
Research article

Prediction of bank credit customers churn based on machine learning and interpretability analysis

Ying Li¹ and Keyue Yan^{2,*}

¹ College of Global Talents, Beijing Institute of Technology (Zhuhai), Zhuhai, China

² Choi Kai Yau College, University of Macau, Macau, China

* **Correspondence:** Email: conneryan@foxmail.com.

Abstract: Nowadays, traditional machine learning methods for building predictive models of credit card customer churn are no longer sufficient for effective customer management. Additionally, interpreting these models has become essential. This study aims to balance the data using sampling techniques to forecast whether a customer will churn, combine machine learning methods to build a comprehensive customer churn prediction model, and select the model with the best performance. The optimal model is then interpreted using the Shapley Additive exPlanations (SHAP) values method to analyze the correlation between each independent variable and customer churn. Finally, the causal impacts of these variables on customer churn are explored using the R-learner causal inference method. The results show that the complete customer churn prediction model using Extreme Gradient Boosting (XGBoost) achieved significant performance, with accuracy, precision, recall, F1 score, and area under the curve (AUC) all reaching 97%. The SHAP values method and causal inference method demonstrate that several variables, such as the customer's total number of transactions, the total transaction amount, the total number of bank products, and the changes in both the amount and the number of transactions from the fourth quarter to the first quarter, have an impact on customer churn, providing a theoretical foundation for customer management.

Keywords: customer churn; machine learning; SHAP; causal inference; R-learner

JEL Codes: C12, C53, C63

1. Introduction

In the financial market, risk management has become one of the key elements of banking. Credit card customer churn prediction plays an important role in the risk management of the banking industry, assisting banks in identifying potential risks and implementing suitable measures to mitigate them

(Butaru et al., 2016). With the development of technology, the influence of customer churn prediction has extended into various fields, including the financial industry. In particular, the rapid progression of machine learning theories and algorithms has offered a fresh perspective and solution for customer churn prediction (Alfaiz and Fati, 2022). Firstly, machine learning technology is capable of automatically extracting useful information from large amounts of historical data, identifying distribution patterns and relationships to predict future trends and behaviors (Pudjihartono et al., 2022). In credit card customer churn prediction, machine learning methods can efficiently identify key variables and underlying relationships related to customer attrition. By analyzing the mass multi-dimensional data, including consumer behavior, credit history, and demographic information of historical users, machine learning can construct an effective prediction model. Secondly, machine learning technology has strong generalization ability, automatically adapting to new and unseen data after training without frequent adjustments and optimizations (Janiesch et al., 2021). Bank customer churn prediction using machine learning methods presents both challenges and opportunities. Through comprehensive studies and the application of machine learning methods, prediction models provide more precise and reliable results, thereby enhancing the accuracy and reliability of risk management standards in the credit card business and supporting the sustainable development of banking.

Despite the significant advantages of machine learning in customer churn prediction, its practical implementation still encounters several challenges. As datasets grow larger and models become more complex, ensuring model interpretability becomes a crucial concern (Chen and Meng, 2020). The SHAP values method is applied to interpret the outcomes of machine learning models. It quantifies the importance of each variable in the dataset, considering both the independent influence of each variable on prediction outcomes and the interactions among different variables. The SHAP values method provides a comprehensive explanation of machine learning model predictions (Gebreyesus et al., 2023). Moreover, the causal inference method has gained significant popularity in recent years for data analysis and machine learning-related research. By analyzing causal effects among variables, the causal inference method excludes variables that are related to prediction results but lack actual causal connections, and filters out variables with a causal impact on prediction results. This method provides a more reliable basis for variable selection in predictions based on machine learning (Li et al., 2023). The purpose of this study is to discuss the construction of credit card customer churn prediction models by applying machine learning techniques. Additionally, the SHAP values method is utilized to complete variable selection based on the importance of influencing the outcomes and to interpret machine learning-based prediction results. The causal inference method aims to explain the prediction results by analyzing the causal relationships between variables. These causal relationships are compared with the results obtained by the SHAP values method.

The remainder of this study is organized as follows. Section 2 reviews research related to machine learning, interpretability analysis, and causal inference. Section 3 describes the concepts of four sampling techniques for data preprocessing, six machine learning methods for credit card customer churn prediction, and the causal inference method in detail. Section 4 summarizes the performance of the machine learning models, interprets the optimal model, and identifies the causal relationships by analyzing the experimental results. Finally, Section 5 presents the conclusion and describes limitations and further research.

2. Literature review

With the development of computer hardware in recent years, technologies such as machine learning and deep learning are now widely used across various industries. Machine learning can be combined with stock price prediction to construct new quantitative investment strategies in financial investment (Yan et al., 2023). In the field of stock market risk, traditional machine learning models and neural networks can also predict Barrier Option prices with relative accuracy (Li and Yan, 2023). Even Bitcoin, a cryptocurrency that has been widely discussed in recent years, can have its price predicted by machine learning methods, with the Support Vector Regression model outperforming other models (Erfanian et al., 2022). In the area of credit card customer churn prediction, there is a growing number of reliable research papers aiming to predict churn rates by analyzing the personal and behavioral data of credit card customers, enabling banks to take effective measures to retain customers in advance. For instance, Zhang et al. used simple preprocessing of customer data and applied only the Random Forest (RF) model, a traditional machine learning method, to classify customers, achieving better classification results (Zhang et al., 2024). In the study by de Lima Lemos et al., several traditional machine learning models were used for prediction, including RF, Decision Tree, k-Nearest Neighbor, Elastic Net, Logistic Regression, and Support Vector Machines, with the RF model achieving the best classification accuracy at 82.8% (de Lima Lemos et al., 2022). Lalwani et al. introduced more advanced ideas into the data preprocessing problem, using the Gravitational Search Algorithm to select variables and improve the efficiency of the machine learning models (Lalwani et al., 2022). Siddiqui et al. used three variable selection methods to observe the performance of different machine learning models by comparing all variables, separating continuous and discrete variables, and selecting variables based on their importance. They ultimately found that using all variables yielded the most prevalent classification results (Siddiqui et al., 2024).

In addition to focusing on constructing different models to improve classification efficacy, researchers have gradually paid more attention to the interpretability of the models and the relationships between variables. In the field of financial investment, some articles focus on the impact of each variable on volatility risk (Yan and Li, 2024). In the study by AL-Najjar et al., they demonstrated the variable importance of the same traditional machine learning model, C5 Tree, under different methods of variable selection. The results showed that the three main variables affecting the model's effectiveness were the total number of transactions, the total credit card revolving balance, and the change in the number of transactions, providing useful references for credit card managers (AL-Najjar et al., 2022). Other studies discuss variables using more advanced and understandable variable visualization tools, such as the SHAP values method. In Peng et al.'s study, the best model for classification, which combined a genetic algorithm with XGBoost, was used to analyze interpretability. The results of the SHAP values method not only showed the order of importance of different independent variables, but also indicated whether changes in the value of each independent variable had a positive or negative effect on the predicted variable (Peng et al., 2023). In addition to the SHAP values method, other interpretable tools for visualization with similar effects, such as Local Interpretable Model-Agnostic Explanations (LIME), can also be used for analysis (Chang et al., 2024).

So far, machine learning algorithms have primarily captured correlations between variables, and inferring their importance from a correlation perspective, often ignoring the causal relationships among them. In other words, the model's determination of variable importance is frequently based on the

strength of the correlation between the dependent and independent variables, which can introduce bias errors in actual predictions (Feuerriegel et al., 2024). To address this issue, this study innovatively uses the R-learner in meta learners to analyze variable importance. This approach combines the fields of machine learning and causal inference (Künzel et al., 2019). The R-learner employs the control variable method to observe how independent variables affect the probability of credit card customer churn and calculates a more accurate conditional average treatment effect (CATE) by removing bias from the data. This helps quantify the effect of different variables on the probability of credit card customer churn.

3. Methodology

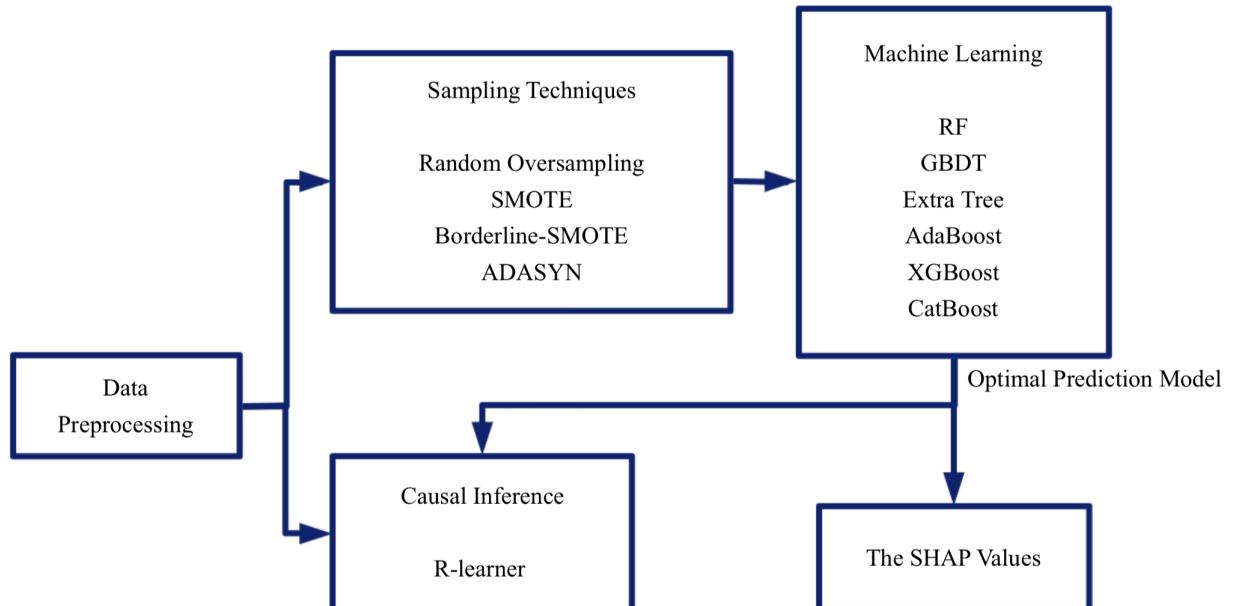


Figure 1. The research framework diagram

The aim of this study is to predict whether bank credit card customers will churn using machine learning methods and to utilize the prediction results to construct models that explain the influencing factors of customer churn. This will provide an effective basis for rationalizing customer management before they leave the bank and for effective risk management in the banking industry. Initially, the relevant dataset of bank customer churn needs to be downloaded and preprocessed. Subsequently, the approach is divided into two parts. The first part involves balancing the class distribution of the training set using four sampling techniques: Random Oversampling, Synthetic Minority Oversampling Technique (SMOTE), Borderline-Synthetic Minority Oversampling Technique (Borderline-SMOTE), and Adaptive Synthetic Sampling (ADASYN). Then, six machine learning methods, including RF, Gradient Boosting Decision Tree (GBDT), Extra Tree, AdaBoost, XGBoost, and CatBoost, are used to predict whether the bank will lose credit card customers. The optimal model is selected by comparing the performance of each model, and the important variables influencing customer churn and their effects are analyzed using the SHAP values method. The second part involves using the causal inference method to investigate the causal impact of variables on customer churn based on the optimal prediction

model mentioned above. In this part, the R-learner is used for continuous variables in the meta learners of the causal inference method. The framework of this study is outlined in Figure 1.

3.1. Data preprocessing

Table 1. The summary of categorical variables.

Categorical Variables	Class	Number of classes	Conversion to numbers
Attrition_Flag	Existing Customer	8500	0
	Attrited Customer	1627	1
Gender	M	4769	0
	F	5358	1
Education_Level	Uneducated	1487	6
	High School	2013	15
	College	1013	18
	Graduate	3128	22
	Post-Graduate	516	24
	Doctorate	451	28
	Unknown	1519	22
Marital_Status	Single	3943	0
	Married	4687	1
	Divorced	748	2
	Unknown	749	1
Income_Category	Less than \$40K	3561	2
	\$40K - \$60K	1790	5
	\$60K - \$80K	1402	7
	\$80K - \$120K	1535	10
	\$120K +	727	14
	Unknown	1112	2
Card_Category	Blue	9436	0
	Silver	555	1
	Gold	116	2
	Platinum	20	3

Historical data of bank customers downloaded from Kaggle is used as the dataset in this study (Kaggle, 2022). The dataset contains twenty three variables, including customers' fundamental personal details, the credit card level they hold, and their credit card usage, totaling 10,127 samples. The variable labeled "CLIENTNUM" represents the unique identification number assigned by the bank to each customer. Since this variable has no effect on the model construction in this study, it is removed from the analysis. Additionally, the last two variables in the dataset represent the outcomes provided by Kaggle for predicting bank customer churn based on Naive Bayes classifiers. These variables have been excluded during the data processing as this study does not compare this probabilistic classification

approach with machine learning methods. The remaining twenty variables, which consist of fourteen quantitative variables and six categorical variables, are all utilized in the experiment in this study to discuss the importance and impact. These variables represent customers' personal information and credit card usage. To build the model, it is necessary to convert the classes of categorical variables into numerical values. These six categorical variables are: "Attrition_Flag", "Gender", "Education_Level", "Marital_Status", "Income_Category", and "Card_Category". For ordered categorical variables, the classes need to be sorted from low to high before conversion. One type of class encountered in the variables "Education_Level", "Marital_Status", and "Income_Category" is "Unknown". There are 3,046 samples that meet the criterion of having at least one of these three variables labeled as "Unknown", constituting 30% of the total sample size. The treatment for "Unknown" in any of these three variables is to change it to the majority class of the respective variable. The details of these six categorical variables are outlined in Table 1.

The variable "Marital_Status" is an unordered categorical variable with more than two distinct values, requiring conversion into dummy variables.

3.2. Data balancing

The variable "Attrition_Flag" indicates whether a customer has churned or not, serving as the target variable of this study. The dataset contains 1,627 samples in the "Attrited Customer" class. The ratio of these samples to those in the "Existing Customer" class is greater than 5:1, indicating a significant class imbalance, as shown in Figure 2.

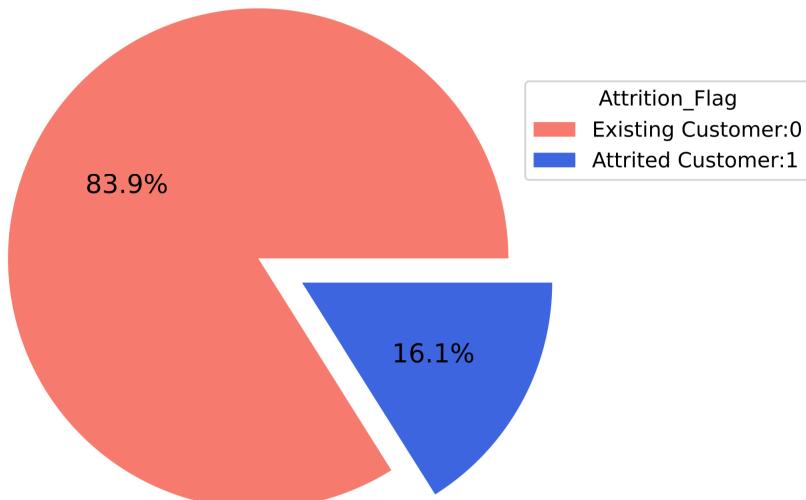


Figure 2. The proportion of classes for "Attrition_Flag"

If the training set is used by models without preprocessing, it will decrease the ability of the machine learning models to recognize churned customers, leading to overfitting in the results and reducing prediction accuracy (Alam et al., 2020). To avoid these problems and improve the performance of the model, improved oversampling has been adopted to balance the class distribution, which involves increasing the number of samples from the minority class. In this study, the following four sampling algorithms are used.

Random oversampling is a straightforward method that randomly selects and replicates samples from the minority class until the classes are balanced.

SMOTE is an improved algorithm for handling imbalanced data based on random oversampling. The first step involves calculating the distance between each sample from the minority class and identifying the k nearest neighbors of the same class. Then, a specified number of samples are randomly selected from these nearest neighbors. A new synthetic sample is generated between the minority class sample and the selected nearest neighbor, lying on the line connecting these two samples, thereby increasing the number of minority class samples (Akin, 2023).

Borderline-SMOTE is an improved version of the SMOTE algorithm. The key difference is an additional preliminary step: a minority class sample is selected to apply the SMOTE algorithm if most of its k nearest neighbors are from the majority class (Gu et al., 2023).

ADASYN is another common oversampling method similar to SMOTE. The key to this algorithm is calculating the ratio of samples between different classes for each minority class sample and using this ratio distribution to determine the number of synthetic samples generated for each minority class sample (Dube and Verster, 2023).

3.3. Machine learning evaluation

This study employs six machine learning methods to predict credit card customer churn. In addition to RF and AdaBoost, which were used in the previous research to predict barrier option prices (Li and Yan, 2023), this study utilizes the extra tree algorithm and three other prevalent boosting algorithms: GBDT, XGBoost, and CatBoost. To enhance the performance of these machine learning models, the grid search method is used to tune the hyperparameters of each model and find the optimal set. A complete customer churn prediction model can be constructed by combining one of the sampling techniques mentioned above with one of these six machine learning methods. The prediction will result in four possible cases:

- True positive (TP): The predicted result is positive, and the actual value is also positive.
- True negative (TN): The predicted result is negative, and the actual value is also negative.
- False positive (FP): The predicted result is positive, whereas the actual value is negative.
- False negative (FN): The predicted result is negative, whereas the actual value is positive.

TP and TN indicate the predicted results are consistent with the actual values, and FP and FN indicate the opposite. To evaluate the performance of models in binary classification problems, the following indicators are typically utilized: accuracy, precision, recall, F1 score, and AUC. The formulas for the first four indicators are defined as follows:

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

$$\text{precision} = \frac{TP}{TP + FP}, \quad (2)$$

$$\text{recall} = \frac{TP}{TP + FN}, \quad (3)$$

$$F1 = 2 \times \frac{precision \times recall}{precision + recall}. \quad (4)$$

The value of AUC is equal to the area under the receiver operating characteristic curve (ROC), which is the curve consisting of recall and false positive rate (FPR). The formula for FPR is as follows:

$$FPR = \frac{FP}{FP + TN}. \quad (5)$$

The closer the values of these five indicators are to 1, the better the model performs.

3.4. Causal inference

Causal inference is a method that effectively explores and analyzes whether a variable is the main factor affecting the target variable (Jiang, 2022). In traditional causal inference, the randomized controlled trial is generally considered as a reliable methodology to determine the influence of variables. However, in practice, due to the cost and ethical concerns of experiment, causal inference is often based on collected observational datasets. In Facure's book, which explores the combination of machine learning and causal inference, the authors propose a modeling approach called R-learner for continuous variables in a dataset to assess the causal significance of an independent variable on a dependent variable (Facure M, 2023). In current industrial applications, various code tools are becoming available for researchers to explore the feasibility of machine learning in causal inference problems (Molak A, 2023). To determine whether a variable has a causal relationship with customer churn and to quantify its impact, this study adopts the causal inference method to understand the causal effects among them.

In causality, the cause variable is the treatment denoted as T , while the dependent variable is the outcome denoted as Y . Additionally, the other independent variables are referred to as characteristic variables, denoted as X . For an individual sample i , the prediction model predicts the corresponding value Y_i as $Y_i(T = 1|X)$ when the treatment is applied (i.e., $T_i = 1$), and as $Y_i(T = 0|X)$ without the treatment (i.e., $T_i = 0$). Individual treatment effect (ITE) calculates the difference in outcomes for an individual between the scenario with treatment and the scenario without treatment. The formula of ITE is shown as follows:

$$ITE_i = Y_i(T = 1|X) - Y_i(T = 0|X). \quad (6)$$

Considering the overall causal effect between the treatment and the outcome, it needs to measure CATE, which is given by the following equation:

$$CATE = E[Y|T = 1, X] - E[Y|T = 0, X]. \quad (7)$$

The value of CATE indicates whether there exists a causal relationship between the treatment variable T on the outcome Y , and the quantification of its effect.

For binary classification problems, this study uses R-learner to estimate the causality and effect of different continuous treatment variables on the outcome Y respectively, based on the optimal customer churn prediction model. The CATE value in R-learner is obtained by calculating the probability of the same outcome occurring with and without the treatment, which is given by the following equation:

$$CATE^* = P(Y = 1|T = 1, X) - P(Y = 1|T = 0, X). \quad (8)$$

For each T , hypothesis testing is applied to CATE* to determine the causal impact of that treatment variable on Y .

4. Experimental results and analysis

The experimental environment for this study is Python, utilizing toolboxes such as imblearn, sklearn, xgboost, and catboost. Following the data preprocessing steps outlined in Section 3.1, the dataset is divided into a training set and a test set in an 8:2 ratio. The training set data is balanced using four sampling techniques, and the independent variables of all the data are normalized. Subsequently, six machine learning methods are trained on the training set data, resulting in twenty-four complete customer churn prediction models. These trained models are then applied to predict the “Attrition_Flag” values in the test set, and the performance of these models is evaluated using the indicators mentioned in Section 3.3.

The results of the accuracy for the training set, evaluated using six machine learning methods, range from 96.06% to 99.92% under random oversampling, 97.33% to 99.88% under SMOTE, 96.95% to 99.94% under Borderline-SMOTE, and 96.88% to 99.91% under ADASYN. The test set results are as follows.

Table 2. The evaluation results of prediction models in the test set.

Sampling techniques	Machine learning	accuracy	precision	recall	F1	AUC
Random Oversampling	RF	0.9650	0.9648	0.9650	0.9449	0.9901
	GBDT	0.9743	0.9746	0.9743	0.9745	0.9954
	Extra Tree	0.9620	0.9613	0.9620	0.9613	0.9890
	AdaBoost	0.9521	0.9568	0.9521	0.9535	0.9893
	XGBoost	0.9748	0.9751	0.9748	0.9749	0.9939
	CatBoost	0.9556	0.9606	0.9556	0.9570	0.9900
SMOTE	RF	0.9615	0.9627	0.9615	0.9620	0.9897
	GBDT	0.9699	0.9701	0.9699	0.9700	0.9945
	Extra Tree	0.9600	0.9596	0.9600	0.9597	0.9881
	AdaBoost	0.9556	0.9565	0.9556	0.9560	0.9856
	XGBoost	0.9704	0.9704	0.9704	0.9704	0.9919
	CatBoost	0.9590	0.9607	0.9590	0.9596	0.9884
Borderline-SMOTE	RF	0.9585	0.9583	0.9585	0.9588	0.9891
	GBDT	0.9709	0.9711	0.9709	0.9710	0.9936
	Extra Tree	0.9556	0.9549	0.9556	0.9551	0.9886
	AdaBoost	0.9516	0.9538	0.9516	0.9524	0.9855
	XGBoost	0.9719	0.9719	0.9719	0.9719	0.9934
	CatBoost	0.9497	0.9532	0.9497	0.9508	0.9857
ADASYN	RF	0.9605	0.9615	0.9605	0.9609	0.9901
	GBDT	0.9709	0.9710	0.9709	0.9709	0.9939
	Extra Tree	0.9566	0.9561	0.9566	0.9563	0.9885
	AdaBoost	0.9576	0.9589	0.9876	0.9581	0.9859
	XGBoost	0.9724	0.9724	0.9724	0.9724	0.9931
	CatBoost	0.9546	0.9575	0.9546	0.9556	0.9878

As shown in Table 2, the values of all indicators for the complete customer churn prediction models are above 0.94 after hyperparameter tuning and training, indicating good performance for each model. The difference in the evaluation results of the prediction models between the training set and the test set is not significant, and there is no instance of complete overfitting. For the random oversampling-XGBoost, SMOTE-XGBoost, Borderline-SMOTE-XGBoost, and ADASYN-XGBoost models, all indicator values exceed 0.97, suggesting that these four models perform better than the others. These four models can be considered the optimal prediction models. Furthermore, the XGBoost model performs better and more stably compared to other machine learning models across different sampling techniques by optimizing both the loss function and regularization term (Guo and Fan, 2024).

The SHAP values method is used to understand the contribution of each variable in the optimal prediction models and its relationship with the target variable better (Wu et al., 2024). When the SHAP value of a variable is positive, it indicates a positive relationship with customer churn, whereas a negative SHAP value indicates a negative relationship. By using the SHAP toolbox in Python, an interpretable explanation of the optimal prediction models is obtained, resulting in corresponding SHAP summary plots. The variables in the SHAP summary plots are sorted from highest to lowest according to their importance in each model, with the top 10 variables selected. The points corresponding to each variable represent their SHAP values.

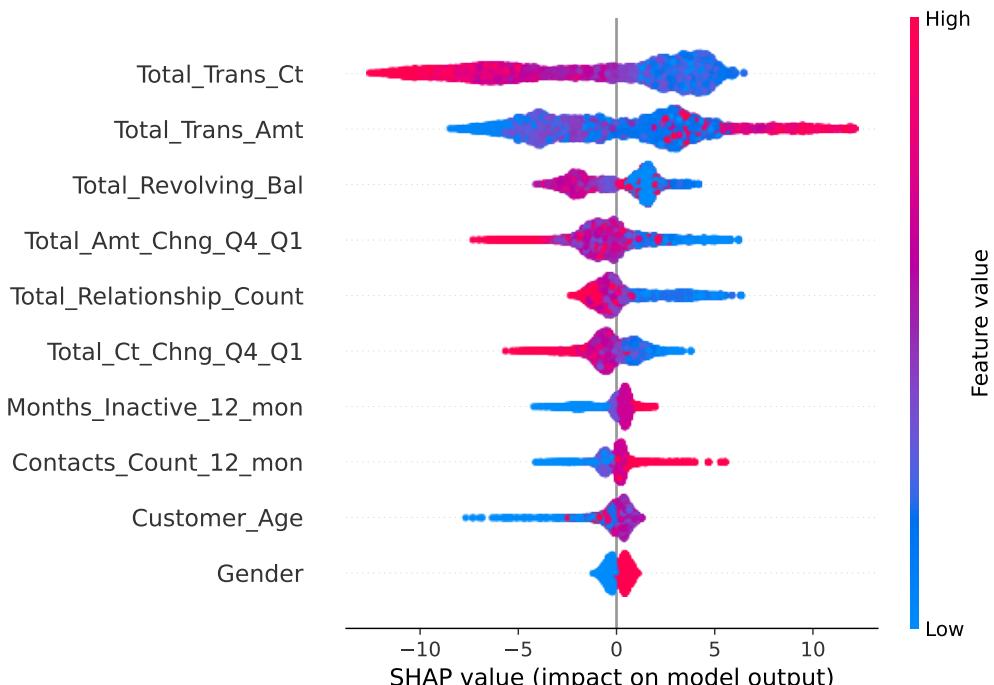


Figure 3. The SHAP summary plot of random oversampling-XGBoost model

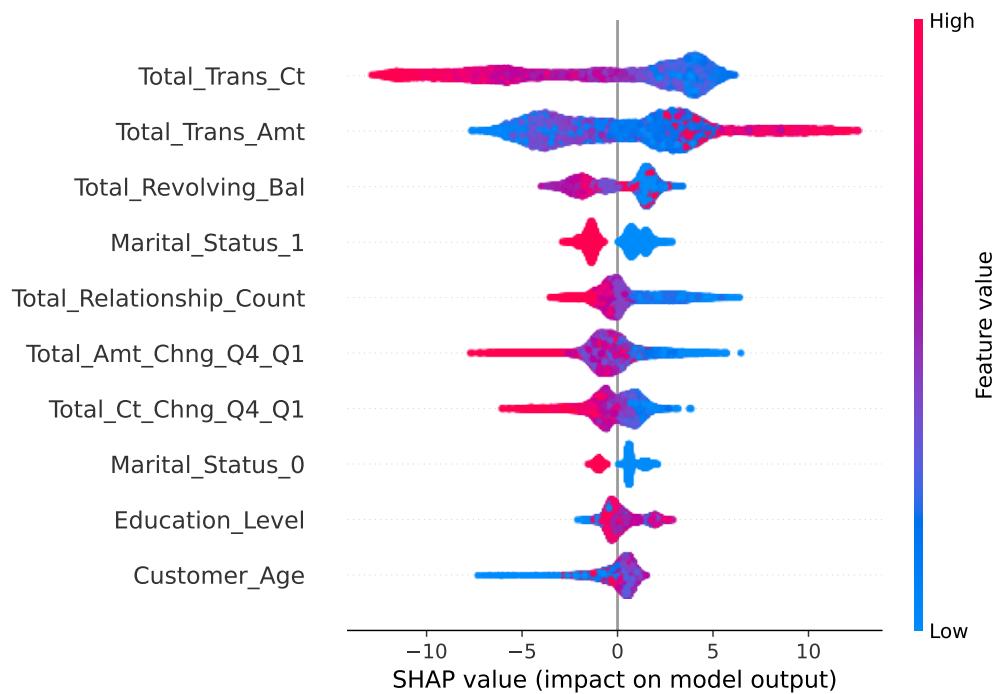


Figure 4. The SHAP summary plot of SMOTE-XGBoost model

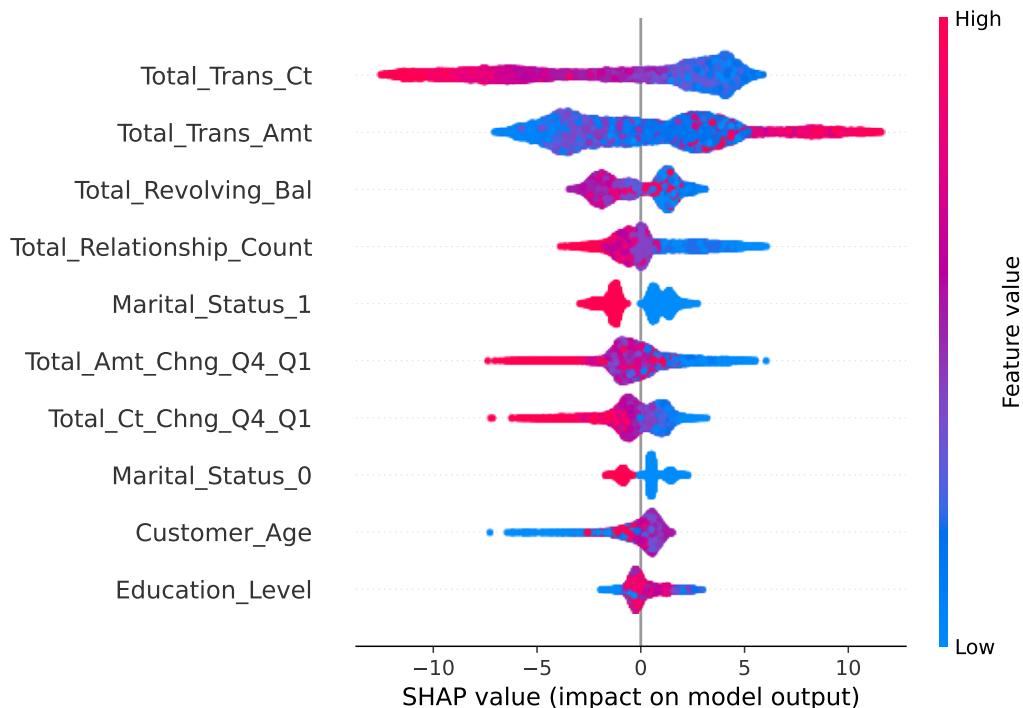


Figure 5. The SHAP summary plot of BorderlineSMOTE-XGBoost model

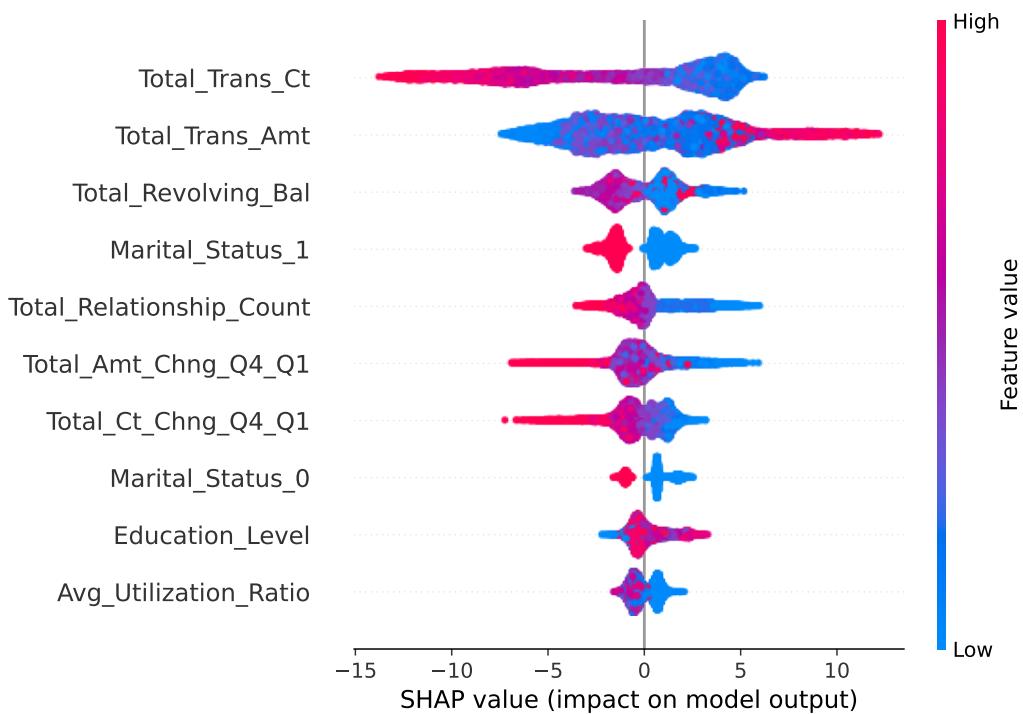


Figure 6. The SHAP summary plot of ADASYN-XGBoost model

Based on the four figures above, in any of the optimal prediction models, the variables “Total_Trans_Ct”, “Total_Trans_Amt”, and “Total_Revolving_Bal” consistently rank in the top three in terms of importance and in the same order. Moreover, the positive or negative relationship between each of these variables and “Attrition_Flag” remains unchanged. These three variables represent the customer’s total number of transactions with the bank in the past year, the total amount of transactions, and the total revolving balance of the credit card, respectively. As the total number of transactions increases, the SHAP value decreases, indicating a negative relationship with customer churn. This suggests that customers with a higher number of transactions are less likely to churn. For the second most important variable, the higher the total amount of transactions, the higher the SHAP value. When the total transaction amount is small, the SHAP value can be either positive or negative. There is a significant positive relationship with customer churn only when the total amount exceeds a certain threshold, indicating that customers are more likely to churn when their total transaction amount is above this threshold. For “Total_Revolving_Bal”, the greater the total revolving balance of the credit card, the lower the SHAP value, indicating that customers with a larger total credit card revolving balance are less prone to churn compared to those with smaller total credit card revolving balance.

Furthermore, for each optimal prediction model, the variables “Total_Relationship_Count”, “Total_Amt_Chng_Q4_Q1”, and “Total_Ct_Chng_Q4_Q1” also rank among the top 10 of the SHAP values, consistently appearing in the middle positions. These three variables represent the total number of bank products held by customers, the change in the amount from the fourth quarter to the first quarter, and the change in the number of transactions, respectively. For these three variables, as their values decrease, their SHAP values increase, leading to the predicted result tending closer to 1, which indicates a negative relationship with customer churn. This means that customers who hold a larger number of

products or have significant changes in amount and transaction frequency across different quarters are less likely to churn.

For the remaining variables in the SHAP summary plots, although some are important in most optimal prediction models, they pertain to customers' personal information. Therefore, it is not convenient to discuss and classify the attributes of variables such as "Customer_Age", "Marital_Status_0", "Marital_Status_1", and "Education_Level". Additionally, "Months_Inactive_12_mon" for the total number of months inactive in the past year, "Contacts_Count_12_mon" for the total number of contacts in the past year, and "Gender" are only among the top 10 in the random oversampling-XGBoost model. Among these, "Gender" is personal information, while the other two variables are positively related to customer churn. Furthermore, the variable "Avg_Utilization_Ratio" only appears in the ADASYN-XGBoost model, ranking 10th in importance. This variable represents the average utilization rate of bank credit cards, with a lower rate indicating a higher likelihood of customer churn.

The statsmodels toolbox in python is used to analyze the causal effects between variables without personal information of customers and customer churn in the SHAP summary plots. The XGBoost model is selected for estimating customer churn in R-learner due to its superior predictive performance. The experimental results of the causal inference method are listed as follows.

Table 3. The experimental results of R-learner.

Variables	coef	std err	t	$P > t $	95 % interval
Total_Trans_Ct	-0.0003	2.48e-05	-13.109	0.000	[-0.000,-0.000]
Total_Trans_Amt	7.163e-06	4.14e-07	17.308	0.000	[6.35e-06,7.97e-06]
Total_Revolving_Bal	-4.396e-06	5.8e-06	-0.758	0.449	[-1.58e-05,6.98e-06]
Total_Relationship_Count	-0.0008	0.000	-6.324	0.000	[-0.001,-0.001]
Total_Amt_Chng_Q4_Q1	-0.0079	0.001	-7.078	0.000	[-0.010,-0.006]
Total_Ct_Chng_Q4_Q1	-0.0051	0.001	-4.819	0.000	[-0.007,-0.003]
Months_Inactive_12_mon	0.0007	0.000	3.968	0.000	[0.000,0.001]
Contacts_Count_12_mon	0.0006	0.000	3.872	0.000	[0.000,0.001]
Avg_Utilization_Ratio	0.0039	0.010	0.384	0.701	[-0.016,0.024]

According to Table 3, the p-values for "Total_Revolving_Bal" and "Avg_Utilization_Ratio" are greater than 0.05, indicating no significant difference in customer churn caused by changes in these two treatment variables. Combined with the SHAP values analysis, this suggests that there is only a correlation between each of these two variables and customer churn, not a causal relationship. For the remaining treatment variables, the p-values are all less than 0.05, indicating that changes in each of these variables have causal relationships with customer churn.

Considering the coefficient values in Table 3, the treatment variables "Total_Trans_Ct", "Total_Relationship_Count", "Total_Amt_Chng_Q4_Q1", and "Total_Ct_Chng_Q4_Q1" have negative coefficients, indicating negative causal relationships with customer churn. Among these variables, "Total_Amt_Chng_Q4_Q1" has the largest causal effect, while "Total_Trans_Ct" has the smallest. Conversely, the coefficients of "Total_Trans_Amt", "Months_Inactive_12_mon", and "Contacts_Count_12_mon" are positive, indicating positive causal relationships with customer churn. "Months_Inactive_12_mon" displays the largest causal effect, while "Total_Trans_Amt" displays the smallest. Compared with the

results of the SHAP values method, the similarity is that these seven variables have the same direction of impact on customer churn, while the difference lies in the order of the quantified effect values.

5. Conclusions

In the current research, to enhance the accuracy of prediction results, a combination of sampling techniques and machine learning models was employed to forecast customer churn in banks. A comparative performance analysis indicates that the XGBoost model consistently outperforms other machine learning models, achieving an accuracy of at least 97%, regardless of the sampling techniques used. Furthermore, the SHAP values method was utilized to interpret the optimized prediction models, while R-learner was used to investigate the causal effects of these variables on customer churn. Based on these two methods, the main important variables affecting customer churn, which include the total number and amount of transactions with the bank in the past year, the total number of bank products held by the customer, and the changes in the amount and number of transactions from the fourth quarter to the first quarter, were identified. Additionally, the analysis found that the total credit card revolving balance does not have a significant causal relationship with customer churn, but there is a strong correlation. The research findings provide valuable recommendations for bank managers to improve customer management strategies.

Due to the limited number of minority class samples in the dataset, the experiment requires sampling techniques to generate synthetic samples. This approach enables the prediction model to more accurately identify the categories of samples. Furthermore, excluding the variables belonging to customers' personal information, the other variables consist of cross-sectional data. They are typically utilized in the analysis that emphasize the differences among individual samples, rather than focusing on changes within a sample over time. Therefore, if additional samples with more extensive variables, such as time series data, are available the feasibility of the model's predictions, so that a better comparison of the interaction between the SHAP values method and the causal inference method in the further research.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this study.

Acknowledgments

The authors have received no financial assistance from any source in the preparation of this study.

Conflict of interest

All authors declare no conflicts of interest in this study.

References

- Akin P (2023) A new hybrid approach based on genetic algorithm and support vector machine methods for hyperparameter optimization in synthetic minority over-sampling technique (SMOTE). *AIMS Math* 8: 9400–9415. <https://doi.org/10.3934/math.2023473>

-
- Alam TM, Shaukat K, Hameed IA, et al. (2020) An Investigation of Credit Card Default Prediction in the Imbalanced Datasets. *IEEE Access* 8: 201173–201198. <https://doi.org/10.1109/access.2020.3033784>
- Alfaiz NS, Fati SM (2022) Enhanced Credit Card Fraud Detection Model Using Machine Learning. *Electronics* 11: 662. <https://doi.org/10.3390/electronics11040662>
- AL-Najjar D, Al-Rousan N, AL-Najjar H (2022) Machine learning to develop credit card customer churn prediction. *J Theor Appl El Comm* 17: 1529–1542. <https://doi.org/10.3390/jtaer17040077>
- Bank Churners (2022) Kaggle. Available from: <http://www.kaggle.com/competitions/bank-churners/overview>.
- Butaru F, Chen Q, Clark B, et al. (2016) Risk and Risk Management in the Credit Card Industry. *J Bank Financ* 72: 218–239. <https://doi.org/10.1016/j.jbankfin.2016.07.015>
- Chang V, Hall K, Xu Qa, et al. (2024) Prediction of Customer Churn Behavior in the Telecommunication Industry Using Machine Learning Models. *Algorithms* 17: 231. <https://doi.org/10.3390/a17060231>
- Chen K, Meng X (2020) Interpretation and Understanding in Machine Learning. *J Comput Res Dev* 57: 1971–1986. <https://doi.org/10.7544/issn1000-1239.2020.20190456>
- de Lima Lemos RA, Silva TC, Tabak BM (2022) Propension to customer churn in a financial institution: a machine learning approach. *Neural Comput Appl* 34: 11751–11768. <https://doi.org/10.1007/s00521-022-07067-x>
- Dube L, Verster T (2023) Enhancing classification performance in imbalanced datasets: A comparative analysis of machine learning models. *Data Sci Financ Econ* 3: 354–379. <https://doi.org/10.3934/dsfe.2023021>
- Erfanian S, Zhou Y, Razzaq A, et al. (2022) Predicting Bitcoin (BTC) Price in the Context of Economic Theories: A Machine Learning Approach. *Entropy* 24: 1487. <https://doi.org/10.3390/e24101487>
- Facure M (2023) *Causal Inference in Python*. O'Reilly Media, Inc.
- Feuerriegel S, Frauen D, Melnychuk V, et al. (2024) Causal machine learning for predicting treatment outcomes. *Nat Med* 30: 958–968. <https://doi.org/10.1038/s41591-024-02902-1>
- Gebreyesus Y, Dalton D, Nixon S, et al. (2023) Machine Learning for Data Center Optimizations: Feature Selection Using Shapley Additive exPlanation (SHAP). *Future Internet* 15: 88. <https://doi.org/10.3390/fi15030088>
- Gu Q, Song S, Zhang X, et al. (2023) Personal Credit Risk Assessment Based on Improved BS-Stacking. *Oper Res Manage Sci* 32: 137–144. <https://doi.org/10.12005/orms.2023.0262>
- Guo K, Fan H (2024) Research on AdaFocal-XGBoost Integrated Credit Scoring Model Based on Unbalanced Data. *Stat Appl* 13: 2204–2214. <https://doi.org/10.12677/sa.2024.136214>
- Janiesch C, Zschech P, Heinrich K (2021) Machine learning and deep learning. *Electron Mark* 31: 685–695. <https://doi.org/10.1007/s12525-021-00475-2>

- Jiang T (2022) Mediating Effects and Moderating Effects in Causal Inference. *China Ind Econ* 5: 100–120. <https://doi.org/10.19581/j.cnki.ciejournal.2022.05.005>
- Künzel SR, Sekhon JS, Bickel PJ, et al. (2019) Metalearners for estimating heterogeneous treatment effects using machine learning. *P Natl Acad Sci* 116: 4156–4165. <https://doi.org/10.1073/pnas.1804597116>
- Lalwani P, Mishra MK, Chadha JS, et al. (2022) Customer churn prediction system: a machine learning approach. *Computing* 104: 271–294. <https://doi.org/10.1007/s00607-021-00908-y>
- Li J, Xiong R, Lan Y, et al. (2023) Overview of the Frontier Progress of Causal Machine Learning. *J Comput Res Dev* 60: 59–84. <https://doi.org/10.7544/issn1000-1239.202110780>
- Li Y, Yan K (2023) Prediction of Barrier Option Price Based on Antithetic Monte Carlo and Machine Learning Methods. *Cloud Comput Data Sci* 4: 77–86. <https://doi.org/10.37256/ccds.4120232110>
- Molak A (2023) *Causal Inference and Discovery in Python*. Packt Publishing Ltd.
- Peng K, Peng Y, Li W (2023) Research on customer churn prediction and model interpretability analysis. *PLOS ONE* 18: e0289724. <https://doi.org/10.1371/journal.pone.0289724>
- Pudjihartono N, Fadason T, Kempa-Liehr AW, et al. (2022) A Review of Feature Selection Methods for Machine Learning-Based Disease Risk Prediction. *Front Bioinform* 2: 927312. <https://doi.org/10.3389/fbinf.2022.927312>
- Siddiqui N, Haque MA, Khan SMS, et al. (2024) Different ML-based strategies for customer churn prediction in banking sector. *J Data Inf Manage* 6: 217–234. <https://doi.org/10.1007/s42488-024-00126-z>
- Wu M, Mao Z, Wang D (2024) Shapley Value and its Application. *Math Model Appl* 13: 110–119. <https://doi.org/10.19943/j.2095-3070.jmmia.2024.01.13>
- Yan K, Li Y (2024) Machine learning-based analysis of volatility quantitative investment strategies for American financial stocks. *Quant Financ Econ* 8: 364–386. <https://doi.org/10.3934/QFE.2024014>
- Yan K, Wang Y, Li Y (2023) Enhanced Bollinger Band Stock Quantitative Trading Strategy Based on Random Forest. *Artif Intell Evol* 4: 22–33. <https://doi.org/10.37256/aie.4120231991>
- Zhang N, Zheng Y, Duan C (2024) Bank Customer Churn Prediction based on Random Forest Algorithm. *In Proceedings of the 5th International Conference on Computer Information and Big Data Applications* : 1031–1035. <https://doi.org/10.1145/3671151.3671331>



AIMS Press

© 2025 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>)