Proceedings of the 3rd International Conference on Management Research and Economic Development
DOI: 10.54254/2754-1169/2025.31298

Determinants of Credit Card Attrition: A Machine Learning-Driven Analysis with Empirical Evidence from Thera Bank

Chengting Jiang[1*], Tongying Chen[2]

[1]School of Mathematics and Statistics, Fuzhou University, Fuzhou, China
[2]Zhengzhou Foreign Language High School, Zhengzhou, China
**Corresponding Author. Email: 072202234@fzu.edu.cn*

Abstract. In the highly competitive banking environment, the issue of bank credit card customer churn is of great concern. The purpose of this study is to develop machine learning models for predicting whether a credit card customer is churned or not and determining the core factors of customer churn. By comprehensively collecting multidimensional data from Thera Bank customers and constructing three models, Logistic Regression, Neural Network, and XG boost, to analyze the data. After the study, it was found that factors such as the frequency of customers' recent contact with the bank and the number of product holdings may have a significant impact on customer churn or not. We accurately assess the characteristics of customers who tend to churn after analyzing these factors based on machine learning models, for example, customers with a lower total number of credit card transactions and a higher total transaction amount are more likely to churn. The results of this research help banks gain a deeper understanding of customer behavior and provide a data-driven basis for formulating targeted customer retention strategies, according to which banks can optimize their products and services and improve their customer management processes to effectively reduce customer churn, thereby enhancing their competitiveness in the banking industry and achieving sustainable development.

Keywords: Credit Card Attrition Prediction, Machine Learning, SHAP Feature Importance Analysis, Logistic Regression, XGBoost

1. Introduction

In recent years, the problem of bank credit card customer churn has received extensive attention from both academia and the industry. Customer churn not only leads to a reduction in the bank's revenue but also has a negative impact on the bank's reputation and long-term growth. Therefore, it is of great significance for the sustainable development of banks to deeply investigate the causes, development, and current situation of the customer churn problem and put forward effective countermeasures and suggestions.

Several articles have been written on the problem of bank credit card customer churn. Liang [1], sorted out the definition of customer churn, the reasons, and the importance of customer churn management. The article defines customer churn as a shift of customers to competitors, resulting in loss of profits. In order to manage churn, it is critical to identify customers who may be switching to

DOI: 10.54254/2754-1169/2025.31298

competing banks. The article also mentions that customer churn can be categorized into two main types: voluntary churn and involuntary churn, in which voluntary churn is relatively difficult to identify because customers consciously terminate their relationship with the bank based on their own subjective will.

In the area of corporate social responsibility (CSR), studies have also examined the impact of CSR practices on customer loyalty. Mehnaz et al. (Mehnaz et al.) [2] investigated how customer loyalty is affected by CSR practices in Pakistani banking sector. The results of the study indicate that there is a significant positive relationship between perceived CSR, bank reputation and customer loyalty and customer satisfaction and that bank reputation plays a key mediating role between perceived CSR and customer loyalty.

Given that whether a bank's credit card customers are churned or not can be regarded as a typical binary prediction problem, a number of well-established models have been derived and utilized in this particular domain. Bahel et al. [3] studied the performance and advantages and disadvantages of five binary classification models which contains Logistic Regression, Naïve Bayes Classifier, K-Nearest Neighbours Classifier, Decision Tree Classifier, Random Forest Classifier. And now there are researchers who have applied machine learning models in the problem of predicting customer churn, Peng et al. [4] by using GA - XGBoost algorithm to construct a bank customer churn prediction model, and with the help of the SHAP framework for the model to carry out detailed local interpretive analysis, aimed at providing valuable decision-making basis and targeted recommendations for the decision-making layer of the bank, helping the banking industry in advance to prevent customer churn risk. It aims to provide valuable decision-making basis and targeted suggestions for banks' decision-makers, and help the banking industry to prevent the risk of customer churn in advance. Siddiqui et al. [5] proposed three models based on different feature set combinations and machine learning algorithms to predict credit card customer churn. These three models cover models that consider all variables: models that separate categorical and continuous features and models that focus on important features using feature selection techniques. The study applied different machine learning models including decision trees, random forests, support vector machines, K nearest neighbors, XGBoost and logistic regression. Their findings provide instructive examples to guide our subsequent research efforts.

In summary, bank credit card customer churn management has become an important part of bank strategic development. Through the application of predictive modeling, cause analysis, retention strategies, and machine learning techniques, banks can more effectively deal with customer churn and enhance customer loyalty, thereby achieving long-term success and sustainable development. These research results not only provide banks with a theoretical basis for customer churn prediction, but also provide practical guidance for the implementation of effective customer relationship management strategies.

Based on the above research, this paper focuses the problem on the credit card customer churn of Thera bank in the US banking industry, tries to predict whether the group is churning or not using different machine learning methods, and focuses on evaluating and analyzing each machine learning model in terms of model performance as well as interpretability. Finally, based on the model results, we suggest to the bank that credit card users with characteristics such as the number of times the customer has recently initiated contact with the bank and the number of product holdings against the bank are more likely to churn, so they should be maintained on a day-to-day basis in order to prevent customer churn.

DOI: 10.54254/2754-1169/2025.31298

2. Method

2.1. Descriptive statistics

The Thera Bank customer churn prediction dataset from the Kaggle database (https://www.kaggle.com/datasets/sakshigoyal7/credit-card-customers), which contains detailed records of over 10,000 credit card holders, was selected for this study.

First we performed descriptive statistics on the data, we plotted a pie chart for the response variable attrition_Flag and found that the number of churned customers in the sample was much less than the number of customers who were churned, which suggests that there is a more pronounced sample imbalance in the dataset, which may have an impact on our subsequent analysis.

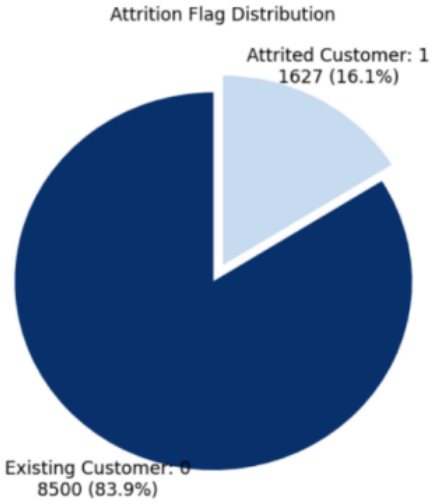

Figure 1. Attrition flag distribution

Since there may be a certain degree of correlation between the covariates, when this correlation is too high a multicollinear type problem occurs, which triggers problems such as wider confidence intervals for the coefficients, unreliable parameter estimation results, difficulty in interpreting the coefficients, inability to screen out variables with significant effects, and reduced predictive power of the model. In order to detect the presence of multicollinearity among covariates and to examine the correlation between the covariates and the response variable, we plotted the correlation coefficient matrices of all variables as shown in Figure 2.

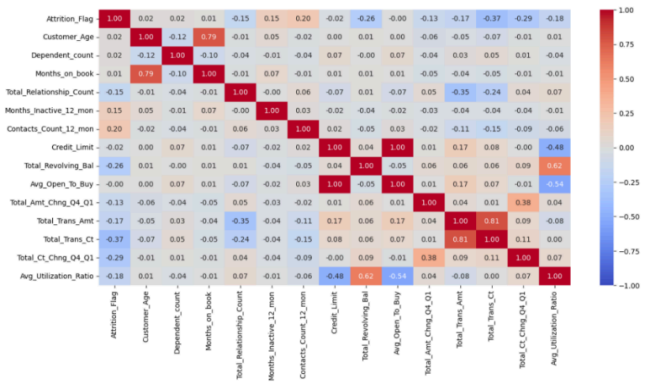

Figure 2. Correlation matrix of features influencing credit card customer attrition

DOI: 10.54254/2754-1169/2025.31298

With Figure 2 we can see that there are several pairs of covariates with high correlation coefficients. Total_Trans_Amt and Total_Trans_Ct have a correlation coefficient of 0.81, Avg_Open_To_Buy and Credit_Limit have a correlation coefficient of 0.62, and Total_Revolving_Bal and Avg_Utilization_Rati have a correlation coefficient of 0.62. However, none of their correlation coefficients exceeded 0.9, indicating that each pair of variables each contains unique information that cannot be fully explained by the other. For example, the existence of mega-transaction orders makes there is no complete linear correlation between the total amount of transactions and the number of transactions, and it is possible that the amount of one mega-transaction will be larger than the total amount of multiple normal transactions; the user's desire to spend may grow with the increase of credit limit, resulting in the remaining available credit per month to be less instead even if the credit limit becomes higher. To summarize, we do not see the need to delete the pairwise variables with high correlation and choose to keep all of them.

In addition, we found that the covariates Contacts_Count_12_mon and Total Relationship Count were strongly correlated with the response variable, so we plotted bar charts of the change in customer churn with each of the two variables.

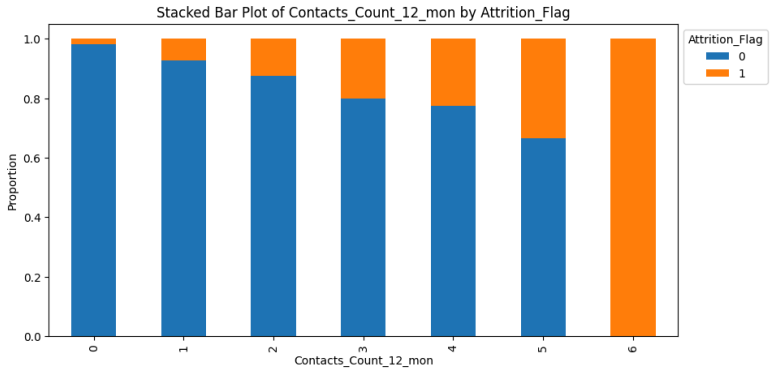

Figure 3. Stacked bar plot of Contacts_Count_12_mon by attrition_flag

Figure 3 shows that the risk of churn increases with the number of customer contacts with the bank. Customers contact the bank for help when they encounter problems or needs, and the more these customers interact with the bank the more problems they encounter, such as customer service inquiries, billing or transaction inquiries, loan or credit card limit increase applications, complaints or feedback, etc. As these problems arise more often, customers will gradually question the bank's professionalism, thus reducing their trust and ultimately choosing to leave.As these problems arise more often, customers will gradually question the bank's professionalism, thereby reducing their trust and ultimately choosing to leave. This suggests that banks need to improve service quality and reduce transaction risks in the process of serving customers, and try to avoid problems in transactions, which in turn reduces the number of times customers contact the bank on their own initiative, improves customer trust in the bank, and stabilizes the relationship between customers and the bank.

DOI: 10.54254/2754-1169/2025.31298

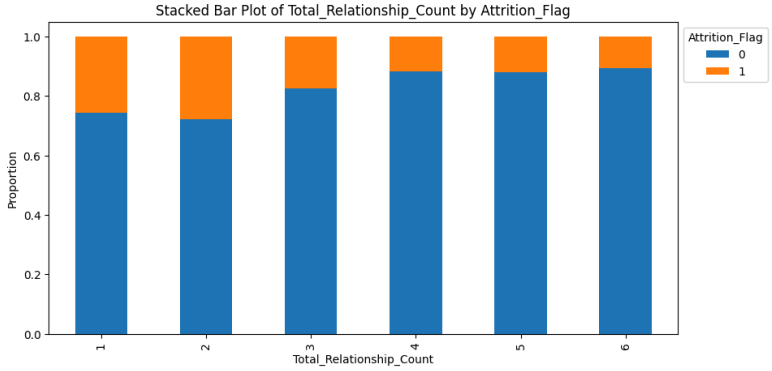

Figure 4. Stacked bar plot of total relationship count by Attrition_Flag

Figure 4 shows that as the number of products owned by the customer increases, the smaller the percentage of customer churn. If the more products they own, the cost of customers leaving their bank to use other banks' credit cards will be greater, and in order to avoid cumbersome processes and procedures, customers are more willing to stay in their bank rather than using other banks' credit cards. This suggests that getting customers to buy more of the bank's products will create a sense of dependence on the bank and make them less likely to turnover, provided that the customers do not resent it.

By plotting the correlations between variables, we found some pairs of covariates with correlations, but since their correlation coefficients did not reach 0.9, we chose to keep all of them, in addition, Contacts_Count_12_mon and Total Relationship Count have higher correlation coefficients with the response variables, and we speculate that they play a more important role in the subsequent modeling process, we hypothesize that they play a more important role in the subsequent modeling process.

2.2. Model construction

In Section 2.1, we delve into the link between the covariates and the response variables and find that the variables Total_Trans_ ct and contacts count_ 12_ mon are significantly related to the phenomenon of bank customer churn. In addition, we also note the presence of sample imbalance in the response variable, which may pose some challenges to our subsequent modeling and analytical work. Therefore, we will discuss and evaluate the model comprehensively in terms of these two dimensions.

In evaluating the model performance, we use two performance metrics, Accuracy and Recall, to select and evaluate the models [6]. The formula for these two performance metrics can be expressed as:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}, Recall = \frac{TP}{TP+FN} \quad (1)$$

Where TP, TN, FP, and FN denote the number of samples with test results of true positive, true negative, false positive, and false negative, respectively. Accuracy denotes the number of samples correctly predicted by the model as a proportion of the total number of samples, while Recall denotes the proportion of samples successfully predicted by the model to be in the positive category as a proportion of all samples that were actually in the positive category. On the other hand, a classifier that judges a customer as about to be lost but in fact the customer will not be lost will

DOI: 10.54254/2754-1169/2025.31298

cause relatively little damage to the bank and should be considered a relatively minor error. Therefore, when evaluating classifiers, more emphasis should be placed on evaluating the Recall performance metric, because the higher the metric, the smaller the proportion of the first type of error that occurs. The Accuracy performance metric, on the other hand, reflects the overall predictive performance of the model and should not be too low.

2.2.1. Logistic regression

Bank customer churn prediction is a typical binary classification problem. We first use Logistic Regression (LR) model to analyze and predict the data. Logistic regression is a widely used statistical method specifically for dichotomous problems [7]. The model is based on a core assumption that there is a linear relationship between the Log Occurrence Ratio (Log Odds) and each covariate, which can be expressed by the following equation:

$$\log\left(\frac{P(Y=1)}{P(Y=0)}\right) = X^T\beta \tag{2}$$

In this equation, represents the covariate and the response variable, i.e., whether the bank customer is churned or not. By solving the logistic regression model, we can determine the regression coefficient between the logit odds and each covariate, which reflect the extent to which each covariate affects the response variable (i.e., customer churn) to the degree of influence on the response variable (i.e., customer churn). This approach provides a quantitative way to assess the impact of different factors on the likelihood of customer churn.

We used the Logistics Regression model for the dataset and obtained the regression coefficients and p-values for each covariate as shown in Table 1:

Table 1. Logistic regression coefficients and p-values based on original data

Variable	Coef	P-value	Variable	Coef	P-value
Const	-3.0579	0.000***	Gender_M	-0.4747	0.000***
Customer_Age	0.0367	0.605	Education_Level_Doctorate	0.1241	0.009**
Dependent_count	0.1784	0.000***	Education_Level_Graduate	0.1454	0.059
Months_on_book	-0.0978	0.165	Education_Level_High School	0.0827	0.226
Total_Relationship_Count	-0.6976	0.000***	Education_Level_Post-Graduate	0.1012	0.049*
Months_Inactive_12_mon	0.523	0.000***	Education_Level_Uneducated	0.0831	0.194
Contacts_Count_12_mon	0.5833	0.000***	Marital_Status_Married	-0.276	0.001***
Credit_Limit	-0.1024	1.000	Marital_Status_Single	0.0004	0.997
Total_Revolving_Bal	-0.7269	1.000	Income_Category_$40K - $60K	-0.2859	0.001***
Avg_Open_To_Buy	-0.0372	1.000	Income_Category_$60K - $80K	-0.1281	0.071
Total_Amt_Chng_Q4_Q1	-0.1036	0.022*	Income_Category_$80K - $120K	-0.0538	0.435
Total_Trans_Amt	1.6178	0.000***	Income_Category_Less than $40K	-0.2972	0.015*
Total_Trans_Ct	-2.7941	0.000***	Card_Category_Gold	0.12	0.004**
Total_Ct_Chng_Q4_Q1	-0.6412	0.000***	Card_Category_Platinum	0.0503	0.133
Avg_Utilization_Ratio	-0.0734	0.334	Card_Category_Silver	0.0534	0.293

DOI: 10.54254/2754-1169/2025.31298

In Logistic Regression, if the regression coefficient is positive, it means that the larger the covariate, the greater the probability that the response variable will be 1, all other things being equal, and if the regression coefficient is negative, it means that the larger the covariate, the smaller the probability that the response variable will be 1, all other things being equal. By analyzing the regression coefficients, we can conclude that customers who are married have a lower probability of churn; Customers with higher levels of education have a relatively greater chance of losing customers; The larger the total transaction amount in the past 12 months, the more likely customers are to be lost, while the more total transactions in the past 12 months, the less likely customers are to be lost.

When making statistical inferences, we often use p-values [8] to determine whether a variable is significant or not. If the p-value of a variable is greater than a given significance level (set to 0.05), we claim that the regression coefficient of that variable is not significant, i.e., not significantly different from zero. We rebuilt the Logistic Regression model after removing the non-significant variables, and the recall of the logistic regression model with the non-significant variables removed dropped by almost 6% in comparison and the correctness rate dropped by almost 1% compared to the previous one. his means that although these variables are not significant, they are still explanatory of the response variable, and we ultimately chose to keep the non-significant variables in order to ensure the predictive performance of the model.

Response variable having sample imbalance may affect the effectiveness of the model [9], it is more likely to misclassify the minority class samples into the majority class, in the dataset we are interested in churning customers are minority class samples so this may lead to higher false negative rate, so we need to balance the dataset. we used both SMOTE and random undersampling methods to obtain oversampled and undersampled data, respectively, before training logistic regression models based on the oversampled and undersampled data and subsequently evaluating them. The evaluation results are shown in Figure 5.

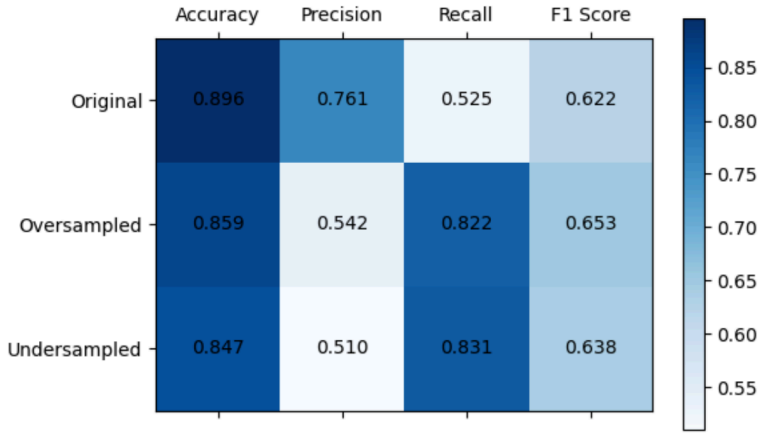

Figure 5. Performance metrics of logistic regression different datasets

From Figure 5 we can see that the lower recall value on Original data indicates that his false negatives are higher, which is not what we want. After the sample balancing process, we found that the recall value of the model can be improved substantially, but accordingly its false positive rate also increased, indicating that after the sample balancing, there are a number of non-churned customers we predicted to be churned, and their characteristics may be more similar to those of the churned customers, and it may be necessary to use other methods of modeling to identify them with the churned customers.

DOI: 10.54254/2754-1169/2025.31298

Again, from Figure 5 it can be found that logistic regression has the highest recall on the undersampled dataset at 83.1%. Therefore, this model was adopted as the final model for logistic regression for this problem. The regression results of the logistic regression model trained on the undersampled data are shown in Table 2.

Table 2. Logistic regression coefficients and p-values based on original data

Variable	Coef	P-value	Variable	Coef	P-value
const	-1.4833	0.000***	Gender_M	-0.3633	0.002**
Customer_Age	0.0363	0.721	Education_Level_Doctorate	0.1377	0.043*
Dependent_count	0.1193	0.056	Education_Level_Graduate	0.1861	0.078
Months_on_book	-0.1547	0.122	Education_Level_High School	0.1405	0.134
Total_Relationship_Count	-0.5694	0.00***	Education_Level_Post-Graduate	0.1555	0.032*
Months_Inactive_12_mon	0.5769	0.000***	Education_Level_Uneducated	0.1674	0.059
Contacts_Count_12_mon	0.5931	0.000***	Marital_Status_Married	-0.1158	0.305
Credit_Limit	-0.1195	1	Marital_Status_Single	0.1275	0.262
Total_Revolving_Bal	-0.5385	1	Income_Category_$40K - $60K	-0.2435	0.051
Avg_Open_To_Buy	-0.0712	1	Income_Category_$60K - $80K	-0.1246	0.215
Total_Amt_Chng_Q4_Q1	0.1799	0.005**	Income_Category_$80K - $120K	-0.0379	0.693
Total_Trans_Amt	1.8334	0.000***	Income_Category_Less than $40K	-0.2526	0.149
Total_Trans_Ct	-3.0761	0.000***	Card_Category_Gold	0.0719	0.197
Total_Ct_Chng_Q4_Q1	-0.539	0.000***	Card_Category_Platinum	0.1227	0.244
Avg_Utilization_Ratio	0.1657	0.104	Card_Category_Silver	0.051	0.456

From the results of Table 2 logistic regression, it can be found that the total number of transactions made by customers using Thera Bank credit cards in the past 12 months has a greater impact on the probability of churn. This may be due to the fact that frequent transactions imply that customers are highly active and have developed strong habits and dependence on the product (credit card). It also reflects a high level of customer satisfaction with the credit card service and therefore a lower likelihood of churning. However, the model also shows that the greater the total amount of Total_Trans_Amt total transactions by customers using their credit cards in the past 12 months, the greater the chances of customer churn. It is also possible to conclude that the higher the number of times a customer has been in contact with the bank in the last 12 months and the higher the number of months of inactive interactions with the bank's related business, the more likely it is that the customer will churn; the higher the total number of customers' holdings of Thera Bank's financial products of all types, and the higher the change in the number of transactions (in the fourth quarter compared to the first quarter), the less likely it is that the customer will churn, and so on.

Finally, we plotted the confusion matrix on the downsampled dataset as shown in Figure 6. It can be seen that the accuracy of the logistic regression results, the recall rate, although has reached a high level, can be better predicted, but there is still room for improvement, mainly in the model will be part of the unchurned customers predicted as churned customers, although for the bank of this type of error can be afforded, however, the maintenance cost required for customers who are about to leave will be higher than that for stable customers. Therefore, in order to save costs, we should also try to correctly predict unchurned (stable) customers.

DOI: 10.54254/2754-1169/2025.31298

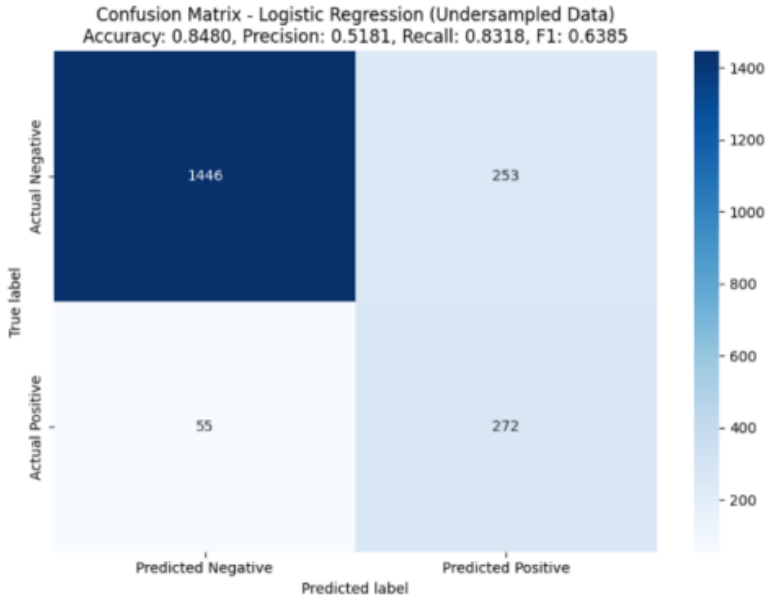

Figure 6. Confusion matrix of logistic regression on undersampled dataset

2.2.2. Feed-forward neural network

The Logistics Regression model has a strong explanatory power, but the results are less satisfactory in terms of model performance, even after we processed it by sample balancing, the recall value is only 83.1% and has a high false positive rate. This suggests that the relationship between covariates and response variables is not simply linear, but may have a more complex functional relationship. Artificial Neural Networks, as a natural heuristic algorithm that can model any kind of complex function, are more effective than traditional models in solving classification problems. As it can generate optimal ANN parameters, rules and topologies, thus providing the best classification performance while taking into account the quality of the solution, computational cost and avoiding local minima [10]. Therefore, we use feed forward neural network to train and predict the credit card customer churn rate of Thera bank. We changed the parameters of the neural network, they are: the number of hidden layers of the neural network, the activation function, the optimizer, and the threshold of classification. By adjusting the above parameters, the neural network with the best possible prediction performance that we can get, its structure and some parameters are shown in Table 3. With this structure, if the complexity of the model is increased further, such as by increasing the number of hidden layers and neurons, the improvement in accuracy and recall is minimal, so we end up using this structure.

Table 3. Neural network architecture

Type of layer	Activation function	Neuron count	Optimizer
Imput layer	relu	512	
1st hidden layer	relu	256	
2nd hidden layer	relu	128	SGD
3rd hidden layer	relu	64	
Output layer	sigmoid	1	

DOI: 10.54254/2754-1169/2025.31298

Meanwhile, to prevent overfitting, we set the L2 regularization parameter in each fully connected layer with a penalty factor of 0.01 to penalize larger weight values and suppress the complexity of the model; we added a Dropout layer after each layer of the neural network to discard 20% of the neurons; and we used the EarlyStopping strategy [11] to stop the training of the validation set when its Recall for successive 10 cycles no longer improves, training is stopped and the model weights are restored to those at the highest Recall.

We oversampled and undersampled the data before. Based on the above neural network structure, we train the neural network using raw, oversampled as well as undersampled data respectively and try to get the highest accuracy and recall of the model. The accuracy and recall of the model trained using the original, upsampled and downsampled data are shown in Figure 7.

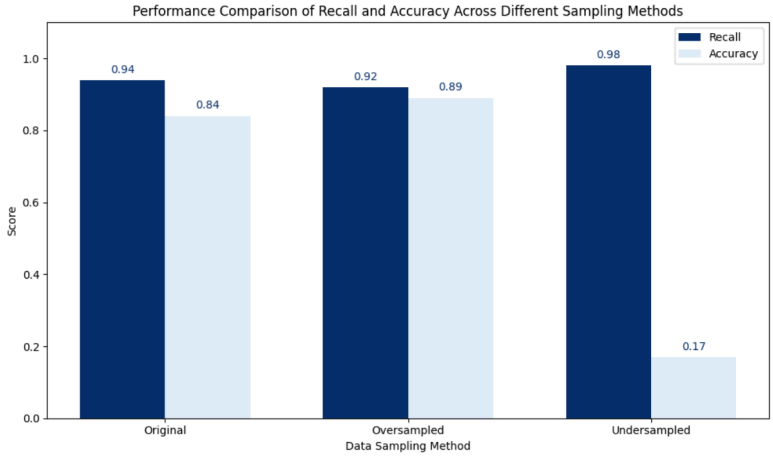

Figure 7. Comparison of model performance under multiple sampling methods

By comparison, we can find that the correct rate of the neural network trained using undersampled data is too low for practical application, so it is excluded. This may be due to the fact that the features of unchurned and churned customers in our undersampled dataset are relatively similar, so the neural network is also unable to accurately distinguish between them. The neural network trained with ORIGINAL data and OVERSAMPLED data performs better. As mentioned before, the problem mainly needs to focus on the recall index, so the neural network trained by original data is still chosen as the final neural network model. The training process and confusion matrix are as follows:

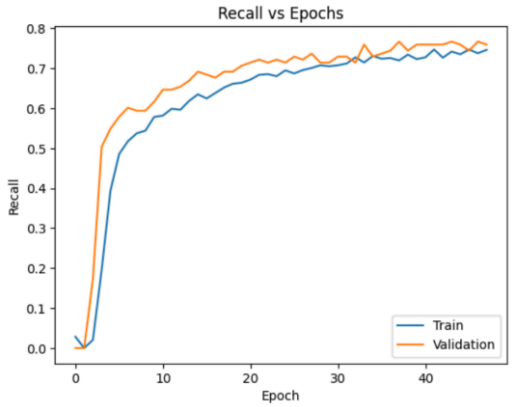

Figure 8. Recall vs epochs - comparison between training set and validation set

DOI: 10.54254/2754-1169/2025.31298

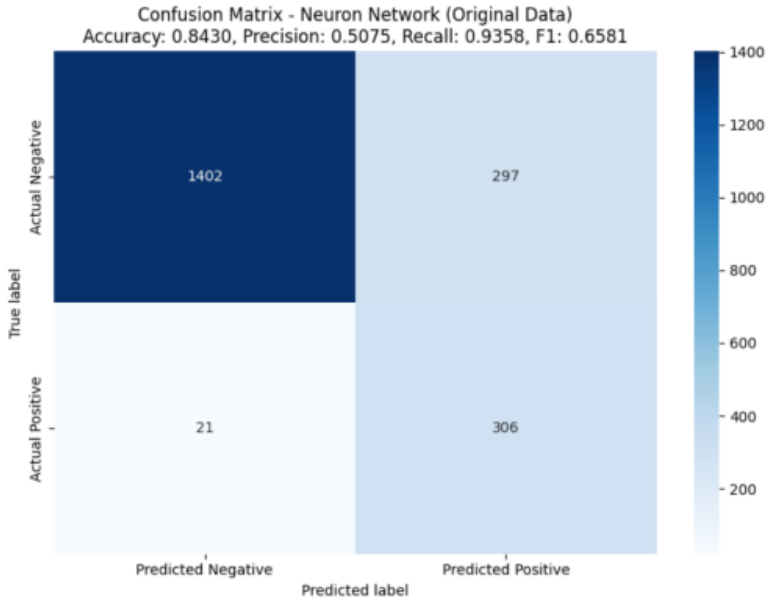

Figure 9. Confusion matrix and performance metrics for neural network

The training process of the model shown by Figure 8. shows that the neural network performs close to the training set and validation set, the model has no overfitting, and the generalization ability is good. From Figure 9. it can be seen that the model achieves better recall (94%) and accuracy (84%), and the model performance is good. From the confusion matrix we can see that the neural network also does not solve the problem of lower PRECISION and higher false positives, and the improvement in the model's effectiveness is that he further reduces the false negative rate.

2.2.3. Xgboost

Logistic regression has better explanatory properties, but its predictive performance is usually not as good as complex nonlinear models; neural networks have better performance in predicting whether a customer has churned or not, but the black-box nature of the model makes it difficult to account for the effects of covariates.In order to find a balance between explanatory and predictive performance, we consider the use of nonparametric model decision trees.

A decision tree model is a model that does not rely on statistical assumptions or parametric assumptions and reflects only the characteristics of the data features themselves [12]. Simply put, decision trees infer the values of response variables by learning simple decision rules from data features, and their model structure is easy to understand and easy to visualize, thus enabling the identification of covariates that play an important role in the decision-making process. However, affected by the irreversibility and randomness of the algorithm's learning rules, the decision tree model can easily fall into local optimal solutions. In recent years, several methods have been used to improve decision trees, such as Bagging [13], Random Forest [14], AdaBoost [15], Gradient Boosting [16], XGBoost [17].The basic idea of these methods is the same: if a decision tree tends to fall into a local optimum, then train more trees, and then vote on the results of all the trees, and take the one that most of the trees agree on. And Bagging and Boosting methods are somewhat different, bagging is to resample the data several times and then train the decision tree on different sub-datasets, while boosting method is to train multiple trees on the whole dataset step by step.

DOI: 10.54254/2754-1169/2025.31298

We train and test the original data, over-sampled data and under-sampled data with different tree models to get the recall values of different methods under different data and visualize them as shown in Figure 10.

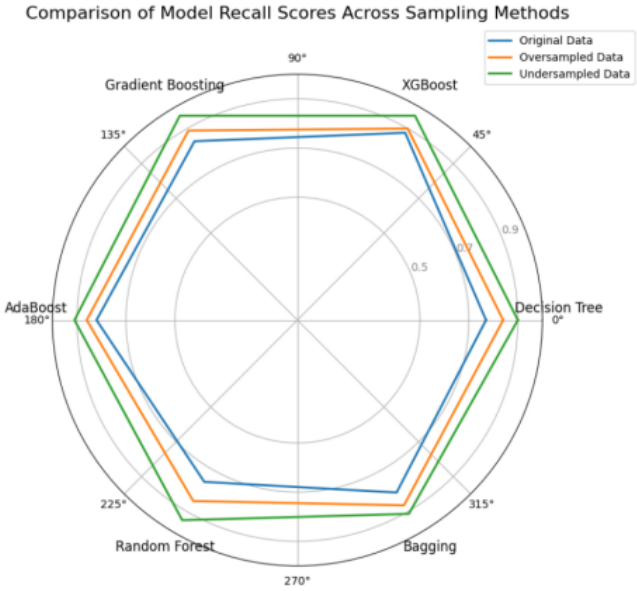

Figure 10. Recall rates of different models under various sampling methods

As can be seen from Fig. 10, the XGBoost method performs relatively better in terms of recall under different sampling methods, and especially shows the best generalization ability on the undersampled dataset. Therefore, in this study, the XGBoost model will be preferred for training on undersampled datasets to further improve the prediction of customer churn. The XGBoost model, developed by Li and Zhang and Chen and Guestrin, is an extremely robust machine learning algorithm known for its excellent utility and efficient execution speed. First, we define the XGBoost model, setting the evaluation metric as log-loss. Subsequently, a hyperparameter grid is defined that covers parameters such as the learning rate, the number of trees, the maximum depth of each tree, the learning rate, the proportion of samples used to train each tree, the percentage of features randomly sampled during the training of each tree, the minimum loss reduction required to divide the nodes of the control tree, the minimum sum of the sample weights in the control sub-nodes, the L1 regularization coefficient, and the L2 regularization coefficient. Each parameter has a set of possible values so that Randomized SearchCV can perform a random search on these parameters to find an optimal set of hyperparameters. The goal of improving the model performance is then achieved by minimizing the objective function. The objective function can be written as:

$$Obj = \sum_{i=1}^{n} \left[y_i \log\left(\widehat{y_i}\right) \log\left(1 - \widehat{y_i}\right) \right] + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^{T} \omega_j^2 \tag{3}$$

Where n is the number of samples, is the true label, is the probability predicted by the model, T is the number of leaf nodes, is the weight of the jth leaf node, and is the regularization parameter [18].

The training results of the above model are shown in Figure 11. Through Figure 11, it can be found that the accuracy and recall of this model have reached 95%, so this model is chosen as the final model. From the confusion matrix we can see that the XGboost model can reduce the false positive rate very well, while the false negative rate is also well controlled, which can help the bank

DOI: 10.54254/2754-1169/2025.31298

to identify both the real churn-prone customers as well as the non-churn-prone customers, thus saving the cost.

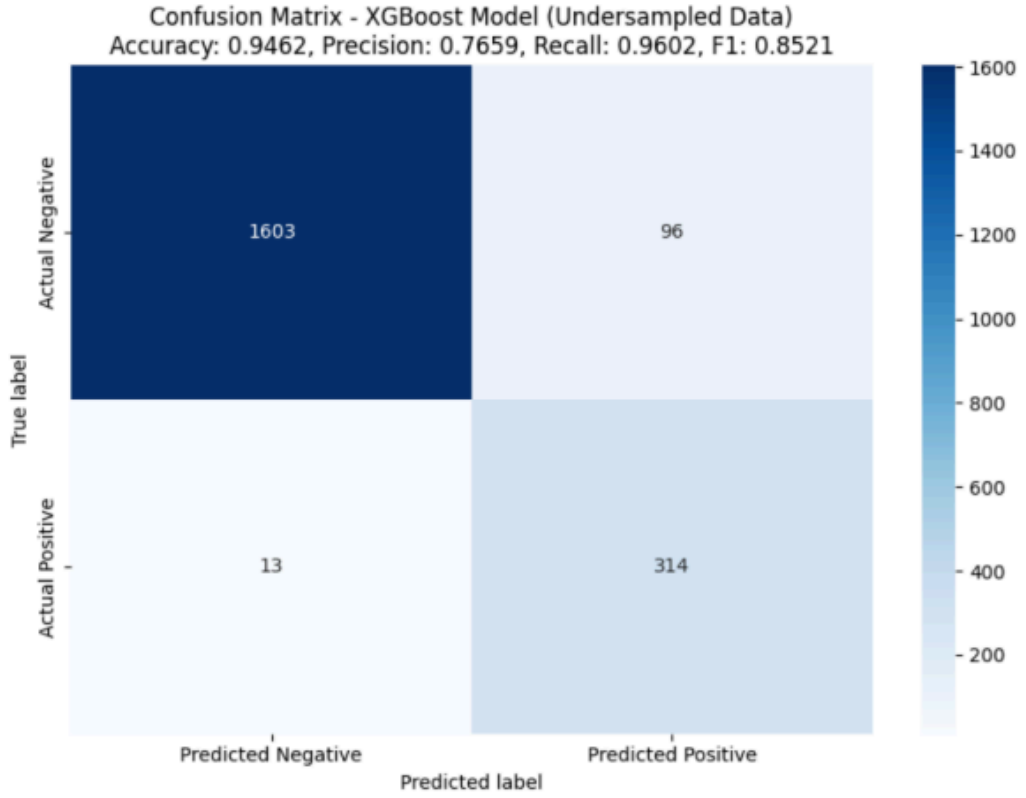

Figure 11. Confusion matrix of the XGBoost model based on undersampled data

In machine learning, we often identify which feature contributes the most to the prediction effect based on feature importance. Feature importance outputs a score or metric that is used to rank features in order of how much they contribute to the model's predictions [19]. In order to better understand the impact of each feature on customer churn and to compare it with the logistic regression results, we use this model to measure the importance of each feature. SHAP is a model-independent feature selection mechanism that assigns importance to features based on their contribution to the model "output", and has been shown to outperform the other common feature selection mechanisms. SHAP is a model-independent feature selection mechanism that assigns importance to features based on their contribution to the model "output", and has been shown to outperform other common feature selection mechanisms [20]. Therefore, this study further employs SHAP to interpret the XGBoost model constructed based on undersampled data. The SHAP interpreter is created by calling the Tree Explainer class and is used to compute the SHAP value for each feature in the training set on each sample. Then, the SHAP values of each feature are taken as absolute values and averaged to quantify the magnitude of their contribution to the model predictions. Finally, the contribution and impact of each feature is visualized and the results are shown in Figure 12.

DOI: 10.54254/2754-1169/2025.31298

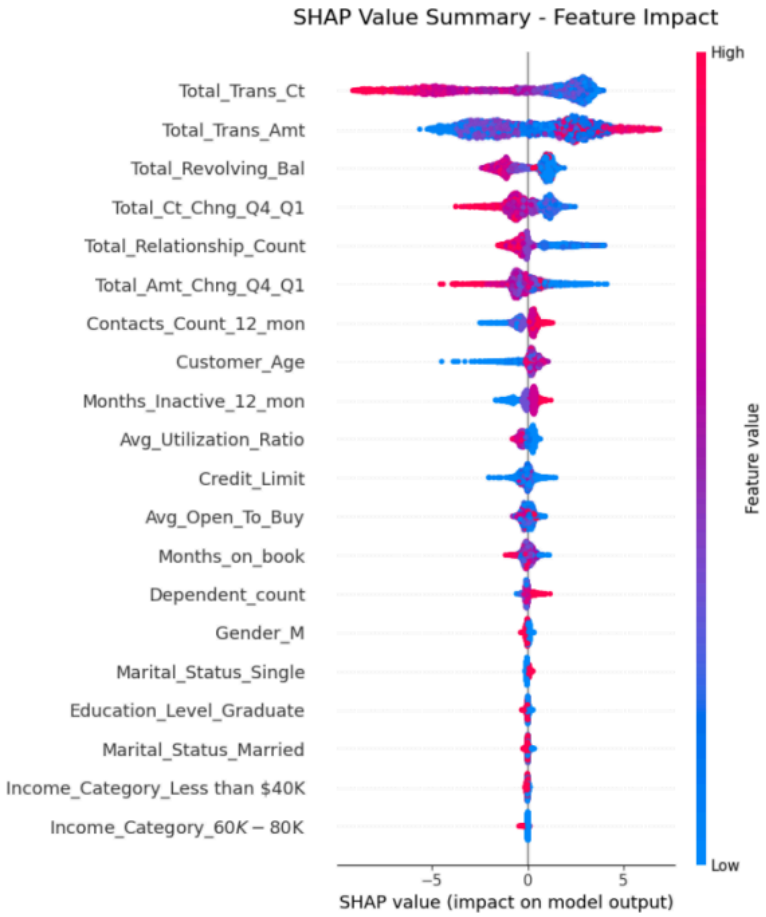

Figure 12. SHAP value summary-feature impact

From Figure 12, it can be found that the total number of transactions made by customers using the bank's credit card in the past 12 months is the feature that has the greatest impact on customer churn, the higher the total number of transactions, the less likely that customers will be churned out; the impact of the total amount of transactions is in the second place, the higher the total amount of transactions, the more likely that customers will be churned out. This is generally consistent with the results of the logistic regression analysis. In addition, the figure shows that there are other factors that also have a greater impact on the churn rate of credit card customers. For example, the higher the total revolving balance, the less likely customers are to churn. The higher the total revolving balance, which is the amount of money owed on a credit card that continues to accrue interest on a rolling basis without being paid in full, the more reliant the customer is on the credit card. At the same time, customers with high revolving balances tend to enjoy installment payments or other benefits, and dropping a bank's credit card may mean that they need to pay off the balance or give up the benefits in one lump sum, which raises the cost of switching cards, so they are less likely to "churn". The impact of the number of total products a customer holds with that bank cannot be ignored. A higher number of total products means that the customer may be involved in a wide range of businesses such as savings accounts, wealth management products, loans, etc., in addition to owning a credit card from that bank, with a wide range of deep business connections. At this point, when a customer wants to give up the bank's credit card, he or she may be more cautious about making a move to change credit card providers, taking into account the adverse impact this will have

DOI: 10.54254/2754-1169/2025.31298

on his or her other business with the bank, enjoyment of privileges, and so on. As a result, customers who hold a greater number of the bank's total products are also less likely to be lost.

3. Conclusion

In this study, we investigate the bank credit card customer churn problem using various machine learning algorithms, including logistic regression, feed-forward neural networks, and XGboost. By systematically comparing the models on two dimensions: Accuracy and Recall and Model Explanatory Power, we find that the logistic regression model trained on undersampled data has achieved better prediction of customers who are actually about to churn, but it is easy to misclassify customers who will not churn as "about to churn", i.e., higher false positives. ", i.e., higher false positives. This will not cause banks to lose credit card customers, but it will increase their customer retention costs. Meanwhile, logistic regression has good explanatory performance. The performance of neural network model is further improved on the basis of the former, but it still fails to solve the problem of high false positives, and it is also difficult to explain the impact of covariates on customer churn due to its "black box characteristics". The XGboost model, on the other hand, is able to achieve accurate prediction of customer churn with low false positives and false negatives after being trained using a multidimensional undersampled dataset. The model is also able to effectively identify the key factors affecting customer churn by extracting feature importance, demonstrating the potential of the model in bank credit card customer churn prediction.

From the results of the logistic regression and XGboost machine learning models for the interpretation of the effects of covariates, the following conclusions can be drawn: both models agree that the total number of transactions a customer has made with the bank's credit card in a 12-month period is the most important factor influencing the prediction of churn, and that the higher the total number of transactions the less likely a credit card customer will be churned. Second, the larger the total amount of transactions a customer has made with the bank's credit card in a 12-month period, the more likely the customer is to be churned. In addition, factors such as changes in total revolving balances, the number of transactions, and the total number of products held by the customer with that bank can also have a significant impact on churn rates. Therefore, based on the model results, the bank needs to pay extra attention to customers who have a low total number of transactions using the bank's credit card in a 12-month period or a high total amount of transactions using the bank's credit card in a 12-month period, as they are more likely to churn. Customers with lower total revolving balances, lower changes in the number of transactions, lower total product holdings, and a higher number of contacts with the bank in the past 12 months also tend to churn, so watch out for that as well.

Acknowledgement

We would like to express our special thanks to Prof. Cosimo Arnesano for his comprehensive guidance on the theoretical and practical aspects of machine learning, which has enabled us to gain a deeper understanding of, and effectively apply, machine learning-related knowledge. We would also like to thank Ms. Zhong Chen Ye and Mr. Deng Yanqi for their valuable support and advice during the project, which helped us to advance the research smoothly. It has been a pleasure to work and discuss with all of you, which has not only broadened our academic horizons, but also benefited us in the field of machine learning.

DOI: 10.54254/2754-1169/2025.31298

References

[1] Liang, Z. (2023). Predict Customer Churn based on Machine Learning Algorithms. Highlights in Business, Economics and Management, 10, 270-275.

[2] Mehnaz, Jin, J., Hussain, A., Warraich, M. A., & Waheed, A. (2024). Impact of perceived CSR practices on customers loyalty. The mediating role of reputation and customer satisfaction. Corporate Social Responsibility and Environmental Management, 31(5), 3724-3734.

[3] Bahel, V., Pillai, S., & Malhotra, M. (2020, June). A comparative study on various binary classification algorithms and their improved variant for optimal performance. In 2020 IEEE Region 10 Symposium (TENSYMP) (pp. 495-498). IEEE.

[4] Peng, K., Peng, Y., & Li, W. (2023). Research on customer churn prediction and model interpretability analysis. Plos one, 18(12), e0289724.

[5] Siddiqui, N., Haque, M. A., Khan, S. S., Adil, M., & Shoaib, H. (2024). Different ML-based strategies for customer churn prediction in banking sector. Journal of Data, Information and Management, 6(3), 217-234.

[6] Juba, B., & Le, H. S. (2019, July). Precision-recall versus accuracy and the role of large data sets. In Proceedings of the AAAI conference on artificial intelligence (Vol. 33, No. 01, pp. 4039-4048).

[7] Mccullagh, P., & Nelder, J. A. (1989). Generalized linear models.

[8] Thiese, M. S., Ronna, B., & Ott, U. (2016). P value interpretations and considerations. Journal of thoracic disease, 8(9), E928.

[9] Wang, L., Han, M., Li, X., Zhang, N., & Cheng, H. (2021). Review of classification methods on unbalanced data sets. Ieee Access, 9, 64606-64628.

[10] Hemeida, A. M., Hassan, S. A., Mohamed, A. A. A., Alkhalaf, S., Mahmoud, M. M., Senjyu, T., & El-Din, A. B. (2020). Nature-inspired algorithms for feed-forward neural network classifiers: A survey of one decade of research. Ain Shams Engineering Journal, 11(3), 659-675.

[11] Prechelt, L. (2002). Early stopping-but when?. In Neural Networks: Tricks of the trade (pp. 55-69). Berlin, Heidelberg: Springer Berlin Heidelberg.

[12] Song, Y. Y., & Ying, L. U. (2015). Decision tree methods: applications for classification and prediction. Shanghai archives of psychiatry, 27(2), 130.

[13] Dietterich, T. G. (2000). An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. Machine learning, 40, 139-157.

[14] Ho, T. K. (1995, August). Random decision forests. In Proceedings of 3rd international conference on document analysis and recognition (Vol. 1, pp. 278-282). IEEE.

[15] Freund, Y., & Schapire, R. E. (1996, July). Experiments with a new boosting algorithm. In icml (Vol. 96, pp. 148-156).

[16] Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. Annals of statistics, 1189-1232.

[17] Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).

[18] Oukhouya, Hassan; Kadir, Hamza; Himdi, Khalid El; Guerbaz, Raby.Forecasting international stock market trends: XGBoost, LSTM, LSTM-XGBoost, and backtesting XGBoost models [J].Statistics, Optimization & Information Computing, 2024, Vol.12(1): 200-209

[19] Musolf, A. M., Holzinger, E. R., Malley, J. D., & Bailey-Wilson, J. E. (2022). What makes a good prediction? Feature importance and beginning to open the black box of machine learning in genetics. Human Genetics, 141(9), 1515-1528.

[20] Marcílio, W. E., & Eler, D. M. (2020, November). From explanations to feature selection: assessing SHAP values as feature selection mechanism. In 2020 33rd SIBGRAPI conference on Graphics, Patterns and Images (SIBGRAPI) (pp. 340-347). Ieee.