International Journal of Advances in Intelligent Informatics
Vol. 12, No. 1, February 2026, pp. 267-282

Predict customer churn in the banking sector: a machine learning approach with imbalanced data handling techniques

Jong-Hwa Lee [a,1], Van-Ho Nguyen [b,2], Hoanh-Su Le [b,3,*]

[a] Dong-Eui University, Busan City, South Korea
[b] University of Economics and Law and Vietnam National University, Ho Chi Minh City, Vietnam
[1] jhlee6050@deu.ac.kr; [2] honv@uel.edu.vn; [3] sulh@uel.edu.vn
* corresponding author

ARTICLE INFO

Article history
Received October 21, 2025
Revised December 3, 2025
Accepted December 12, 2025
Available online February 28, 2026

Keywords
Churn Prediction
Customer churn
Ensemble Learning
Imbalance data handling
Machine Learning

ABSTRACT

Customer value analysis is a critical component in formulating effective marketing and customer relationship management (CRM) strategies, especially in sectors where client movement and strong competition are prevalent. A key element of this process lies in enhancing customer retention rates, as retaining existing clients is typically more cost-effective than acquiring new ones and directly contributes to improving overall profitability. In today's banking environment, where customers can choose from a broad range of financial services, customer churn has become a critical challenge. Predicting and understanding attrition enables financial institutions to implement proactive and targeted interventions to protect market share and strengthen customer loyalty. This study analyzes a real-world dataset comprising 10,127 customer records from a commercial bank, where only 1,627 entries correspond to churned customers, thereby presenting a notable class imbalance problem. To address this, several data balancing techniques were applied, including class-weight adjustment, SMOTE, SMOTE-Tomek Links, and SMOTE-ENN. Multiple machine learning models - Support Vector Machine, Random Forest, Decision Tree, Logistic Regression, AdaBoost - were evaluated to identify the most effective approach for churn prediction. The Random Forest model achieved an 86% F1-score after applying SMOTE-Tomek Links, demonstrating strong predictive capability. The key contribution of this study lies in integrating advanced resampling techniques with ensemble learning and customer behavioral insights to improve churn prediction performance and support data-driven retention strategies in the banking sector.

1. Introduction

Customer churn continues to represent a major concern for financial institutions, where the cost of acquiring new clients often exceeds the investment required to retain existing ones [1], [2], [3]. In the ongoing digital and heightened customer autonomy, the precision and ethical accountability in customer churn prediction have become a critical strategic priority. Nevertheless, traditional churn models often prioritize predictive accuracy while neglecting essential dimensions, including model interpretability, fairness, and theoretical grounding in consumer behavior. This limitation reduces their managerial utility. Theoretically, churn is a volitional act framed by the Theory of Planned Behavior [4] and Switching Costs Theory [5], in which decisions are shaped by attitudes, social norms, and the relational or financial costs of leaving. Recent empirical work suggests that while relational costs reinforce loyalty, adverse social norms can accelerate switching behavior [6], [7], [8], [9]. Strategically, each customer constitutes a potential stream of long-term revenue, underscoring the significance of Customer Lifetime Value as a foundational construct for designing effective churn reduction strategies [10].

In response, many financial institutions are integrating Artificial Intelligence (AI) into CRM systems to enhance operational efficiency and customer engagement [11], [12]. According to [13] and [14], predictive modeling - leveraging Machine Learning (ML) algorithms and statistical methods - plays a pivotal role in anticipating churn behavior. These models enable banks to estimate the likelihood of churn at the individual customer level, thus allowing for timely and targeted retention efforts [15]. An increasing amount of research uses deep learning, hybrid architectures, and ensemble approaches to significantly improve prediction performance. [16] implement a voting ensemble model comprising K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Trees (DT), Random Forest (RF), and XGBoost, augmented by SMOTE-based class balancing, resulting in an improvement of F1-score from 87% to 90%. Similarly, [17] employ a genetic algorithm to fine-tune an XGBoost model with SMOTE+ENN sampling strategies. Applied to banking-sector datasets, their optimized GA-XGBoost method achieves an F1-score of about 90% and an AUC of 99%.

Despite this progress, their real-world application has been critiqued for lacking transparency in the decision-making process [18], [19]. High-performance models such as ensemble methods and deep neural networks often function as "black boxes", complicating efforts to understand their predictions. To address this, recent advances in Model-Agnostic Explainability over the past decade [20], [21], [22] have introduced interpretability techniques, notably SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME), which help translate complex model logic into actionable insights. These methods enable both holistic model comprehension and granular insight into specific predictions, such as counterfactual explanations [23] and telecom churn interpretation templates [24], thereby highlighting their usefulness. Concurrently, the imperative of algorithmic fairness emerges as a central concern among researchers [25], as models trained on historical data can reinforce social biases [26], especially in banking, where features may indirectly relate to sensitive attributes. While recent studies employ Group and Individual fairness criteria [27] and demonstrate that fairness-aware preprocessing can enhance both equity and predictive outcomes [28], a critical gap persists because current research seldom unifies high-performance prediction on imbalanced data with robust model interpretability and comprehensive fairness auditing.

To provide a comprehensive overview of the current landscape, Table 1 summarizes recent studies (2022–2024) on churn prediction. While these studies demonstrate the high predictive power of ensemble methods such as Random Forest and XGBoost, they predominantly focus on accuracy metrics within the telecom and banking sectors, often overlooking the integration of fairness auditing and granular model explainability. This study adopts a comprehensive strategy by integrating robust resampling approaches with multiple machine learning models to improve churn prediction performance under class imbalance conditions. The study focuses on evaluating the effectiveness of different imbalance-handling techniques and machine learning algorithms in predicting customer churn within the banking sector.

Table 1. Summary of recent churn prediction studies.

Reference	Year	Algorithm	Dataset	Results
Muneer et al [15]	2022	RF, AdaBoost, SVM	Banking sector	Random Forest outperformed single classifiers, achieving 88.7% accuracy
Lukita et al [29]	2023	RF, LR, AdaBoost, XGBoost	Banking sector	XGBoost achieved the highest performance, reaching 87% accuracy
Wagh et al [30]	2023	DT and RF	Telecom sector	Focusing on feature importance, Random Forest achieved 99% accuracy
Maduna et al [31]	2024	LR, DT, RF SVM, KNN	Banking sector	Random Forest attained 87% accuracy
Krishna et al [32]	2024	SVM, GBM, XGBoost, LightGBM LR, RF	Telecom sector	Random Forest achieved its highest performance with 95.6% accuracy

Consequently, the main goal of this paper is to develop a machine learning framework for customer churn prediction in the banking sector. To address the primary research question, "To what extent can machine learning models accurately predict customer churn in banking under class imbalance conditions?" this study outlines the following specific contributions:

- Developing some state-of-the-art ML models and assessing their predictive power using an imbalanced banking dataset.
- Highlighting the potential role of explainable AI techniques such as SHAP in improving the interpretability of churn prediction models and identifying directions for future research on model transparency.
- Discussing the importance of fairness considerations in machine learning applications within the banking domain and identifying fairness-aware modeling as an important direction for future research.

Providing managerial insights into how data-driven churn prediction models can support customer retention strategies in the banking sector.

2. Method

2.1. Proposed model

Fig. 1 illustrates the research stages designed to address the issue of imbalanced data, comprising four main phases: pre-analysis, model tuning, model training, and model evaluation and visualization. The pre-analysis phase involves data exploration and preparation, including label encoding, feature extraction, and data scaling. In the model tuning phase, various techniques are applied to optimize model parameters and improve performance on imbalanced datasets. Once the data is prepared, the model training phase experiments with several algorithms, including SVM, RF, DT, LR, AdaBoost, and Easy Ensemble. Finally, the model evaluation and visualization phase assesses the models' performance and presents the results through visualizations to support business interpretation and decision-making.

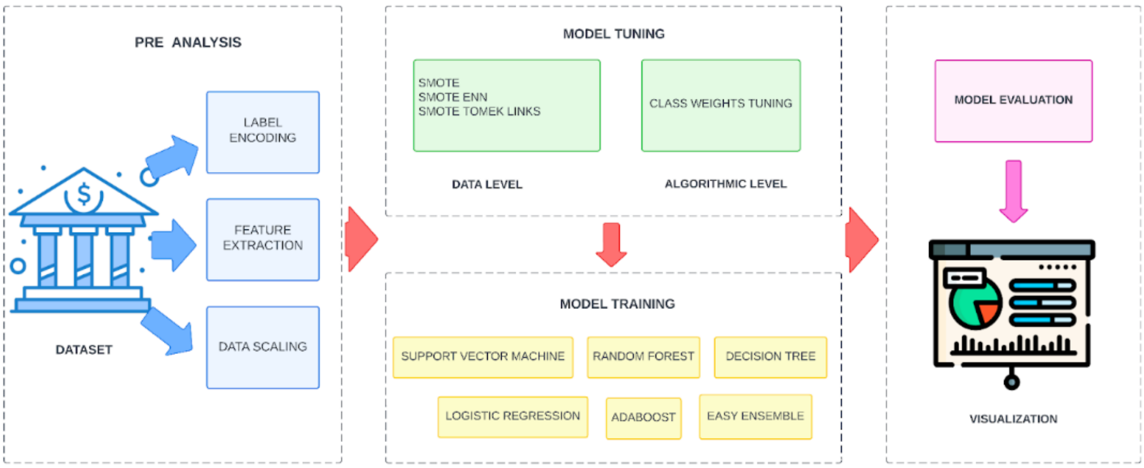

Fig. 1. Proposed model for an imbalanced dataset.

2.2. Dataset collection

The experimental data for this study are drawn from the "Credit Card Customers" dataset, which is publicly available via the Kaggle repository (Goyal, 2020). The dataset consists of 10,127 records of bank customers, including their demographic and transaction information, as well as whether they have churned. With only 1627 churned customers, representing just 16% of the dataset, it becomes more challenging to classify customers due to the imbalanced data. There are 20 variables and no null values present in the dataset, as shown in Fig. 2.

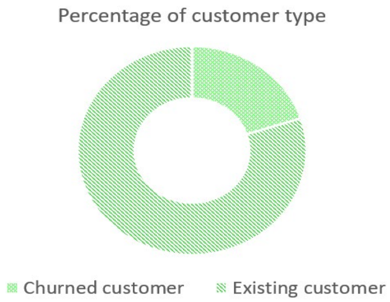

Fig. 2. Proportion of customers in the dataset.

The dataset combines transactional, financial, behavioral, and demographic characteristics, which facilitates predictive modeling, customer segmentation, fairness learning, and churn analysis. Table 2 describes the original variables included in the experimental dataset. A comprehensive dictionary of the original input variables is provided in Appendix A.

To enhance the model's predictive power and capture latent behavioral patterns not explicitly present in the raw data, we applied domain-specific theoretical constructs to engineer 34 derived features. While raw data provides static snapshots of customer status, the engineered features are designed to reveal dynamic behavioral trends and interaction effects. A detailed taxonomy of these 34 engineered features, including their mathematical definitions and the logic used to calculate them, is presented in Appendix B.

Table 2. Variables definition.

Variable Name	Category	Academic Description
Total_Trans_Ct	Transactional	Represents the total number of individual transactions carried out by the customer within a 12-month timeframe.
Total_Ct_Chng_Q4_Q1	Transactional	Reflects the relative change in transaction frequency from the first to the fourth fiscal quarter.
Avg_Utilization_Ratio	Financial	Represents the mean ratio of credit card balance to credit limit, indicating average card utilization over a specified time horizon.
Attrition_Flag	Behavioral	Indicates customer churned from the business

2.3. Data Encoding

Customer demographic information is qualitative, so it needs to be converted to numerical data using Label Encoding and One Hot Encoding. The process of label encoding involves transforming each qualitative value within a column into a corresponding numerical value. This method can be proposed simply, but the drawback is that numeric values can lead to misinterpretation, as the model may be unsure whether one value is greater or less than the others [28]. One Hot Encoding is the process of creating new columns for each unique categorical value in a column and assigning a binary value to each based on its presence in the original data. The One Hot Encoding method can represent categorical values equally, but it significantly increases the number of dimensions, thereby increasing computational complexity [33].

2.4. Feature selection

Recursive Feature Elimination with Cross-Validation (RFECV) is a wrapper-based feature selection method that systematically eliminates a specified number of features in each iteration to identify the optimal subset for a given machine learning model. The selection process is guided by cross-validation performance scores, ensuring the robustness of the chosen features. Beyond its primary function of feature selection, RFECV also helps mitigate multicollinearity and remove redundant dependencies among variables [34]. This iterative refinement aims to retain only the features that significantly enhance model performance, thereby improving overall accuracy and generalizability [35]. In this study, RFECV was employed to select features from the training dataset, ensuring that only the most relevant variables were retained for model development.

2.5. Imbalanced data resampling

Imbalanced datasets are a well-recognized challenge in machine learning, characterized by a disproportionate distribution of instances across classes, typically with a dominant majority class and an underrepresented minority class. Conventional machine learning algorithms often aim to maximize overall accuracy, which can result in biased models that favor the majority class while neglecting the minority class, leading to poor generalization for the latter [36]. To address this issue, the present study employs two main strategies: oversampling and class-weight adjustment during model training. Oversampling involves increasing the representation of the minority class by synthetically generating new instances or replicating existing ones. Specifically, we employed SMOTE, SMOTE-ENN, and SMOTE-Tomek Links to improve the minority-class distribution. However, oversampling techniques can introduce challenges, such as increased variance in classifier predictions and the potential distortion of posterior probabilities [37]. As a complementary approach, class weight tuning was also used, which involves adjusting each class's contribution to the loss function, thereby mitigating the dominance of the majority class and improving model fairness [38]. Oversampling configuration and its application timeline are critical for reproducibility and proper evaluation. In the revised manuscript, we would like to clarify that oversampling was applied only to the training set, after the train-test split, to avoid data leakage and ensure an unbiased evaluation of model performance.

2.6. Data scaling

Scalability is an important issue in real-world ML applications, especially when the data includes features with very different value ranges. When this happens, some features may have much smaller values than others, which can cause the model to ignore them during training [39]. To solve this, feature scaling is used to bring all features to a similar range, helping the model treat them equally. Scaling also helps reduce the distance between data points, making the learning process more stable and efficient. Common scaling techniques include Min-Max Scaler, Max-Abs Scaler, and Quantile Transformer Scaler, each suited for different types of data and tasks [40]. In this study, we utilized Min-Max normalization, which transforms the values of each feature to a fixed range - usually between 0 and 1 - so that all features have equal importance during model training

2.7. Machine Learning algorithm

LR is a commonly employed statistical technique designed for analyzing relationships between multiple independent variables and a binary dependent variable [41], [42]. It is used in machine learning as a simple probabilistic classifier that estimates class probabilities using the Sigmoid function. LR tends to perform effectively when the data is linearly separable [43]. SVM is one of the most common machine learning algorithms used in both classification and regression. By defining an ideal hyperplane in a multidimensional space and optimizing the margin while minimizing errors, it divides classes [40]. With the flexibility that kernel functions offer [43], SVMs excel in high-dimensional settings and have a wide range of applications across different fields [33].

DT represents one of the earliest and most influential machine learning algorithms, with foundational contributions by researchers such as [44] and [45] . It classifies data from root nodes to leaf nodes by constructing a hierarchy of decision rules. DTs are useful for both classification and regression because they are straightforward, quick, and simple to understand [40]. According to [46], their transparency aids in identifying significant features and class relationships. RF is an ensemble learning method that constructs multiple DTs, each built on a randomly selected subset of the training data and features [47]. The final classification decision for a test instance is determined through majority voting across all individual trees in the forest [48]. RF offers high accuracy, robustness to noise, and strong generalization, effectively handling non-linear and correlated features [49]. It has shown strong performance in customer churn prediction, outperforming several other methods [50]. Adaptive Boosting (AdaBoost) is an ensemble technique that combines multiple weak learners through weighted voting to form a strong classifier. It adjusts instance weights to focus on difficult cases, improving performance iteratively [46], [51]. AdaBoost is widely used for its high accuracy and adaptability, including in customer churn prediction.

The EasyEnsemble classifier, introduced by [52], was developed to address the limitations of traditional undersampling methods while maintaining their computational efficiency. It trains multiple classifiers on balanced subsets of the majority class and combines them using AdaBoost. This ensemble approach effectively addresses class imbalance and has demonstrated strong performance on imbalanced classification tasks [53].

2.8. Model evaluation metrics

The performance of the classification models was evaluated using accuracy rate (Accuracy), precision (P), recall (R), and F1 scores as shown in Table 3. The dataset was split into 70/30, with 70% for training and 30% for testing.

Table 3. Confusion matrix.

Actual class	Predicted class	
	Positive (Attrited customers)	Negative (Existing customers)
Positive (Attrited customers)	True Positive (TP)	False Negative (FN)
Negative (Existing customers)	False Positive (FP)	True Negative (TN)

3. Results and Discussion

Table 4 presents the original dataset before any imbalanced data processing. It reflects the raw distribution of customer classes without applying balancing techniques. In this dataset, class 0 represents existing (retained) customers, while class 1 denotes customers who have churned.

Table 4. Original dataset without imbalanced data processing results.

Original dataset	Accuracy	Precision		Recall		F1-score	
		Class 0	Class 1	Class 0	Class 1	Class 0	Class 1
Logistic Regression (LR)	90%	92%	71%	96%	52%	94%	60%
Decision Tree (DT)	92%	95%	73%	95%	73%	95%	73%
Random Forest (RF)	95%	96%	91%	99%	76%	97%	82%
Support Vector Machine (SVM)	92%	94%	81%	97%	64%	96%	71%
Adaboost	93%	95%	83%	97%	70%	96%	76%
EasyEnsemble	90%	98%	62%	90%	89%	94%	73%

According to Fig. 3, the RF and EasyEnsemble ensemble algorithms performed better without any data imbalance treatment. The Random Forest models achieved a remarkable accuracy of 95% and an F1-score of 82% for class 1, both of which were the highest among all models. On the other hand, the LR model achieved the lowest accuracy (90%) and F1-score (60%) for class 1, indicating inferior performance relative to the ensemble algorithms, as shown in Table 5.

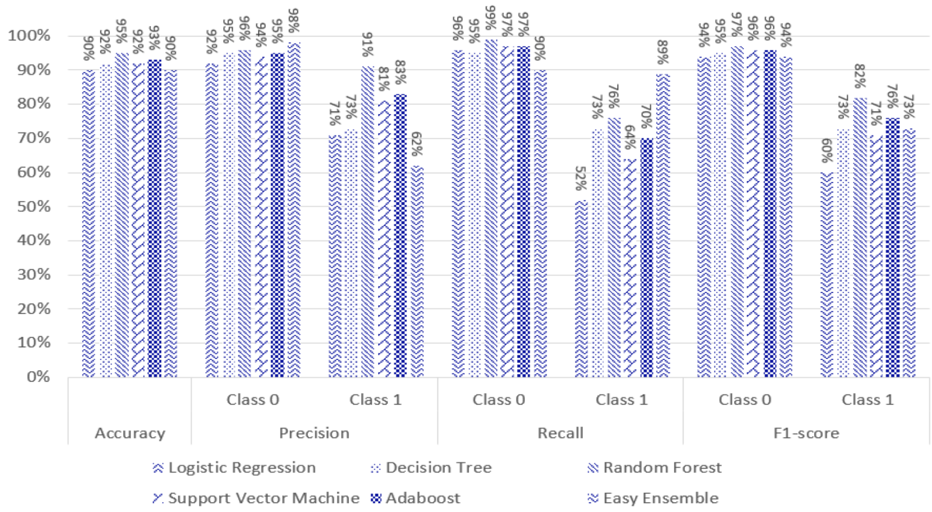

Fig. 3. Original dataset results visualization.

Fig. 4 and Table 4 show that, except for the SVM model, all models experienced a 26% decrease in F1-score, and all models demonstrated significant improvements in their F1-scores for attrited customers. The application of class weights particularly benefited the LR and DT models, resulting in a 7% increase in their F1 scores. Despite these improvements, the RF model still achieved the highest F1 score for class 1.

Table 5. Original dataset with class weights tuning results.

Original dataset with class weights	Accuracy	Precision		Recall		F1-score	
		Class 0	Class 1	Class 0	Class 1	Class 0	Class 1
Logistic Regression (LR)	90%	95%	63%	93%	72%	94%	67%
Decision Tree (DT)	94%	96%	**80%**	**96%**	80%	96%	80%
Random Forest (RF)	**96%**	**99%**	79%	**96%**	**93%**	**97%**	**85%**
Support Vector Machine (SVM)	84%	93%	41%	89%	52%	91%	45%
Adaboost	93%	95%	79%	**96%**	77%	96%	78%
EasyEnsemble	93%	95%	79%	**96%**	77%	96%	78%

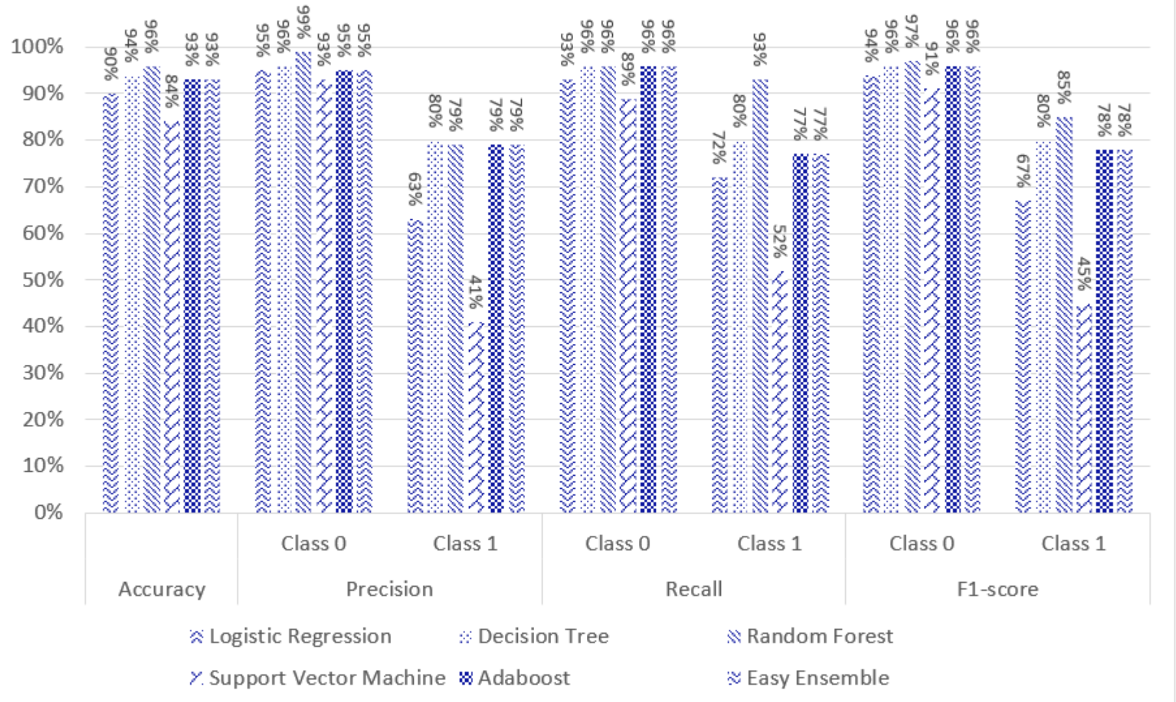

Fig. 4. Original dataset with class weights tuning results visualization.

Table 6 shows the evaluation metrics of the models trained on data balanced using SMOTE, a synthetic oversampling technique. Overall, Random Forest achieved the best F1-score of 76% for churned customers, along with high recall and precision, indicating reliable identification of the minority class. Decision Tree and SVM also performed moderately well, achieving F1 scores of 70% and 68%, respectively. Meanwhile, Logistic Regression's performance remained modest, with an F1-score of 54% for Class 1. These outcomes, as visualized in Fig. 5, suggest that while SMOTE improves recall for most models, it may not sufficiently boost precision, especially for linear models. The results also confirm that ensemble classifiers remain more effective under synthetic balancing conditions

Table 6. SMOTE dataset results.

SMOTE	Accuracy	Precision		Recall		F1-score	
		Class 0	Class 1	Class 0	Class 1	Class 0	Class 1
Logistic Regression (LR)	78%	97%	40%	77%	84%	86%	54%
Decision Tree (DT)	89%	97%	60%	90%	84%	93%	70%
Random Forest (RF)	**92%**	**98%**	**67%**	**92%**	88%	**95%**	**76%**
Support Vector Machine (SVM)	87%	97%	55%	88%	87%	92%	68%
Adaboost	87%	**98%**	54%	87%	**89%**	92%	68%

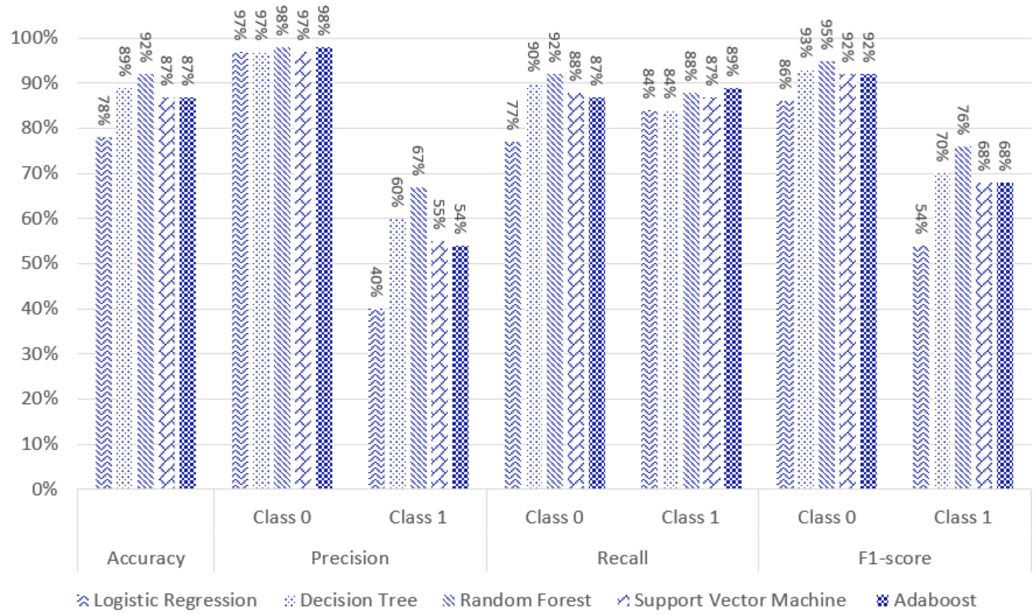

Fig. 5. SMOTE dataset results visualization.

The performance of the 5 ML algorithms on the SMOTE-ENN dataset is summarized in Table 7 and visualized in Fig. 6. Among the models, RF demonstrated the strongest performance, achieving an accuracy of 93% and an F1-score of 80% for churned customers (Class 1). This balance between precision (75%) and recall (86%) reflects its robustness in detecting instances of the minority class without overfitting.

Table 7. SMOTE ENN dataset results.

SMOTE ENN	Accuracy	Precision		Recall		F1-score	
		Class 0	Class 1	Class 0	Class 1	Class 0	Class 1
Logistic Regression (LR)	81%	95%	45%	81%	80%	88%	58%
Decision Tree (DT)	91%	97%	69%	92%	86%	95%	76%
Random Forest (RF)	93%	97%	75%	94%	86%	96%	80%
Support Vector Machine (SVM)	71%	93%	33%	71%	74%	81%	46%
Adaboost	92%	98%	70%	93%	90%	95%	79%

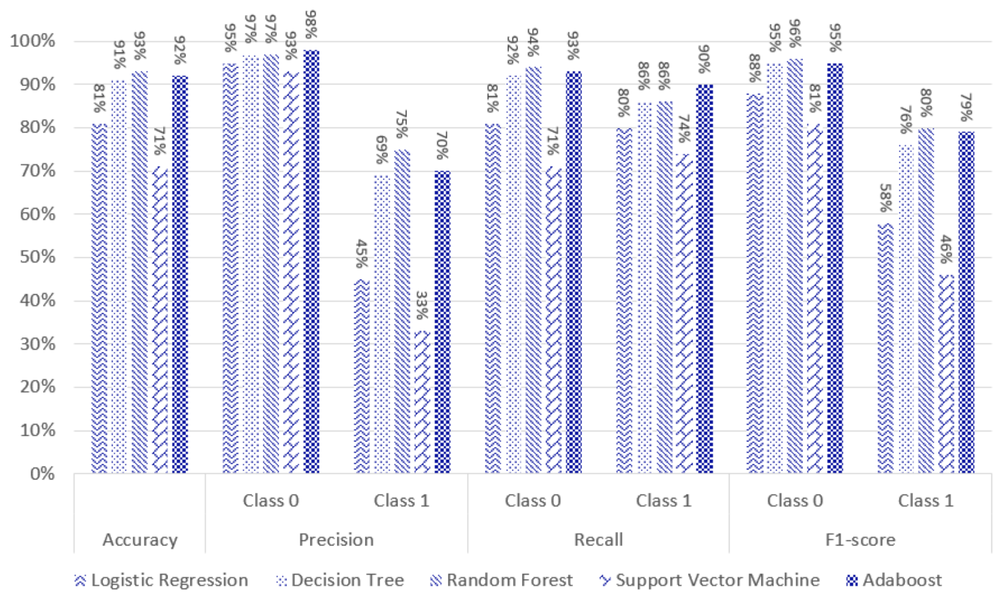

Fig. 6. SMOTE ENN dataset results visualization.

Adaboost and DT also performed competitively, attaining F1-scores of 79% and 76%, respectively, for Class 1. In contrast, LR yielded a more modest F1-score of 58%, hindered by its relatively low precision. SVM exhibited the weakest performance across all metrics for Class 1, with an F1-score of only 46%, indicating limited suitability for this resampling context. As shown in Fig. 6, these results are consistent across all evaluation metrics, with tree-based models and boosting techniques outperforming simpler linear models such as LR and models that rely on decision boundaries, like SVM. The figure clearly highlights the superior balance achieved by RF and Adaboost, while also illustrating the significant performance drop of SVM, especially in precision and recall for the churn class.

Table 8 and Fig. 7 show the results of applying the same classifiers to the SMOTE-Tomek Links dataset, which combines synthetic oversampling with boundary cleaning to improve class separability. In this setting, RF again emerged as the top-performing model, reaching 95% accuracy and achieving the highest F1-score for Class 1 at 86%. This performance reflects a strong synergy between the model's tree-structured learning approach and the cleaner data distribution produced by the Tomek Links algorithm. Adaboost and DT followed closely with F1-scores of 83% and 79%, confirming their reliability in this experimental configuration. LR showed a slight improvement, with an F1-score of 57% for Class 1, while SVM continued to struggle, managing only 45% in the same metric. Fig. 7 offers a clear visual confirmation of these findings, with RF maintaining a consistent lead across all metrics, particularly in churn detection. The bar chart also emphasizes the performance gaps between models based on decision trees and their simpler linear counterparts or those depending heavily on boundary optimization, underlining the challenges faced by SVM in adapting to synthetic data distributions even after boundary refinement..

Table 8. SMOTE Tomek links dataset results.

SMOTE Tomek links	Accuracy	Precision		Recall		F1-score	
		Class 0	Class 1	Class 0	Class 1	Class 0	Class 1
Logistic Regression (LR)	82%	94%	46%	83%	75%	88%	57%
Decision Tree (DT)	93%	**97%**	75%	94%	84%	96%	79%
Random Forest (RF)	**95%**	**97%**	**87%**	**97%**	85%	**97%**	**86%**
Support Vector Machine (SVM)	72%	93%	33%	72%	72%	81%	45%
Adaboost	94%	**97%**	79%	96%	**87%**	**97%**	83%

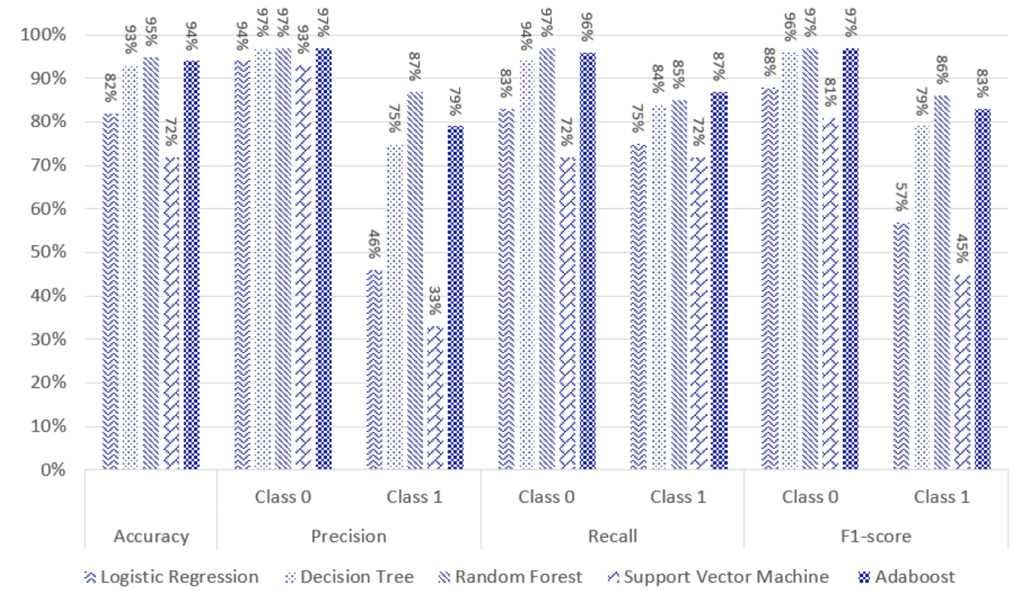

Fig. 7. SMOTE Tomek Links dataset results visualization.

It is worth noting that oversampling methods may not always yield better results in handling imbalanced data. In some instances, such as with the SMOTE and SMOTE ENN datasets, these techniques resulted in lower F1 scores, particularly for Support Vector Machine models. Hence, oversampling may not be the ideal approach to address imbalanced data. However, the Random Forest model trained by the SMOTE Tomek links dataset achieved an impressive F1 score of up to 86% for class 1, demonstrating the potential effectiveness of oversampling when combined with the correct method.

4. Conclusion

This study builds upon the Theory of Planned Behavior [4] and Switching Costs Theory [5] together with the strategic importance of Customer Lifetime Value (CLV) [10], to frame customer churn as a behavior shaped by psychological, relational, and economic factors. The empirical findings highlight the critical impact of class imbalance on churn prediction within the banking sector. Conventional algorithms such as LR showed notable performance deterioration under severe imbalance, whereas ensemble-based models - particularly RF and Easy Ensemble - demonstrated consistently superior predictive performance and robustness. Among the imbalance-handling techniques evaluated, class weight tuning proved to be a highly effective and computationally efficient solution. It significantly improved minority-class F1-scores without generating synthetic samples. Notably, the results challenge the common assumption that oversampling inherently enhances performance. In several cases - especially when applying SMOTE or SMOTE-ENN to SVM - oversampling led to reduced classification effectiveness. These findings underscore the context-dependent nature of oversampling and the importance of aligning imbalance strategies with model-specific behavior and dataset characteristics. Despite these contributions, several limitations must be acknowledged. First, the experimental results are derived from a single dataset, suggesting that the findings may reflect specific regional or demographic biases and may not generalize immediately to financial institutions with different customer profiles. Second, the model's performance was assessed using retrospective historical data; this offline evaluation does not account for real-time data-streaming challenges or the "concept drift" often observed in live production environments. Finally, due to the computational scope of benchmarking 20 distinct model configurations, rigorous statistical validation (e.g., paired t-tests) was not performed in this study. This study contributes to the banking domain by providing a comprehensive comparison of imbalance mitigation strategies and by proposing a practical roadmap for real-world implementation. Future research should further investigate how imbalance mitigation techniques interact with advanced machine learning architectures, including deep learning and hybrid ensemble approaches. Additionally, expanding the analysis across multiple financial institutions, incorporating temporal behavioral data, and exploring real-time or streaming-based churn prediction systems would enhance the generalizability and operational applicability of future models. Future studies may also incorporate explainable AI techniques and fairness-aware evaluation to further improve model transparency and ethical accountability in banking applications.

Acknowledgment

This research is funded by University of Economics and Law, Vietnam National University Ho Chi Minh City, Vietnam.

Declarations

Author contribution. All authors contributed equally to this paper.
Funding statement. This research is funded by University of Economics and Law, Vietnam National University Ho Chi Minh City, Vietnam.
Conflict of interest. The authors declare no conflict of interest.
Additional information. No additional information is available for this paper.

References

[1] E. Kaya, X. Dong, Y. Suhara, S. Balcisoy, B. Bozkaya, and A. S. Pentland, "Behavioral attributes and financial churn prediction," *EPJ Data Sci.*, vol. 7, no. 1, p. 41, Dec. 2018, doi: 10.1140/epjds/s13688-018-0165-5.

[2] V. Gkonis and I. Tsakalos, "Deep Dive Into Churn Prediction in the Banking Sector: The Challenge of Hyperparameter Selection and Imbalanced Learning," *J. Forecast.*, vol. 44, no. 2, pp. 281–296, Mar. 2025, doi: 10.1002/for.3194.

[3] P. Jiang, Z. Liu, M. Z. Abedin, J. Wang, W. Yang, and Q. Dong, "Profit-driven weighted classifier with interpretable ability for customer churn prediction," *Omega*, vol. 125, no. June, p. 103034, Jun. 2024, doi: 10.1016/j.omega.2024.103034.

[4] I. Ajzen, "The theory of planned behavior," *Organ. Behav. Hum. Decis. Process.*, vol. 50, no. 2, pp. 179–211, Dec. 1991, doi: 10.1016/0749-5978(91)90020-T.

[5] T. A. Burnham, J. K. Frels, and V. Mahajan, "Consumer Switching Costs: A Typology, Antecedents, and Consequences," *J. Acad. Mark. Sci.*, vol. 31, no. 2, pp. 109–126, Apr. 2003, doi: 10.1177/0092070302250897.

[6] T. A. Ganaie and M. A. Bhat, "Impact of Perceived Switching Costs on Customer Loyalty in Banks: An Empirical Investigation," *Int. J. Res. Trends Innov.*, vol. 8, no. 4, p. 9, 2023, Accessed: Mar. 09, 2026. [Online]. Available at: https://www.ijrti.org/papers/IJRTI2304170.

[7] G. Batmunkh, "An investigation of the switching behavior and why customers switch banks in the retailing banking sector," *Front. Commun.*, vol. 10, p. 1548050, May 2025, doi: 10.3389/fcomm.2025.1548050.

[8] M. Georgiou, S. Daskou, A. Anastasiou, and M. Siakalli, "The effects of the theory of planned behaviour on the switching propensity of retail banking customers at different critical switching incidents," *Int. J. Bank Mark.*, vol. 41, no. 7, pp. 1872–1898, Dec. 2023, doi: 10.1108/IJBM-12-2022-0532.

[9] Ø. Bortne, J. Bjornestad, M. N. Arnestad, T. Tjora, and K. K. Brønnick, "Self-efficacy or perceived behavioral control: which influences bank-switching intention?," *J. Mark. Anal.*, pp. 1–14, May 2025, doi: 10.1057/s41270-025-00408-4.

[10] J. S. Thomas, R. C. Blattberg, and E. J. Fox, "Recapturing Lost Customers," *J. Mark. Res.*, vol. 41, no. 1, pp. 31–45, Feb. 2004, doi: 10.1509/jmkr.41.1.31.25086.

[11] A. Bilal Zoric, "Predicting Customer Churn in Banking Industry using Neural Networks," *Interdiscip. Descr. Complex Syst.*, vol. 14, no. 2, pp. 116–124, 2016, doi: 10.7906/indecs.14.2.1.

[12] C. Ledro, A. Nosella, and I. Dalla Pozza, "Integration of AI in CRM: Challenges and guidelines," *J. Open Innov. Technol. Mark. Complex.*, vol. 9, no. 4, p. 100151, Dec. 2023, doi: 10.1016/j.joitmc.2023.100151.

[13] P. Supsermpol, V. N. Huynh, S. Thajchayapong, and N. Chiadamrong, "Predicting financial performance for listed companies in Thailand during the transition period: A class-based approach using logistic regression and random forest algorithm," *J. Open Innov. Technol. Mark. Complex.*, vol. 9, no. 3, p. 100130, Sep. 2023, doi: 10.1016/j.joitmc.2023.100130.

[14] P. P. Singh, F. I. Anik, R. Senapati, A. Sinha, N. Sakib, and E. Hossain, "Investigating customer churn in banking: a machine learning approach and visualization app for data science and management," *Data Sci. Manag.*, vol. 7, no. 1, pp. 7–16, Mar. 2024, doi: 10.1016/j.dsm.2023.09.002.

[15] A. Muneer, R. Faizan Ali, A. Alghamdi, S. Mohd Taib, A. Almaghthawi, and E. A. A. Ghaleb, "Predicting customers churning in banking industry: A machine learning approach," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 26, no. 1, p. 539, Apr. 2022, doi: 10.11591/ijeecs.v26.i1.pp539-549.

[16] R. Bhuria *et al.*, "Ensemble-based customer churn prediction in banking: a voting classifier approach for improved client retention using demographic and behavioral data," *Discov. Sustain.*, vol. 6, no. 1, p. 28, Jan. 2025, doi: 10.1007/s43621-025-00807-8.

[17] K. Peng, Y. Peng, and W. Li, "Research on customer churn prediction and model interpretability analysis," *PLoS One*, vol. 18, no. 12, p. e0289724, Dec. 2023, doi: 10.1371/journal.pone.0289724.

[18] J. Tao *et al.*, "Explainable AI for Cheating Detection and Churn Prediction in Online Games," *IEEE Trans. Games*, vol. 15, no. 2, pp. 242–251, Jun. 2023, doi: 10.1109/TG.2022.3173399.

[19] A. Manzoor, M. Atif Qureshi, E. Kidney, and L. Longo, "A Review on Machine Learning Methods for Customer Churn Prediction and Recommendations for Business Practitioners," *IEEE Access*, vol. 12, pp. 70434–70463, 2024, doi: 10.1109/ACCESS.2024.3402092.

[20] M. T. Ribeiro, S. Singh, and C. Guestrin, "Model-Agnostic Interpretability of Machine Learning," 2016. doi: 10.48550/arXiv.1606.05386.

[21] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why Should I Trust You?,'" in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA: ACM, Aug. 2016, pp. 1135–1144. doi: 10.1145/2939672.2939778.

[22] S. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," Neural information processing systems foundation, Nov. 2017. doi: 10.48550/arXiv.1705.07874.

[23] S. Tonati, M. Di Vece, R. Pellungrini, and F. Giannotti, "Ensemble Counterfactual Explanations for Churn Analysis," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 15244 LNAI, Springer Science and Business Media Deutschland GmbH, 2025, pp. 332–347. doi: 10.1007/978-3-031-78980-9_21.

[24] S. S. Poudel, S. Pokharel, and M. Timilsina, "Explaining customer churn prediction in telecom industry using tabular machine learning models," *Mach. Learn. with Appl.*, vol. 17, no. September, p. 100567, Sep. 2024, doi: 10.1016/j.mlwa.2024.100567.

[25] C. Ashurst and A. Weller, "Fairness Without Demographic Data: A Survey of Approaches," in *Equity and Access in Algorithms, Mechanisms, and Optimization*, New York, NY, USA: ACM, Oct. 2023, pp. 1–12. doi: 10.1145/3617694.3623234.

[26] C. Denis, R. Elie, M. Hebiri, and F. Hu, "Fairness guarantee in multi-class classification," Mar. 2023. doi: 10.48550/arXiv.2109.13642.

[27] C. Dwork, M. Hardt, T. Pitassi, O. Reingold, and R. Zemel, "Fairness through awareness," in *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, New York, NY, USA: ACM, Jan. 2012, pp. 214–226. doi: 10.1145/2090236.2090255.

[28] E. Jackson and R. Agrawal, "Performance Evaluation of Different Feature Encoding Schemes on Cybersecurity Logs," in *2019 SoutheastCon*, IEEE, Apr. 2019, pp. 1–9. doi: 10.1109/SoutheastCon42311.2019.9020560.

[29] C. Lukita, "Predictive and Analytics using Data Mining and Machine Learning for Customer Churn Prediction," *J. Appl. Data Sci.*, vol. 4, no. 4, pp. 454–465, Dec. 2023, doi: 10.47738/jads.v4i4.131.

[30] S. K. Wagh, A. A. Andhale, K. S. Wagh, J. R. Pansare, S. P. Ambadekar, and S. H. Gawande, "Customer churn prediction in telecom sector using machine learning techniques," *Results Control Optim.*, vol. 14, no. March, p. 100342, Mar. 2024, doi: 10.1016/j.rico.2023.100342.

[31] M. Maduna, A. Telukdarie, I. Munien, U. Onkonkwo, and A. Vermeulen, "Smart Customer Churn Management System Using Machine Learning," *Procedia Comput. Sci.*, vol. 237, pp. 552–558, Jan. 2024, doi: 10.1016/j.procs.2024.05.139.

[32] R. Krishna, D. Jayanthi, D. S. Shylu Sam, K. Kavitha, N. K. Maurya, and T. Benil, "Application of machine learning techniques for churn prediction in the telecom business," *Results Eng.*, vol. 24, no. 4, p. 103165, Dec. 2024, doi: 10.1016/j.rineng.2024.103165.

[33] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, "A comprehensive survey on support vector machine classification: Applications, challenges and trends," *Neurocomputing*, vol. 408, no. September, pp. 189–215, Sep. 2020, doi: 10.1016/j.neucom.2019.10.118.

[34] R. A. Mowri, M. Siddula, and K. Roy, "A Comparative Performance Analysis of Explainable Machine Learning Models With And Without RFECV Feature Selection Technique Towards Ransomware Classification," Dec. 2022. doi: 10.48550/arXiv.2212.04864.

[35] A. Bhat, V. Lavanya, and T. Shahana, "Ensemble Classifiers for Bankruptcy Prediction Using SMOTE and RFECV," *Int. J. Enterp. Netw. Manag.*, vol. 15, no. 1, pp. 109–132, 2024, doi: 10.1504/IJENM.2024.10058997.

[36] P. Vuttipittayamongkol, E. Elyan, and A. Petrovski, "On the class overlap problem in imbalanced data classification," *Knowledge-Based Syst.*, vol. 212, no. Januari, p. 106631, Jan. 2021, doi: 10.1016/j.knosys.2020.106631.

[37] A. Fernandez, S. Garcia, F. Herrera, and N. V. Chawla, "SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary," *J. Artif. Intell. Res.*, vol. 61, pp. 863–905, Apr. 2018, doi: 10.1613/jair.1.11192.

[38] B. Bakirarar and A. H. Elhan, "Class Weighting Technique to Deal with Imbalanced Class Problem in Machine Learning: Methodological Research," *Turkiye Klin. J. Biostat.*, vol. 15, no. 1, pp. 19–29, 2023, doi: 10.5336/biostatic.2022-93961.

[39] H. Henderi, "Comparison of Min-Max normalization and Z-Score Normalization in the K-nearest neighbor (kNN) Algorithm to Test the Accuracy of Types of Breast Cancer," *IJIIS Int. J. Informatics Inf. Syst.*, vol. 4, no. 1, pp. 13–20, Mar. 2021, doi: 10.47738/ijiis.v4i1.73.

[40] S. Uddin, A. Khan, M. E. Hossain, and M. A. Moni, "Comparing different supervised machine learning algorithms for disease prediction," *BMC Med. Inform. Decis. Mak.*, vol. 19, no. 1, p. 281, Dec. 2019, doi: 10.1186/s12911-019-1004-8.

[41] M. E. Shipe, S. A. Deppen, F. Farjah, and E. L. Grogan, "Developing prediction models for clinical use using logistic regression: an overview," *J. Thorac. Dis.*, vol. 11, no. S4, pp. S574–S584, Mar. 2019, doi: 10.21037/jtd.2019.01.25.

[42] T. G. Nick and K. M. Campbell, "Logistic Regression," in *Methods in molecular biology (Clifton, N.J.)*, vol. 404, Humana Press, 2007, pp. 273–301. doi: 10.1007/978-1-59745-530-5_14.

[43] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," *SN Comput. Sci.*, vol. 2, no. 3, p. 160, May 2021, doi: 10.1007/s42979-021-00592-x.

[44] J. R. Quinlan, "Learning efficient classification procedures and their application to chess end games," in *Machine Learning*, Elsevier, 1983, pp. 463–482. doi: 10.1016/B978-0-08-051054-5.50019-4.

[45] B. de Ville, "Decision trees," *WIREs Comput. Stat.*, vol. 5, no. 6, pp. 448–455, Nov. 2013, doi: 10.1002/wics.1278.

[46] Y. Wang and L. Feng, "Improved Adaboost Algorithm for Classification Based on Noise Confidence Degree and Weighted Feature Selection," *IEEE Access*, vol. 8, pp. 153011–153026, 2020, doi: 10.1109/ACCESS.2020.3017164.

[47] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.

[48] A. Parmar, R. Katariya, and V. Patel, "A Review on Random Forest: An Ensemble Classifier," in *Lecture Notes on Data Engineering and Communications Technologies*, vol. 26, Springer, Cham, 2019, pp. 758–763. doi: 10.1007/978-3-030-03146-6_86.

[49] Y. Liu, Y. Wang, and J. Zhang, "New Machine Learning Algorithm: Random Forest," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7473 LNCS, Springer, Berlin, Heidelberg, 2012, pp. 246–252. doi: 10.1007/978-3-642-34062-8_32.

[50] I. Ullah, B. Raza, A. K. Malik, M. Imran, S. U. Islam, and S. W. Kim, "A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector," *IEEE Access*, vol. 7, pp. 60134–60149, 2019, doi: 10.1109/ACCESS.2019.2914999.

[51] D. Do, P. Huynh, P. Vo, and T. Vu, "Customer churn prediction in an internet service provider," in *2017 IEEE International Conference on Big Data (Big Data)*, IEEE, Dec. 2017, pp. 3928–3933. doi: 10.1109/BigData.2017.8258400.

[52] Xu-Ying Liu, Jianxin Wu, and Zhi-Hua Zhou, "Exploratory Undersampling for Class-Imbalance Learning," *IEEE Trans. Syst. Man, Cybern. Part B*, vol. 39, no. 2, pp. 539–550, Apr. 2009, doi: 10.1109/TSMCB.2008.2007853.

[53] T. Wang, C. Lu, W. Ju, and C. Liu, "Imbalanced heartbeat classification using EasyEnsemble technique and global heartbeat information," *Biomed. Signal Process. Control*, vol. 71, no. January, p. 103105, Jan. 2022, doi: 10.1016/j.bspc.2021.103105.

Appendices

Appendix A. Variables definition

Variable Name	Category	Academic Description
Customer Age	Demographic	Denotes the chronological age of the account holder, measured in completed years.
Gender	Demographic	Represents the biological sex of the customer, typically categorized as Male (M) or Female (F).
Dependent_count	Demographic	Indicates the number of individuals financially dependent on the account holder.
Education_Level	Demographic	Captures the highest level of formal education attained by the customer, such as high school diploma or college degree.
Marital_Status	Demographic	Refers to the legal marital condition of the account holder, including categories such as Married, Single, Divorced, or Unknown.
Income_Category	Demographic	Reflects the customer's estimated annual income, classified into ordinal income brackets (e.g., "< $40K", "$40K–$60K", etc.).
Card_Category	Product	Specifies the type or tier of credit card issued to the customer, typically segmented into categories like Blue, Silver, Gold, or Platinum.
Months_on_book	Relationship Tenure	Measures the duration (in months) since the customer began their relationship with the financial institution.
Total_Relationship_Count	Engagement	Quantifies the total number of distinct products or services held by the customer with the institution.
Months_Inactive_12_mon	Behavioral	Indicates the number of months during the past 12-month period in which the customer exhibited no account activity.
Contacts_Count_12_mon	Behavioral	Records the number of interactions or contact points between the customer and the bank within the preceding 12 months.
Credit_Limit	Financial	Represents the maximum credit limit authorized on the customer's credit card account.
Total_Revolving_Bal	Financial	Denotes the outstanding revolving balance currently carried on the credit card account.
Avg_Open_To_Buy	Financial	Reflects the average amount of available credit that the customer can still utilize, calculated over a 12-month period.
Total_Amt_Chng_Q4_Q1	Transactional	Captures the proportional change in total transaction amount from the first to the fourth fiscal quarter.
Total_Trans_Amt	Transactional	Measures the aggregate monetary value of all transactions conducted by the customer over the last 12 months.

Appendix B. Variables definition

Variable Name	Category	Academic Description
TPB_Card_Utilization	Attitude	Operationalizes customer satisfaction through credit card usage intensity, where higher utilization reflects stronger positive attitudes toward the banking relationship. Derived from Avg_Utilization_Ratio × 100.
TPB_Transaction_Engagement	Attitude	Measures average monthly transaction value as a proxy for engagement quality. Calculated as Total_Trans_Amt / (Months_on_book + 1).
TPB_Frequency_Engagement	Attitude	Reflects transactional frequency normalized by tenure, indicating consistent engagement patterns. Defined as Total_Trans_Ct / (Months_on_book + 1).
TPB_Behavior_Consistency	Attitude	Captures spending pattern consistency across fiscal periods. Higher values indicate more stable, predictable customer behavior - operationalized as abs(Total_Amt_Chng_Q4_Q1 - 1) × -1.
TPB_Product_Diversity	Subjective Norms	Reflects relational bonds through product portfolio breadth; customers with multiple products exhibit stronger institutional ties. Directly mapped from Total_Relationship_Count.
TPB_Financial_Commitment	Subjective Norms	Normalizes credit limit relative to maximum portfolio value, reflecting perceived financial commitment and stakes in the relationship. Calculated as Credit_Limit / Credit_Limit_max.

Variable Name	Category	Academic Description
TPB_Switching_Difficulty	Perceived Behavioral Control	Combines account complexity and tenure to measure perceived switching barriers. Higher values indicate greater difficulty in switching institutions. Operationalized as (Total_Relationship_Count × Months_on_book) / 100.
TPB_Account_Inactivity	Perceived Behavioral Control	Represents account dormancy over a 12-month window; greater inactivity suggests lower perceived control over the relationship. Direct measure: Months_Inactive_12_mon.
TPB_Bank_Communication	Perceived Behavioral Control	Quantifies customer-bank interaction frequency as a proxy for relationship integration and perceived control. Derived from Contacts_Count_12_mon.
SCT_Credit_Exposure	Economic	Normalizes authorized credit limit relative to portfolio maximum, operationalizing financial stake magnitude. Calculated as Credit_Limit / Credit_Limit_max.
SCT_Active_Balance	Economic	Measures revolving balance utilization relative to credit limit, reflecting active financial commitment. Operationalized as Total_Revolving_Bal / (Credit_Limit + 1).
SCT_Transaction_Investment	Economic	Normalizes cumulative transaction volume relative to portfolio maximum, reflecting total financial investment. Derived as Total_Trans_Amt / (Total_Trans_Amt_max + 1).
SCT_Account_Complexity	Procedural	Captures switching effort through product portfolio complexity; multiple products necessitate coordinated account migration. Mapped from Total_Relationship_Count.
SCT_Tenure_Lock	Procedural	Normalizes account tenure relative to maximum tenure in the dataset, operationalizing "lock-in" through institutional embeddedness. Calculated as Months_on_book / Months_on_book_max.
SCT_Integration_Level	Procedural	Combines transaction frequency and tenure to measure operational integration depth. Higher values reflect greater interdependence with the institution. Operationalized as (Total_Trans_Ct × Months_on_book) / 1000.
SCT_Relationship_Strength	Psychological	Synthesizes product diversity and tenure as a loyalty proxy; long-standing multi-product relationships develop psychological attachment. Calculated as (Total_Relationship_Count × Months_on_book) / 100.
SCT_Engagement_Loyalty	Psychological	Normalizes transaction frequency by tenure to capture consistent engagement patterns, reflecting habitual loyalty. Operationalized as Total_Trans_Ct / (Months_on_book + 1).
SCT_Card_Status	Psychological	Binary indicator of card prestige; premium card tiers (Platinum) signal status investment and psychological commitment. Encoded as Card_Category == 'Platinum' (1 = Platinum, 0 = Other).
SCT_Economic_Cost_Index	Composite	Aggregate normalized index synthesizing all economic switching costs. Calculated as (SCT_Credit_Exposure + SCT_Active_Balance + SCT_Transaction_Investment) / 3.
SCT_Procedural_Cost_Index	Composite	Aggregate normalized index synthesizing all procedural switching costs. Derived as (SCT_Account_Complexity_norm + SCT_Tenure_Lock + SCT_Integration_Level_norm) / 3.
SCT_Psychological_Cost_Index	Composite	Aggregate normalized index synthesizing all psychological switching costs. Calculated as (SCT_Relationship_Strength_norm + SCT_Engagement_Loyalty_norm + SCT_Card_Status) / 3.
CLV_Historical_Transactions	Historical Value	Normalizes transaction count relative to portfolio maximum, operationalizing past behavioral intensity. Calculated as Total_Trans_Ct / Total_Trans_Ct_max.
CLV_Historical_Revenue	Historical Value	Normalizes cumulative transaction amount relative to portfolio maximum, reflecting historical profitability. Derived as Total_Trans_Amt / Total_Trans_Amt_max.
CLV_Revenue_Consistency	Historical Value	Measures spending pattern stability across fiscal quarters; consistent behavior indicates reliable revenue streams. Operationalized as abs(Total_Amt_Chng_Q4_Q1 - 1) × -1.

Variable Name	Category	Academic Description
CLV_Account_Age_Value	Current Engagement	Normalizes tenure relative to maximum tenure in dataset, reflecting loyalty duration and relationship investment. Calculated as Months_on_book / Months_on_book_max.
CLV_Monthly_Spending	Current Engagement	Quantifies average monthly profitability; normalized by tenure to isolate spending intensity. Operationalized as Total_Trans_Amt / (Months_on_book + 1).
CLV_Transaction_Frequency	Current Engagement	Measures transaction frequency normalized by tenure, capturing current engagement intensity. Derived as Total_Trans_Ct / (Months_on_book + 1).
CLV_Credit_Capacity	Potential Value	Normalizes available credit limit relative to portfolio maximum, representing untapped revenue potential through future utilization. Calculated as Credit_Limit / Credit_Limit_max.
CLV_Product_Expansion_Potential	Potential Value	Normalizes product count relative to maximum products in dataset, operationalizing cross-sell and upsell potential. Derived as Total_Relationship_Count / Total_Relationship_Count_max.
CLV_Balance_Availability	Potential Value	Measures revolving balance utilization relative to credit limit, reflecting capacity for increased product adoption. Operationalized as Total_Revolving_Bal / (Credit_Limit + 1).
CLV_Historical_Index	Composite	Aggregate normalized index synthesizing historical profitability components. Calculated as (CLV_Historical_Transactions + CLV_Historical_Revenue + CLV_Revenue_Consistency) / 3.
CLV_Current_Index	Composite	Aggregate normalized index synthesizing current engagement components, weighted toward past behavior. Derived as (CLV_Account_Age_Value + CLV_Monthly_Spending_norm + CLV_Transaction_Frequency_norm) / 3.
CLV_Score	Composite Weighted	Comprehensive CLV assessment integrating all three value components with differential weighting: historical (70%), current (20%), and potential (10%). Formula: 0.70 × CLV_Historical_Index + 0.20 × CLV_Current_Index + 0.10 × CLV_Potential_Index.

Appendix C. Computational Environment and Reproducibility

All data processing, feature engineering, and model training experiments were conducted on a local workstation. To ensure the reproducibility of the reported results, the specific hardware specifications and software environment are as follows:

C.1. Hardware Specifications

The experiments were executed on a system equipped with an 11th-generation Intel Core processor and integrated graphics.

Component	Specification
Processor (CPU)	Intel® Core™ i5-1135G7 (4 Cores, 8 Threads, up to 4.20 GHz)
Graphics (GPU)	Intel® Iris® Xe Graphics
RAM	16 GB DDR4

C.2. Software Environment

The implementation was carried out using Python 3.10.