model. After each loop, Boosting will likely reduce the following error exponentially for the model.

Gradient Boosting aims to provide solutions for minimizing the loss of model, which are:

$$min_{c_n=1:N, w_n=1:N} L(y, \sum_{n=1}^{N} c_n w_n) \quad (2)$$

where, L is the loss function value, y is the model prediction value, w_n is the n^{th} weak model, c_n is the n^{th} weighted coefficient of n^{th} weak model.

Boosting attempt to calculate solution after adding a new model to the model chain with the expectation of getting closer to the global minimum:

$$min_{c_n, w_n} L(y, W_{n-1} + c_n w_n) \quad (3)$$

$$W_{n-1} = \sum_{n=1}^{N-1} c_n w_n \quad (4)$$

A typical approach to finding the global minimum is Gradient Descent, i.e., the model will start from a point that is considered close to the global minimum and then use iterative operations to gradually get to the target–the point that makes the derivative of the model close to 0. Gradient Descent for a multivariable function is as follows:

$$\theta_{t+1} = \theta_t - \eta \nabla_\theta f(\theta_t) \quad (5)$$

where, $f(\theta_t)$ is the function that needs to find the global minimum where θ is a vector often used to denote the set of parameters of a model to be optimized, $\nabla_\theta f(\theta_t)$ is the derivative of the function $f(\theta_t)$ at any θ, η is learning rate.

The selection of learning rate is very important in practical problems. On the off chance that the learning rate is too small, the convergence speed of the model is sluggish, or in some cases, the model cannot arrive at the nearby ideal point. On the contrary, if the learning rate is too large, the algorithm gets close to the goal very quickly after entering the loop, but the algorithm cannot converge because the leap is too large, making the model go around and not arrive at the global minimum.

Applying Gradient Descent to compute the global minimum in Gradient Boosting through minimizing the Loss function L(y, W):

$$W_n = W_{n-1} - \eta \frac{\partial}{\partial w} L(W_{n-1}) \quad (6)$$

From there, we have the relationship as follows:

$$c_n w_n \approx -\eta \frac{\partial}{\partial w} L(W_{n-1}) \quad (7)$$

where, w_n is the n^{th} weak model and its weighted coefficient to be fed into the main model needs to match the value $-\eta \frac{\partial}{\partial w} L(W_{n-1})$–pseudo-residuals.

In summary, to deploy Gradient Boosting need to carry out the followings:

- Build equal pseudo residual value for each data point
- Create the i^{th} loop:
- New model training was added to match the value of pseudo-residuals
- Calculate the new weight ci of the trained model
- Update main model W = W + c_i * w_i
- Calculate the pseudo-residuals value as the basis for the next new model training
- Repeat with the $(i+1)^{th}$ loop

4. Data and methodology

4.1. Dataset

Dataset is collected information of customers using mobile banking services of Vietnam International Commercial Joint Stock Bank. The collection in the review is records of clients' activity and the status of their financial services, a total of 66,736 records in 2019. 35.59% of records are labeled Churn as 1 (Customers leave the bank) and 64.41% of records are marked Churn as 0 (Clients do not leave the bank). Table 1 shows model features.

4.2. Data transformation

Data transformation is the process of changing the structure while keeping the original value of the data. The transformed data will help the model to organize the dataset better and manage the data for humans and computers more easily. Properly formatted and validated values improve data quality and help the model avoid nulls, duplicates, or problems with incompatible formats. Features such as Client_Gender, Staff_VIB, SMS, Verify_Method, and EB_Register_Channel are encoded for the model to learn well.

4.3. Data cleaning

Data cleaning is an important process in optimizing model accuracy. This is the process in which you need to go through all the data and delete or update incomplete, missing, or duplicate information. Employed models do not handle missing data, this value can be replaced with mean, median, or 0. Fortunately, the dataset the authors use is not affected by the missing data.

4.4. Dataset separation

Dataset is divided into train set and test set. The train set is used to build the model and the test set is used to test the above model. The correct prediction of the model will be affected by random separation and different for each different partition.

In this study, we randomly select 30% of the test set from the total dataset, and the remaining 70% is utilized for model training purposes. For each sample, about 36% of customers left and were labeled as 1. As mentioned above, the train set will be estimated and checked from the test set to confirm whether the model is good or not.

Table 2 shows non-churner and churner proportions in the dataset.

Table 1: Model features

Features	Description
Customer_number	Encoded customer
Client_gender	Gender
Age	Customer age
Staff_VIB	Is VIB staff?
Tenure	Using e-banking period
SMS	Does customer register SMS?
Verify_method	Online verification method
EB_register_channel	Ebanking register channel
No_Activity_Name	Total kind of activities in mobile banking application
Type_Transactions	Total kind of transactions
Total_trans_no	Numbers of transactions
Avg_Trans_no_month	Average of transactions/ month
Avg_Trans_Amount	Average of transactions amount
Max_Trans_Amount	Maximum of transactions amount
Min_Trans_Amount	Minimum of transactions amount
No_CurrentAccount	Numbers of Current accounts
Avg_CurrentAccount_Balance	Average of Current account balance
Max_CurrentAccount_Balance	Maximum of Current account balance
Min_CurrentAccount_Balance	Minimum of Current account balance
No_TermDeposit	Numbers of Term deposits
Avg_TermDeposit_Balance	Average of Term deposit balance
Max_TermDeposite_Balance	Maximum of Term deposit balance
Min_TermDeposit_Balance	Minimum of Term deposit balance
No_Loan	Numbers of loans
Avg_Loan_Balance	Average of loan balance
Max_Loan_Balance	Maximum of loan balance
Min_Loan_Balance	Minimum of loan balance
No_CC	Numbers of credit cards
No_DC	Numbers of debit cards
Churn	Is customer churner or non-churner?

Table 2: Non-churner and churner proportion in the dataset

	Non-churner	Churner	% Churn rate
y_test	12863	7166	36%
y_train	30140	16594	36%
Total	43003	23760	36%

4.5. Feature importance

Feature importance shows the influence of each chosen feature on the classification ability of the model. By XGBoost, the rank of feature importance is shown in Fig. 1. Fig. 1 shows the score of the impact of the feature on classification. Tenure, the period time customer on the book, significantly affected to a classifier with the highest score.

4.6. Performance model evaluation of criteria Accuracy

Accuracy is defined as the ratio of the number of correct predictions to the total number of datasets. The accuracy rate is calculated based on the confusion matrix (Table 3).

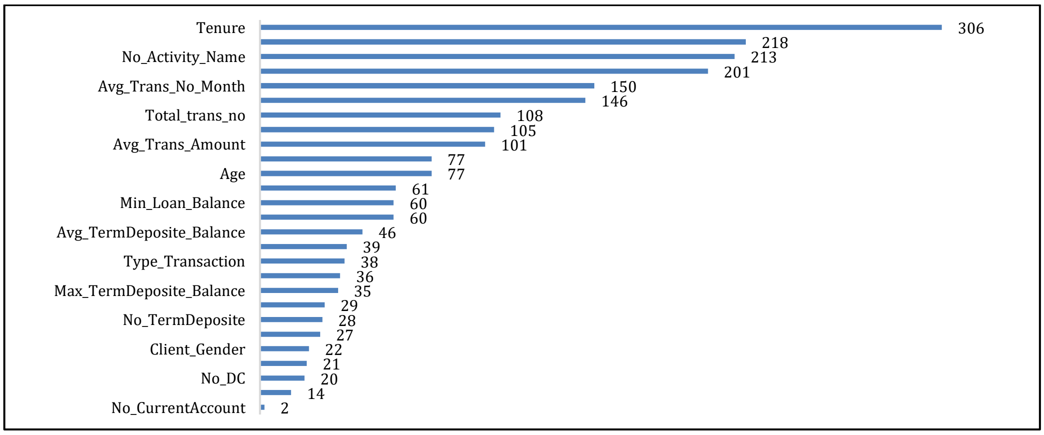

Fig. 1: Feature importance

Table 3: Confusion matrix

	Actually positive (1)	Actually negative (0)
Predicted positive (1)	True positive (TP)	False negative (FN)
Predicted negative (0)	False positive (FP)	True negative (TN)

Accuracy formula:

$$Accuracy = \frac{True\ positive + True\ negative}{T\hat{\acute{o}}ng\ s\acute{\hat{o}}\ l\text{ưỡ}ng\ m\tilde{\hat{a}}u\ d\text{ữ}\ \text{đ}o\acute{a}n} \tag{8}$$

4.7. Recall

Sensitivity is determined by the ratio of the number of true predictions to the total number of cases where the predictions are correct:

$$Sensitivity = Recall = \frac{True\ positive}{True\ positive + False\ negative} \tag{9}$$

4.8. Precision

Unlike Accuracy, Precision also measures accuracy, but only in one row of the confusion matrix. It measures the determinism or correct Positive classification of the model.

$$Precision = \frac{True\ positive}{True\ positive + False\ postitive} \tag{10}$$

4.9. F–score

Since the Recall and Precision coefficients are two different concepts, the F_1 index allows a balance between these two metrics. $F_1 - Score$ is defined as the harmonic mean between Precision and Recall.

$$F_1 = 2 * \frac{1}{\frac{1}{Recall} + \frac{1}{Precision}} \tag{11}$$

$$F_1 = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{12}$$

where, F_1 is a special case in the general formula F_β:

$$F_\beta = (1 + \beta^2) * \frac{Precision * Recall}{(\beta^2 * Precision) + Recall} \tag{13}$$

with F_2 we put a higher weight on Recall than Precision, and with $F_{0.5}$, we put a higher weight on Precision than Recall. F_1 is used when we are interested in the role of both Precision and Recall, or the model has both sensitivity and accuracy.

4.10. ROC_AUC

The ROC curve is a graph showing the performance of a classifier at all classification thresholds. The curve is drawn from two parameters, Recall and (1-Specificity). The ROC curve represents how well the model classifies according to defined thresholds (the boundary between classes). The lower the threshold, the more Positive points will increase the Recall. There exists a point on the ROC line that is close to the point with coordinates (0,1) on the graph (top left corner). The closer the ROC line is to that point, the more effective the model is.

5. Result and discussion

Results of the implementation of three models: KNN, Random forest, and Gradient Boosting are presented in the following subsections.

5.1. Confusion matrix

A confusion matrix is a matrix containing information about actual customer status and predicted results from the classifier. The performance of the model is assessed utilizing information (Polat et al., 2008). The confusion has four types of one:

- True Positive (TP) are actual customer churners and the model correctly predicts that.
- False Positives (FP) are customers, customers are actually non-churners and the model correctly predicts that.
- True Negative (TN) are customers who are actually churners but the predictive model is non-churner.
- False Negative (FN) are actual customers that are non-churners but the predictive model is churner.

Confusion matrix presents confusion matrix with 3 models that the authors use: K-NN, Random Forest, and Gradient Boosting. From the above matrix, the authors implement performance metrics. Fig. 2 shows the confusion matrix.

Table 4 shows some important criteria for evaluation and model comparison and they are visualized in Fig. 3.

Regarding Accuracy rate, the KNN model only achieved 61.84% while the Random Forest predicted accuracy to 77.07%, more efficient than them is the Gradient Boosting model with a rate of 79.45%.

About Roc_Auc–the area under the ROC curve, the KNN model is only 58.84%, Random Forest has an efficiency of 82.66% and Gradient Boosting reaches 86.13%.

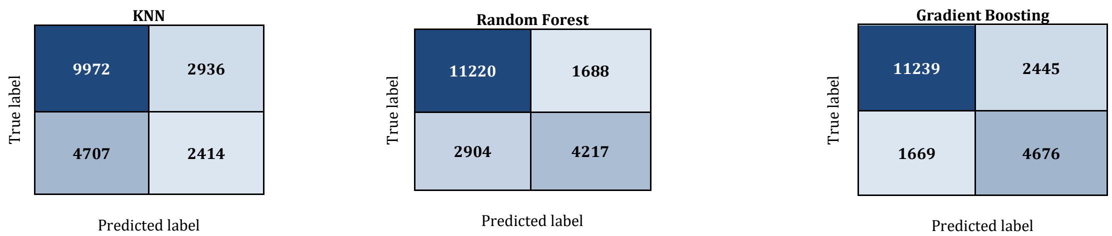

Fig. 2: Confusion matrix

Table 4: Effective criteria of models

Algorithm	Accuracy	Roc_Auc	F1_Score	Recall	Precision
KNN	61.84%	58.84%	28.71%	33.89%	45.12%
Random Forest	77.07%	82.66%	64.75%	59.22%	71.41%
Gradient Boosting	79.45%	86.13%	69.45%	65.66%	73.69%

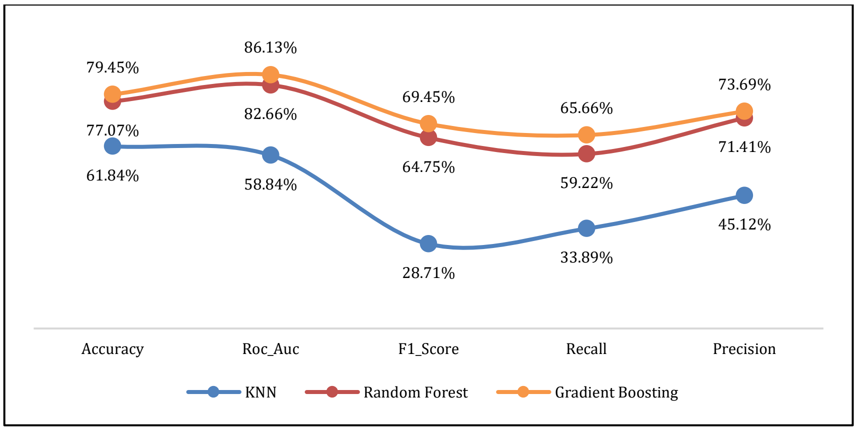

Fig. 3: Effective criteria of models

About Precision-The rate of correctly predicting leaving customers out of the total number of customers predicted to be leaving customers, KNN reaches a low of 45.12%, Random Forest reaches 71.41% and Gradient Boosting reaches 73.69%.

Regarding Recall-The ratio of correctly predicting leaving customers out of the total number of customers actually leaving, the effective KNN gradually decreases to only 33.89% while the remaining models predict about 60% and more (Random Forest–59.22%, Gradient Boosting–65.66%).

Therefore, in this study, the KNN model is less effective and different from the other three models. It is expected that Recall and Precision are as high as possible, however, if we adjust the model to increase Recall too much, it can lead to a decrease in Precision and vice versa. Therefore, we need to balance these two ratios through the F1-Score index. F1-Score–harmonic average between Precision and Recall.

In terms of F1-Score, Gradient Boosting has the highest rate (69.45%). It is possible to ask for the best model to predict customer churn in this data set when the authors apply the algorithm.

From Fig. 4, the ROC curve of Gradient Boosting is closest to the graph top left which means this model works perfectly. ROC curve of KNN is nearby to the random guessing line, so it is not an effective classifier in this study. ROC curve of Random forest has a trend as the same as one of Gradient Boosting.

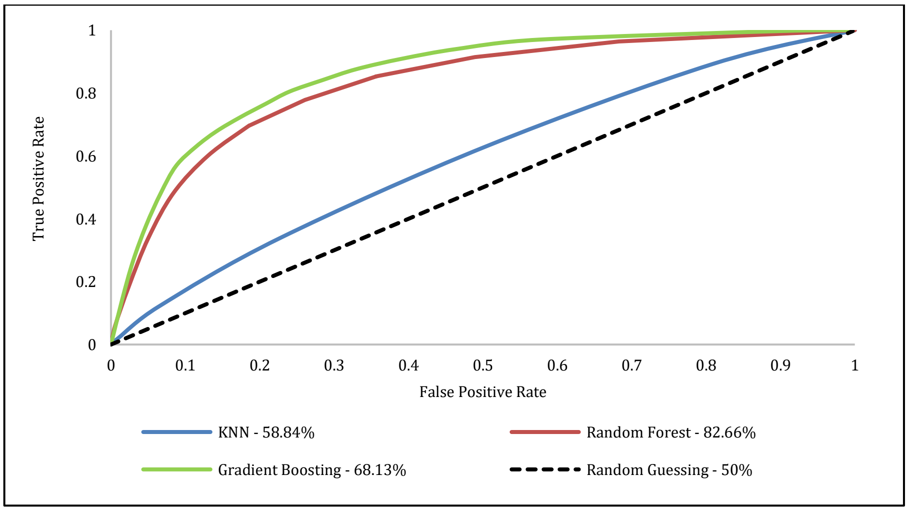

Fig. 4: ROC curve

5.2. Probability density graph

Fig. 5 illustrates the probability density that four models predict for each customer churn rate. The probability spectrum of KNN is plainly not the same as the other three models. Fig. 5 shows the probability density that 3 models predict for each customer. With the Random Forest model, the probability is spread from 0.0 to 1.0, highest at the point of 0.0 and decreasing as the probability increases. Gradient Boosting has relatively similar probability densities with the highest densities falling at the ends of 0.0 and 1.0.

Comparing the authors' results with the prediction probability density of models in Li et al.'s (2011) study as shown in Fig 6, the scores of the

Gradient Boosting model quite match the scores of the ANN, Naïve Bayes, and ANN models. KNN, Decision Tree, Logistic Regression, and RIPPER in Li et al.'s (2011) study, that is, the high probability density gradually approaches the two landmarks 0.0 and 1.0. Thus, with the authors' experimental results, K-NN and Random Forest have the opposite spectrum from other spectra and other studies.

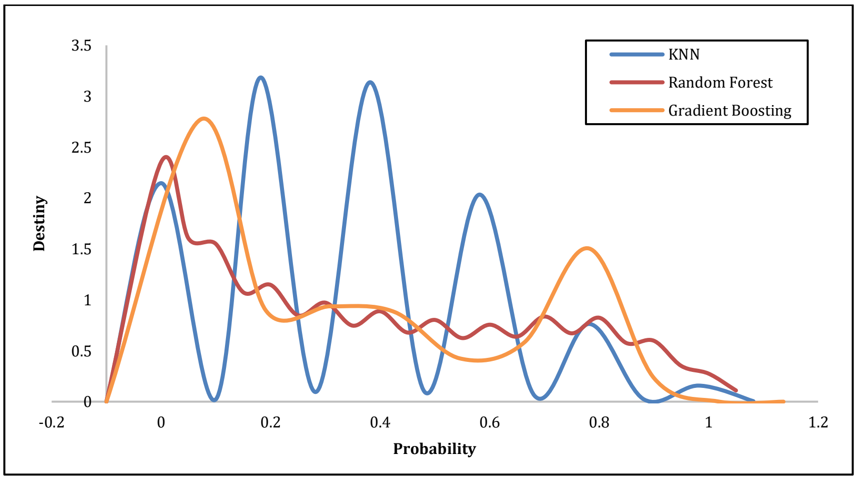

Fig. 5: Probability density of predicting customer churn rate

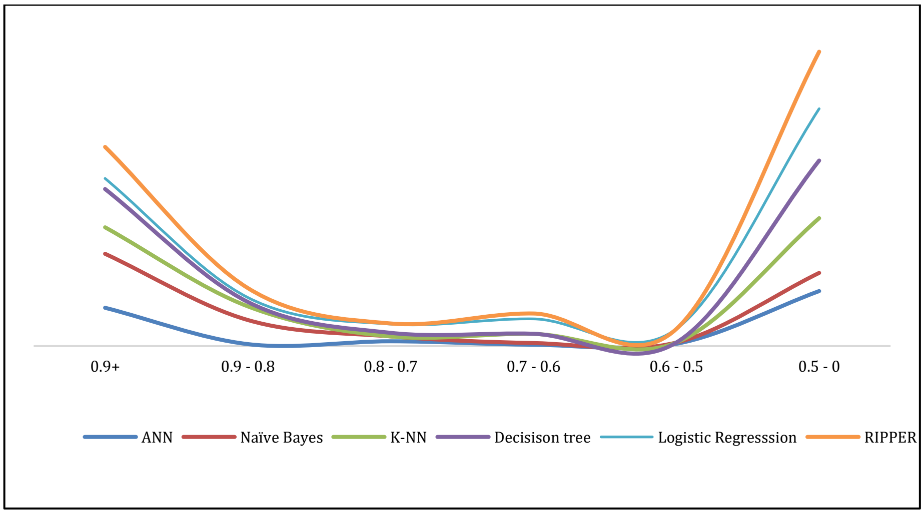

Fig. 6: Probability density of models in Li et al.'s (2011) study

5.3. Tuning models

Tuning is a trial-and-error process that has the purpose to determine the most efficient hyperparameters of the algorithm. By changing each parameter, the performance of the model could be also changed. We have performed hyperparameter finding of Gradient Boosting–The model has the highest Roc_auc ratio and the highest accuracy rate among the three models. After many times of tuning some parameters which are learning_rate, min_samples_split, min_samples_leaf, max_depth, subsample, and random_state. Up to the quality of the computer and how the system can handle it, it should be considered the number of values that can be tested. Because we did some very heavy-duty grid searches in these sections, they took us one hour or more to process. The new set of hyperparameters helps the model be better with a 79.71% accuracy rate and 86.23% Roc_Auc ratio. The new accuracy rate and Roc_auc ratios are 79.71% and 86.23% respectively at a set of hyperparameters as shown below:

- Gradient Boosting Classifier (learning_rate=0.1
- min_samples_split=500,min_samples_leaf=50
- max_depth=8,max_features='sqrt', subsample=0.8
- random_state=10)

5.4. Rules of classify churner or non-churner extracted from decision tree

To establish classification rules, this paper uses a 5-branch decision tree model based on the original data set to determine the cut-off points to rank customers leaving and not leaving. The rule is presented as an "If-Then" structure and makes a conclusion about the client's status. The rule table achieves 80% accuracy and 85% Roc_Auc. Rules are presented as an "If Then" structure and provide customer status results, the rule table includes information from the model and is worded into sentences for managers to read and understand. With more branches, the elements of the "If" condition will be more specific and categorized in more detail. Kumar and Ravi (2008) also got a set of rules on categorizing customers into two groups which are churners or non-churner. Table 5 shows the rules of classifying churner or non-churner.

Table 5: Rules of classifying churner or non-churner

No.	Rule	Result
1	Tenure<=6.41 months and Average of transactions/month>1.5 times and Tenure<=3.05 months and Total kind of activities in mobile banking application<=1.5 and Tenure<=1.55 months	Non-churner
2	Tenure<=6.41 months and Average of transactions/month>1.5 times and Tenure<=3.05 months and Total kind of activities in mobile banking application<=1.5 and Tenure>1.55 months	Churner
3	Tenure<=6.41 months and Average of transactions/month<=1.5 times and Tenure<=3.19 months and Total kind of activities in mobile banking application<=1.5	Churner
4	Tenure<=6.41 months and Average of transactions/month<=1.5 times and Tenure <=3.19 months and Total kind of activities in mobile banking application>1.5	Non-churner
5	Tenure<=6.41 months and Average of transactions/month>1.5 times and Tenure<=3.05 months and Total kind of activities in mobile banking application>1.5	Non-churner
6	Tenure<=6.41 months and Average of transactions/month>1.5 times and Tenure>3.05 months	Non-churner
7	Tenure>6.41 months and Average of transactions/month<=11.5 times and Total kind of activities in mobile banking application<=1.5	Churner
8	Tenure>6.41 months and Average of transactions/month<=11.5 times and Total kind of activities in mobile banking application>1.5 and Total kind of activities in mobile banking application<=6.5	Non-churner
9	Tenure>6.41 months and Average of transactions/month<=11.5 times and Total kind of activities in mobile banking application<=7.5 and Total kind of activities in mobile banking application>6.5	Churner
10	Tenure > 6.41 months and Average of transactions/month<=11.5 and Total kind of activities in mobile banking application>7.5	Churner
11	Tenure > 6.41 months and Average of transactions/month<=11.5 times and Total kind of activities in mobile banking application>7.5 and Average of transactions/month<=4.5 times	Churner
12	Tenure>6.41 months and Average of transactions/month<=11.5 times and Total kind of activities in mobile banking application>7.5 and Average of transactions/month>4.5 times	Churner
13	Tenure>6.41 months and Average of transactions/month>11.5 times and Numbers of transactions<=149.5 times and Maximum of Current account balance<=78.204.904.00 VND and Tenure<=7.41 months	Non-churner
14	Tenure>6.41 months and Average of transactions/month>11.5 times and Numbers of transactions<=149.5 times and Maximum of Current account balance<=78.204.904.00 VND and Tenure>7.41 months	Churner
15	Tenure>6.41 months and Average of transactions/month>11.5 times and Numbers of transactions<=49.5 times and Maximum of Current account balance>78.204.904.00 VND and Minimum of transaction amount<=394.145.00 VND	Churner
16	Tenure>6.41 months and Average of transactions/month>11.5 times and Numbers of transactions<=149.5 times and Maximum of Current account balance>78.204.904.00 VND and Minimum of transaction amount>394.145.00 VND	Churner
17	Tenure>6.41 months and Average of transactions/month>11.5 times and Numbers of transactions>149.5 times and Minimum of Current account balance<=595.474.00 VND and Total kind of transactions<=3.5 types	Non-churner
18	Tenure>6.41 months and Average of transactions/month>11.5 times and Numbers of transactions>149.5 times and Minimum of Current account balance<=595.474.00 VND and Total kind of transactions>3.5 types	Non-churner
19	Tenure>6.41 months and Average of transactions/month>11.5 times and Numbers of transactions>149.5 times and Minimum of Current account balance>595.474.00 VND	Non-churner

6. Conclusion and limitations

The authors conduct a statistical analysis of the data set's users and determine which attributes contribute to the model's classification and which attributes do not, after which they will be included in the cleaned dataset model, the attributes are discrete variables and will be assigned a code, and the variable that has little impact on the model will be removed. The research eventually picked Gradient Boosting with Accuracy rate equal to 79.71%, Roc_Auc equal to 86, 23%, and F1-Score equal to 70.01% based on model assessment indices such as Accuracy, Roc_Auc, F1-Score, Recall, Precision to choose the most optimum model.

A bank is an institution that collects a lot of data from clients. How banks may use data and develop the most efficient customer churn prediction model requires particular implementation approaches, but in principle, model creation will follow the stages below.

6.1. Collect data sets

Olle and Cai (2014) believed that the data set should be gathered on the real number of bank customers utilizing mobile banking services observed over a period of at least 4 months, and the outcomes of customer desertion at the final follow-up date should be detected. The amount of records in the dataset is determined by the number of clients that the bank follows, therefore the optimal model should create a dataset of at least 10000 records.

6.2. Select the appropriate attribute

The model's variable selection is critical since each variable has a different influence on the model's categorization level. According to the study's findings, the authors discovered that information regarding customers' credit, such as loans and savings, had no effect on the predictive model of users leaving mobile banking. Meanwhile, demographic information attained an average high categorization, and the information about the customer's behavior on the application had the greatest influence on the model. As a result, collecting additional characteristics that record client behaviors on the application will assist the categorization model is becoming more accurate and efficient.

6.3. Data set segmentation

As discussed in earlier sections, the machine learning model must be trained on a dataset and the projected outcomes must be validated against the actual data set. Typically, the train/test set ratio is 70/30. Because the data is split at random, the results will differ based on the division.

6.4. Selecting an appropriate algorithm

This study discovered three ways to select the best algorithm for the data.

- Utilize a standard algorithm, then tweak the model to improve model performance.
- Use a variety of methods, then measure and assess the performance of the models before selecting the model with the best performance.
- Use several separate algorithms and construct models that integrate multiple methods, then analyze the performance of the models and select the model with the best performance.

Banks must comprehend trends that identify and warn about future changes in consumer intentions, such as the decision to discontinue banking services. Learn about existing classification and clustering models in terms of technical foundation and approach performance, assessing the benefits and drawbacks of each model while concentrating on learning and developing new data mining approaches.

Following the selection of an appropriate model, the bank should test the model using the test data set to reevaluate the model's accuracy.

Based on the findings of the investigation, the author recommends the model Gradient Boosting with the following parameters: GradientBoostingClassifier (learning_rate=0.1, min_samples_split=500, min_samples_leaf=50, max_depth=8, max_features='sqrt', subsample=0.8, random_state=10)

6.5. Implement the model and utilize it on a regular basis

The bank must save the model and use it to provide quarterly projections for clients who use mobile banking services. There are several reasons to adopt the periodic model. For starters, we must update and enhance the model over time; the more the model is learned, the more predictive the later model will be; the model can even estimate the rate of customer turnover with new and unlearned features. Furthermore, the bank will be able to be proactive and build customized business and customer care plans for each client group as a result of the quarterly customer forecast. Finally, when machine learning evolves, the model is further enhanced, and the bank must assess the selected and in-use model with the models for opportune updates.

Although the initial study met the goal of developing a model to predict customer churn for the VIB mobile banking user data set, due to the limited time of research, we will combine models and collect more social factors of customers such as the number of dependents, social accounts, relationships with friends, information about time and the frequency of chatting with people, to develop predictive stronger models in the future. With the ongoing advancement of technology, particularly machine learning, this study expects that in the future, numerous forecasting models will be used in various parts of banking activities in particular and companies generally.

Funding

This paper is funded by Research Project coded NTCS2021-30 from Foreign Trade University, Vietnam.

Compliance with ethical standards

Conflict of interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Amin A, Khan C, Ali I, and Anwar S (2014). Customer churn prediction in telecommunication industry: With and without counter-example. In the Mexican International Conference on Artificial Intelligence. Springer, Tuxtla Gutiérrez, Mexico: 206-218. https://doi.org/10.1007/978-3-319-13650-9_19

Athanassopoulos AD (2000). Customer satisfaction cues to support market segmentation and explain switching behavior. Journal of Business Research, 47(3): 191-207. https://doi.org/10.1016/S0148-2963(98)00060-5

Bardab SN, Ahmed TM, and Mohammed TAA (2021). Data mining classification algorithms: An overview. International Journal of Advanced and Applied Sciences, 8 (2): 1-5. https://doi.org/10.21833/ijaas.2021.02.001

Bhattacharya CB (1998). When customers are members: Customer retention in paid membership contexts. Journal of the Academy of Marketing Science, 26(1): 31-44. https://doi.org/10.1177/0092070398261004

Colgate M, Stewart K, and Kinsella R (1996). Customer defection: A study of the student market in Ireland. International Journal of Bank Marketing, 14(3): 23-29. https://doi.org/10.1108/02652329610113144

Colgate MR and Danaher PJ (2000). Implementing a customer relationship strategy: The asymmetric impact of poor versus excellent execution. Journal of the Academy of Marketing Science, 28(3): 375-387. https://doi.org/10.1177/0092070300283006

Dahiya K and Bhatia S (2015). Customer churn analysis in Telecom industry. In the 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions), IEEE, New York, USA: 1-6. https://doi.org/10.1109/ICRITO.2015.7359318 **PMid:26827852**

Forhad N, Hussain MS, and Rahman RM (2014). Churn analysis: Predicting churners. In the 9th International Conference on Digital Information Management (ICDIM 2014), IEEE, New York, USA: 237-241. https://doi.org/10.1109/ICDIM.2014.6991433

Ganesh J, Arnold MJ, and Reynolds KE (2000). Understanding the customer base of service providers: An examination of the differences between switchers and stayers. Journal of

Marketing, 64(3): 65-87. https://doi.org/10.1509/jmkg.64.3.65.18028

Gareth J, Daniela W, Trevor H, and Robert T (2013). An introduction to statistical learning: With applications in R. Springer, Berlin, Germany.

Hudaib A, Dannoun R, Harfoushi O, Obiedat R, and Faris H (2015). Hybrid data mining models for predicting customer churn. International Journal of Communications, Network and System Sciences, 8: 91-96. https://doi.org/10.4236/ijcns.2015.85012

Jinbo S, Xiu L, and Wenhuang L (2007). The application of AdaBoost in customer churn prediction. In the 2007 International Conference on Service Systems and Service Management, IEEE, New York, USA: 1-6. https://doi.org/10.1109/ICSSSM.2007.4280172

Kaur M, Singh K, and Sharma N (2013). Data mining as a tool to predict the churn behaviour among Indian bank customers. International Journal on Recent and Innovation Trends in Computing and Communication, 1(9): 720-725.

Kumar DA and Ravi V (2008). Predicting credit card customer churn in banks using data mining. International Journal of Data Analysis Techniques and Strategies, 1(1): 4-28. https://doi.org/10.1504/IJDATS.2008.020020

Li F, Lei J, Tian Y, Punyapatthanakul S, and Wang YJ (2011). Model selection strategy for customer attrition risk prediction in retail banking. In the Ninth Australasian Data Mining Conference (AusDM'11), Australian Computer Society, Ballarat, Australia, 121: 119-124.

Mahmoud MA, Hinson RE, and Duut DM (2019). Market orientation and customer satisfaction: The role of service quality and innovation. International Journal of Business and Emerging Markets, 11(2): 144-167. https://doi.org/10.1504/IJBEM.2019.10022489

Makruf M, Bramantoro A, Alyamani HJ, Alesawi S, and Alturki R (2021). Classification methods comparison for customer churn prediction in the telecommunication industry. International Journal of Advanced and Applied Sciences, 8(12): 1-8. https://doi.org/10.21833/ijaas.2021.12.001

Mizerski RW (1982). An attribution explanation of the disproportionate influence of unfavorable information. Journal of Consumer Research, 9(3): 301-310. https://doi.org/10.1086/208925

Mutanen T, Nousiainen S, and Ahola J (2010). Customer churn prediction–A case study in retail banking. In: Soares C and Ghani R (Eds.), Data mining for business applications: 77-83. Volume 218, IOS Press, Amsterdam, Netherlands.

Nohuddin PN, Noormanshah WM, and Zainol Z (2021). Content analytics based on random forest classification technique: An empirical evaluation using online news dataset. International Journal of Advanced and Applied Sciences, 8(2): 77-84. https://doi.org/10.21833/ijaas.2021.02.011

Olle GDO and Cai S (2014). A hybrid churn prediction model in mobile telecommunication industry. International Journal of e-Education, e-Business, e-Management and e-Learning, 4: 55-62. https://doi.org/10.7763/IJEEEE.2014.V4.302

Paulin M, Perrien J, Ferguson RJ, Salazar AMA, and Seruya LM (1998). Relational norms and client retention: External effectiveness of commercial banking in Canada and Mexico. International Journal of Bank Marketing, 16(1): 24-31. https://doi.org/10.1108/02652329810197807

Phan DTT, Nguyen TTH, and Bui TA (2019). Going beyond border? Intention to use international bank cards in Vietnam. The Journal of Asian Finance, Economics and Business, 6(3): 315-325. https://doi.org/10.13106/jafeb.2019.vol6.no3.315

Poel VDD and Lariviere B (2004). Customer attrition analysis for financial services using proportional hazard models. European Journal of Operational Research, 157(1): 196-217. https://doi.org/10.1016/S0377-2217(03)00069-9

Polat K, Güneş S, and Arslan A (2008). A cascade learning system for classification of diabetes disease: Generalized discriminant analysis and least square support vector machine. Expert Systems with Applications, 34(1): 482-487. https://doi.org/10.1016/j.eswa.2006.09.012

Rasmusson E (1999). Complaints can build relationships. Sales and Marketing Management, 151(9): 89-89.

Reichheld FF (1996). Learning from customer defections. Harvard Business Review, 74(2): 56-67.

Samuel AL (1967). Some studies in machine learning using the game of checkers. II—Recent progress. IBM Journal of Research and Development, 11(6): 601-617. https://doi.org/10.1147/rd.116.0601

Stum DL and Thiry A (1991). Building customer loyalty. Training and Development Journal, 45(4): 34-36.

Zeithaml VA, Berry LL, and Parasuraman A (1996). The behavioral consequences of service quality. Journal of Marketing, 60(2): 31-46. https://doi.org/10.1177/002224299606000203