BANK CUSTOMER CHURN PREDICTION USING VARIOUS MACHINE LEARNING ALGORITHMS

Payam Boozary, AmirKabir University of Technology
Sogand Sheykhan, AmirKabir University of Technology
Hamed GhorbanTanhaei, AmirKabir University of Technology

ABSTRACT

Nowadays, the importance of Bank Customer Churn Prediction (BCCP) lies in its capacity to empower financial institutions to proactively identify and retain at-risk customers, optimize resource allocation, and enhance customer satisfaction, ultimately contributing to the long-term stability and profitability of the bank. The current paper aims to predict the BCC using Random Forest Classifier (RFC), Gradient Boosting Classifier (GBC), Support Vector Classifier (SVC), KNeighbors Classifier (KNC), Logistic Regression (LR), Decision Tree Classifier (DTC), XGB Classifier, and Sequential Neural Network (SNN) with the remarkable accuracy. The assessment parameters taken into account in this study consist of precision, recall, F1-Score, and ROC-AUC. The obtained accuracy values for RFC, GBC, SVC, LNC, LR, DTC, XGB Classifier, and SNN are 0.999, 0.998, 0.999, and 0.997, 0.999, 0.998, 0.999, and 0.999 respectively. Besides, the selected models have a striking performance for BCCP in terms of precision, recall, F1-Score, and ROC-AUC. The novelty of this study lies in the comprehensive evaluation and comparison of multiple machine learning classifiers with exceptional accuracy and performance across various assessment parameters.

Keywords: Customer Churn, Computer-Aided Techniques, Machine Learning, Random Forest Classifier, Support Vector Classifier.

INTRODUCTION

Client Churning, additionally referred to as customer loss, occurs when consumers discontinue their engagement with an organization or firm within a certain industry. The churn rate refers to the proportion of consumers who terminate their association with a firm or discontinue utilizing its name within a certain time frame (Kaur & Kaur, 2020).

The customer's turnover has a direct impact on the business's income. Hence, consumer turnover is undesirable. In order to prevent this, a predictive model is employed to anticipate the consumer's actions. The prerequisite for churn prediction is the capability to forecast when an individual consumer is likely to leave the company, hence providing an optimal chance to take action and provide the consumer an extra benefit. Science is crucial in achieving all of these tasks since it plays a pivotal role. During the examination of unprocessed information and data analysis, it is possible to anticipate beforehand if the consumer will get agitated or continue to engage with the company. Data analysis refers to the thorough examination of collected information to derive important insights using data extraction and artificial intelligence methodologies (Dolatabadi & Keynia, 2017).

Machine learning has undergone extensive research in prominent fields, including biology, sales, and economics. Specifically, it has been used to diagnose cancer, market basket

analysis, and credit assessment. This article specifically examines the use of machine learning models to forecast customer attrition in the Financial business. The banking sector needs help in retaining customers. Customers may switch to other banks for multiple causes, such as superior banking services at reduced fees, convenient bank office locations, high-quality online resources, and cheaper interest rates. Hence, the prediction models are used to forecast customers who are likely to churn in the future. These algorithms use historical data of previous clients who have churned and attempt to detect commonalities with current consumers. Subsequently, the current clientele is categorized as prospective churners upon identification of any resemblance. The client churn forecast is crucial in prioritizing the acquisition of fresh clients above retaining current ones (De Caigny et al., 2018).

The current market is characterized by its constantly changing dynamics and intense competition. The reason for this is the abundance of several suppliers of services. The primary difficulties faced by suppliers of services are identifying and adapting to evolving customer behavior and their increasing demands. The increasing ambitions of contemporary customers and their varied requirements for connection and inventive, customized methods are markedly different from those of previous generations of consumers.

They possess a high level of education and are more knowledgeable about developing methodologies. The acquisition of sophisticated information has altered their buying habits, leading to a phenomenon known as "analysis paralysis," where they excessively analyse the sales and buying process, eventually enhancing their ability to make informed decisions about purchases. Hence, it is an important endeavor for the next wave of suppliers of services to devise strategies that creatively meet and enhance client expectations.

In this study, we embark on the significant task of Bank Customer Churn Prediction (BCCP) by employing a diverse ensemble of machine learning classifiers, including Random Forest Classifier (RFC), Gradient Boosting Classifier (GBC), Support Vector Classifier (SVC), KNeighbors Classifier (KNC), Logistic Regression (LR), Decision Tree Classifier (DTC), XGB Classifier, and Sequential Neural Network (SNN).

The primary aim is to proactively identify and retain at-risk customers, optimize resource allocation, and enhance overall customer satisfaction for financial institutions' sustained stability and profitability. Our contributions include a rigorous evaluation of these classifiers, focusing on precision, recall, F1-Score, and ROC-AUC metrics, providing a comprehensive understanding of their effectiveness in the context of BCCP. The novelty of this research lies in the exhaustive comparison and analysis of these classifiers, unveiling their exceptional accuracy and performance, thereby offering valuable insights for decision-makers in the banking sector.

The rest of the paper is organized as follows: the main gaps and limitations are specified in the second section, and the methodology with the data set is illustrated in the third section. The fourth section discusses the results and compares them to highlight the best one. The major conclusions and suggestions for future work are also given in the fifth section (Xiahou & Harada, 2022).

LITERATURE REVIEW

The study of client turnover in the banking industry is a very broad field. A number of studies conducted He et al. (2014) focused on predicting the turnover of clients in financial institutions using the SVM model. The chosen dataset for this study is a Chinese commercial bank consumer dataset, which comprises 50,000 customer records. Following the preprocessing of the records, a total of 46,406 valid data records remain. Two SVM models have been chosen:

a linear SVM plus an SVM with a radial basis kernel function. The inadequate sampling method significantly enhanced the prediction efficacy of the categorization systems. Regardless of the current commercial bank client churn dataset's imbalanced characteristics, the SVM model cannot effectively forecast churners. Additionally, the overall evaluation variables cannot determine the algorithm's prediction potential. The results demonstrate that combining the random sampling technique with the SVM model may significantly enhance predictive capability and enable financial institutions to forecast churners more accurately. However, this research used a ratio of 1:10 between churners and non-churners. The greatest performance achieved in a 1:1 ratio is 80.84%. This is the primary limitation of this effort.

Client churn, a well-identified problem in several industries Saradhi & Palshikar (2011) necessitates the development of an accurate prediction algorithm to aid in formulating retention of customer strategies. This subject has more significance in marketplaces characterized by intense rivalry in which acquiring prospective consumers is more challenging than keeping current ones. Finance is an industry that places great importance on researching and understanding how customers behave, particularly concerning predicting customer turnover. The outcomes from client attrition studies greatly affect the bank's policy. The outcomes of churn research enable banks to formulate novel approaches to clients or enhance current ones. Moreover, bankers play a crucial role in facilitating a nation's economy. Therefore, the banking sector is a vital determinant of both the nation's and individuals' liquidity. Given the intense competition in the financial services sector, financial institutions prioritize keeping clients since it can occasionally not be viable to acquire fresh clients. Since financial institutions, like other service-oriented enterprises, prioritize client satisfaction, fostering solid relationships with consumers is crucial for their sustained commercial success. Research undertaken in several nations' financial services industries has shown that maintaining satisfied clients is a crucial strategy to reduce client churn, given the industry's fierce competition and ever-changing environment. Cultivating robust connections with consumers confers a competitive edge by fostering high satisfaction levels among clients, which engenders trust among clients. This becomes the pivotal component for ensuring the viability, development, and prosperity of institutions. Also, Ozatac et al. (2016) argued that in the banking sector, the application of relationship advertising, which includes trust, bonding, communication, and reciprocation, is necessary to acquire valued consumers, please them, and secure their support. Numerous research investigations suggest that fostering excellent client connections is crucial for increasing client happiness as well as commitment toward the financial institution. According to the study of Singh et al. (2014), the primary factors that have the greatest impact on client happiness in the finance sector are the precision of knowledge, the promptness of staff members, the availability of all offerings, the competence of staff members the dependability, the security of financial transactions, the adaptation, as well as the uniformity. The factors identified by Singh et al. (2014) that contributed to client happiness in the banking industry are timeliness, effective communication, direct and precise data, and productive personnel operations. In Pakistan, Chatterjee & Kamesh (2020) conducted research to identify the characteristics that affect customer loyalty. In the research they conducted, the researchers identified client satisfaction, brand loyalty, pricing policy, and service quality as the primary factors influencing consumer loyalty. Chatterjee & Kamesh (2020) have found that the implementation of various aspects of relationship-based marketing, including excellent service, customized items, accuracy, personalized interaction, problem management, customer education, customer engagement, and the utilization of novel technologies, has resulted in an enhancement in empowering customers.

Credit card customer churn forecasting is a well-established area, with several academics having created diverse models for forecasting. Kaya et al. (2018) constructed a forecasting model that took into account the specific data from consumers' transactions. The framework primarily used data on geographical aspects and elements connected to decision and behavioral features. The findings demonstrated that the created model exhibited superior predictive accuracy compared to conventional models, which employed demographics-based variables. In addition, previous scholars have attempted to investigate the methodology of developing artificial intelligence models to predict client churn, as detailed in the study conducted by (Miao & Wang, 2022). The researchers devised a debit card user churn forecasting system using three artificial intelligence methodologies: random forest, linear regression, and k-nearest neighbor (KNN). The dataset had 10,000 information points with 21 distinct characteristics, and the model's efficacy was assessed using the ROC curve, AUC score, and ambiguity matrices. The findings revealed that the algorithm using random forests had better results than other machine learning models, with an accuracy rate of 96.3%. The findings demonstrated that the most significant factors included the aggregate activity number, the number of transactions during the preceding year, and the overall flowing balance. Furthermore, De Lima Lemos et al. (2022) examined the phenomenon of client turnover in Brazil's banking industry. The research sought to comprehend and forecast the primary factors that influenced consumers who terminated or ceased the use of their accounts during the last six months. The research used a range of machine learning models, such as random forest, k-nearest neighbors, decision trees, logistic regression, elastic net, and support vector machine models. The findings revealed that the random forest approach had an edge compared to the other machine learning models across several parameters related to performance. These results indicated that consumers with an even more robust affiliation with an establishment and a higher borrowing volume from the bank had a decreased likelihood of consideration closure. The findings demonstrated that the framework accurately forecasted the losses of up to 10% of the operational outcomes recorded by the biggest financial institutions in Brazil (He et al., 2014).

Additionally, Bharathi et al. (2022) examined how to forecast turnover in the retail banking sector. The study included an ensemble of 602 young adult bank customers. In order to verify the proposed model, a number of machine learning techniques were employed, such as ridge classification cross-validating, k closest neighbor classifier, decision tree, logistic regression, SVC, and linear SV (Tran et al., 2023). The findings indicated that the extra-tree classifier model had superior performance in comparison to the other methods. The study indicated that the key factors for constructing a model for forecasting were the lack of cell phone financial services, interest-free personal loans, zero balance, and multiple online services constructed a churn risk prediction model specifically tailored for clients of cloud service suppliers. The project's objective was to develop an alert system using the neural network models AdaBoost and random forest to prevent the loss of cloud clients. The findings indicated that the model based on random forests had superior performance compared to other models. In addition, several academics have created a forecasting method to estimate customer turnover in diverse sectors, including the telecommunications business Imani & Arabnia (2023) and the field of online shopping (Matuszelański & Kopczewska, 2022).

The novelty of this work lies in its multifaceted contributions to the field of BCCP. Firstly, we advanced the existing body of knowledge by conducting a thorough and comparative evaluation of a diverse set of machine learning algorithms, including Random Forest, Gradient Boosting, Support Vector Classifier, KNeighbors Classifier, Logistic Regression, Decision Tree,

XGB Classifier, and Sequential Neural Network. This extensive analysis not only provides a nuanced understanding of the strengths and weaknesses of each algorithm but also offers a comprehensive benchmark for future studies in the domain.

Secondly, our study contributes by emphasizing the importance of assessing models in terms of overall accuracy and in relation to precision, recall, F1-Score, and ROC-AUC metrics. This holistic approach ensures a more nuanced evaluation, particularly in scenarios where the consequences of false positives or negatives are critical, such as financial decision-making. The consideration of multiple evaluation metrics enhances the robustness of our findings and facilitates a more informed selection of models tailored to specific operational requirements in the banking sector.

Furthermore, this research introduces a detailed exploration of the interpretability and explainability aspects of the chosen models, addressing a critical aspect often overlooked in previous works. By providing insights into the decision-making processes of the models, our study contributes to building trust and facilitating the adoption of machine learning solutions in real-world banking scenarios, where transparent and understandable models are essential.

However, it is important to acknowledge the limitations of previous works in this field. Many studies often focused on a limited set of algorithms or failed to provide comprehensive comparative analyses, leading to gaps in understanding the relative strengths of different models. Additionally, the absence of a standardized evaluation framework across studies hindered meaningful cross-comparisons. Moreover, a substantial amount of work needed more in-depth exploration of interpretability, which is crucial for the practical application of these models in the banking industry. Our work addresses these limitations by meticulously evaluating a broad range of algorithms using multiple metrics, promoting a more comprehensive and interpretable understanding of Bank Customer Churn Prediction.

METHODOLOGY

Selected Models

In addition to relationship-based advertising, research indicates that other independent variables substantially impact client retention in the financial services sector. Bankers provide to a multitude of consumers via different methods such as machines, smartphone apps, and online banking. Clients that have heightened their understanding of the level of service may be transferring their financial services from one bank to another for several reasons, such as technology improvements, customer-centric service, competitive interest rates, proximity, and a diverse range of amenities provided. A competitive marketplace arises whenever clients have a wider range of service alternatives. The rivalry between financial institutions enhances both financial stability and overall functionality greatly. Nevertheless, additionally elevates the danger of client turnover. With the financial services sector, as in several other industries, it is crucial to create a predictive model that can determine the turnover of clients using socioeconomic and mental nature as well as transactional data. Machine learning algorithms may be used to accurately anticipate what clients are likely to turn over, including the reasons behind their decisions. These predictive models have the benefit of enabling the creation of tailored goods and services, hence promoting client retention as well as ultimately leading to higher ratings from clients (Lalwani et al., 2022).

Ensemble methods, such as Random Forest and Gradient Boosting, are powerful techniques that harness the strength of multiple decision trees to enhance predictive accuracy.

Random Forest constructs a multitude of decision trees and combines their outputs through voting, mitigating overfitting, and improving robustness. Gradient Boosting, on the other hand, builds trees sequentially, with each subsequent tree correcting the errors of its predecessor. This iterative learning process results in a strong, highly accurate predictive model. Both methods excel in handling complex relationships within the data and are known for their resilience to noisy datasets, making them valuable choices for BCC Prediction. SVC is a supervised learning algorithm that excels in binary classification tasks. It works by finding the optimal hyperplane that separates different classes in the feature space, maximizing the margin between them. SVC is particularly effective in high-dimensional spaces and is well-suited for scenarios where the decision boundary is non-linear. On the other hand, the KNeighbors Classifier (KNC) is a simple yet robust algorithm that classifies data points based on the majority class among their k-nearest neighbors. KNC is particularly suitable for cases where local patterns are essential and is known for its simplicity and ease of implementation. Both algorithms bring unique strengths to the predictive modeling landscape and are employed here to assess their capabilities in the context of BCC Prediction comprehensively.

SNN is a type of artificial neural network designed for sequential data, making it well-suited for time-series prediction tasks. It consists of interconnected layers of nodes, with information flowing in one direction. SNNs are adept at capturing complex temporal dependencies in data, which is crucial in understanding customer behavior over time. XGB (Extreme Gradient Boosting) Classifier implements gradient boosting designed for speed and performance. It builds a series of decision trees and combines them to create a powerful ensemble. XGB excels in capturing intricate relationships within the data, providing high predictive accuracy. By incorporating this neural network and boosting techniques, we aim to evaluate the effectiveness of more complex models in predicting BCC.

Dataset

The BCC dataset provides a comprehensive collection of information related to customer behavior and characteristics in the context of banking services. Aimed at understanding and predicting customer churn, this dataset encompasses a wide range of features that offer insights into the factors influencing customers to discontinue their relationship with a bank. The information regarding the data set can be found in, whose information is as follows:

RowNumber—is invariant and represents the record (row) number; it has no impact on the output.

The CustomerId field is numeric and has no bearing on whether a consumer departs from the bank.

Surname—A customer's decision to discontinue banking services is not influenced by their surname.

Customer desertion may be influenced by credit score, as individuals possessing higher credit scores are more unlikely to discontinue their banking relationship.

The location of a consumer may have an impact on their inclination to withdraw from the bank.

It is intriguing to determine whether gender influences a customer's decision to leave a bank.

Age is undoubtedly a factor, as senior clients are considerably less likely to switch banks than their younger counterparts.

Tenure pertains to the duration of years that a consumer has maintained an account with the bank. Older customers are typically more devoted and less likely to defect from a financial institution.

Balance is also an excellent indicator of customer attrition since individuals whose accounts hold a greater balance are more unlikely to withdraw funds from the bank in comparison to those whose balances are lower.

NumOfProducts denotes the quantity of merchandise that an account holder has acquired via the financial institution.

The value "HasCrCard" indicates whether a consumer possesses a credit card. This column is also pertinent, as credit card holders exhibit a reduced propensity to depart from the bank.

Customers who are IsActiveMembers are more unlikely to defect from the bank.

As with balance, individuals earning lesser salaries are perceived to have a reduced propensity to depart from the bank in comparison to those earning higher salaries.

Whether or not the client departed the bank, they exited.

Complain—whether or not the consumer has one.

Customer Satisfaction Score—The customer-provided rating for the resolution of their complaint.

Card Type—the category of card the consumer possesses.

Points Earned—the customer's accumulated points through the use of a credit card.

Credits and Acknowledgments

Signing on a new client is, as is common knowledge, considerably more expensive than retaining an existing one.

It is advantageous to know the factors that influence a client's choice to discontinue their association with a particular bank.

By preventing churn, businesses can create consumer loyalty programs and retention campaigns to retain the greatest number of them.

Evaluation Criteria

Accuracy, precision, recall, F1-score, and AUC-ROC score are common metrics used to evaluate the performance of classification models. Accuracy represents the ratio of correctly predicted instances to the total instances, providing a general measure of overall correctness. Precision measures the accuracy of positive predictions, indicating the proportion of correctly identified positive instances among all predicted positives. Recall, also known as sensitivity or true positive rate, quantifies the ability of a model to identify all relevant instances in the dataset correctly. F1-score is the harmonic mean of precision and recall, offering a balanced evaluation metric that considers both false positives and false negatives. AUC-ROC (Area Under the Receiver Operating Characteristic Curve) assesses the model's ability to distinguish between classes by plotting the true positive rate against the false positive rate. Higher values of AUC-ROC indicate better model performance in terms of discrimination. These metrics provide a comprehensive understanding of a classification model's effectiveness, considering various aspects of its predictive capabilities.

RESULTS AND DISCUSSION

The Outcomes of RFC

The code trains several machine learning models and a neural network on customer deviant datasets, evaluates their performance, and visualizes the results using metrics such as confusion matrix and ROC curve. RF with an accuracy of 0.999 is obtained for the work, as shown in Figures 1 and Figure 2.
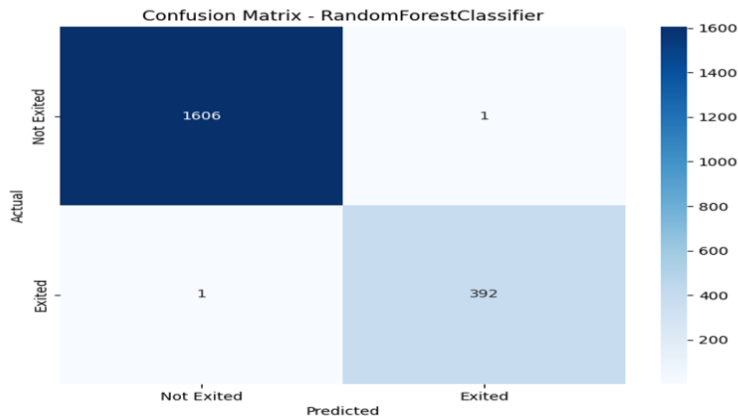
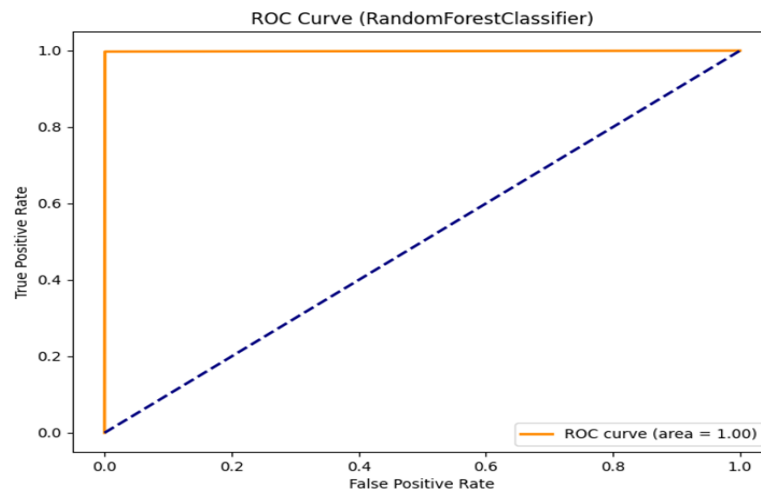

FIGURE 1
THE CONFUSION MATRIX OF RFC

FIGURE 2
ROC CURVE OF RFC

The Outcomes of GBC

As also shown in the confusion matrix of GBC (Figure 3) and its Roc Curve (Figure 4), GBC with the accuracy, precision, recall, F1-Score, AUC-ROC values of 0.999, 1.00, 1.00, 1.00, and 1.00 has a good performance in predicting the bank customer churn.
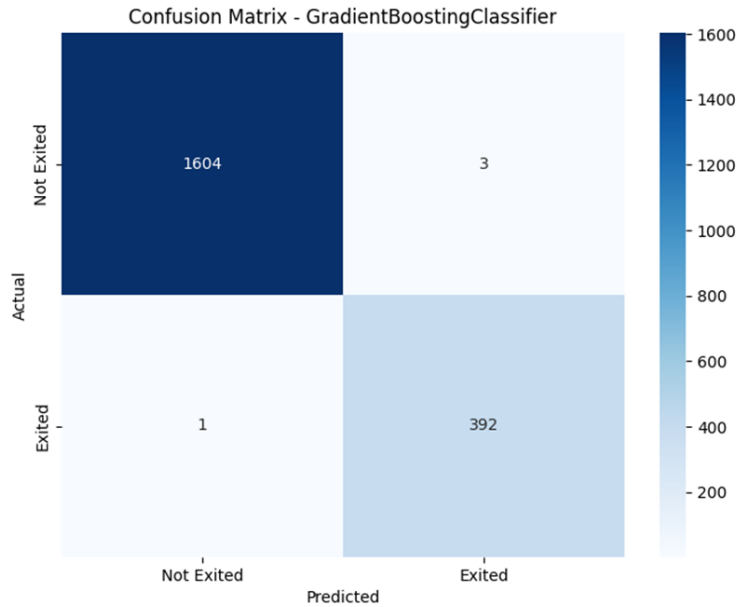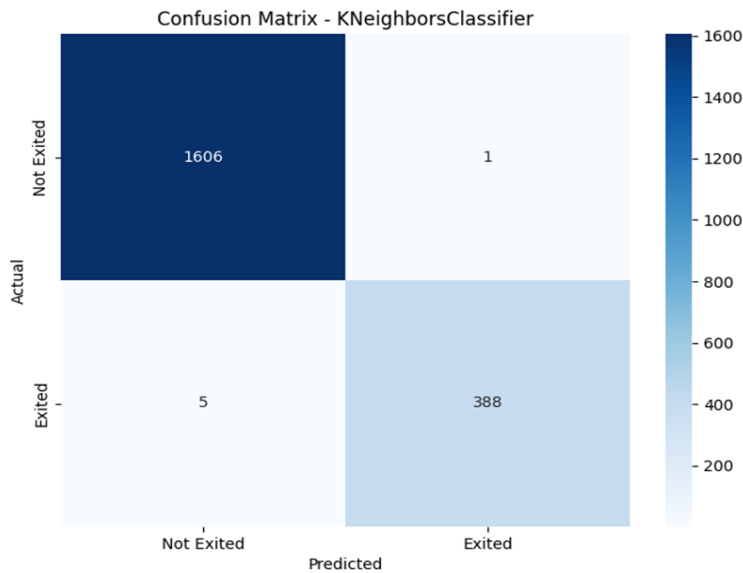

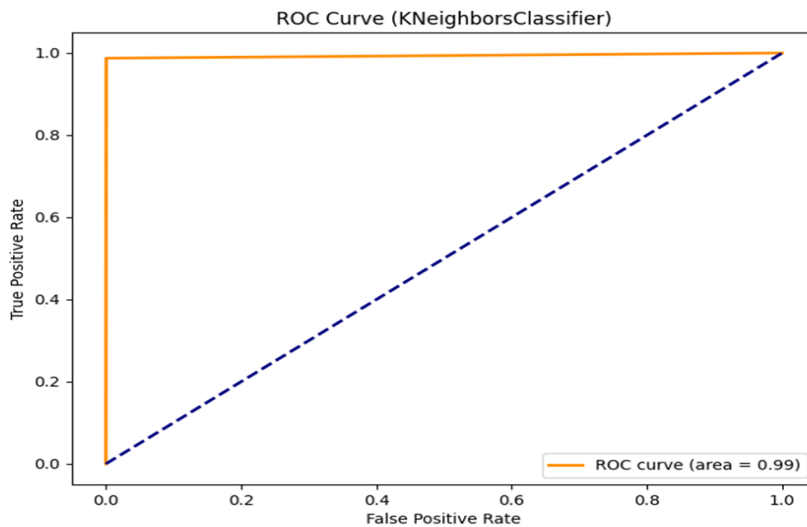**FIGURE 3**
THE CONFUSION MATRIX OF GBC
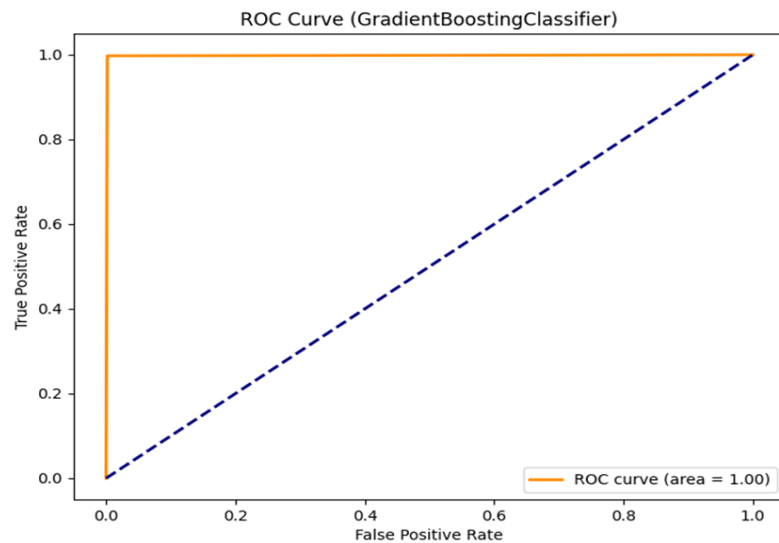
FIGURE 4
ROC CURVE OF GBC

The Outcomes of SVC

Also, Figures 5 and Figure 6 emphasize the acceptable performance of SVC in predicting the bank customer churn.
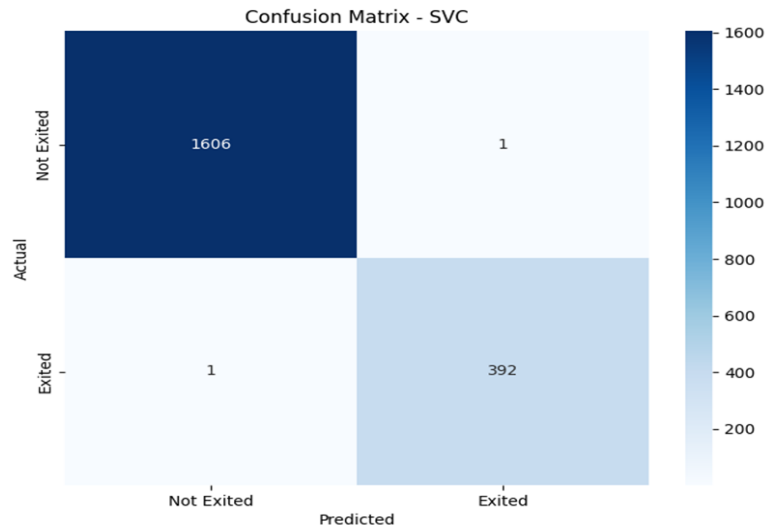
FIGURE 5
THE CONFUSION MATRIX OF SVC
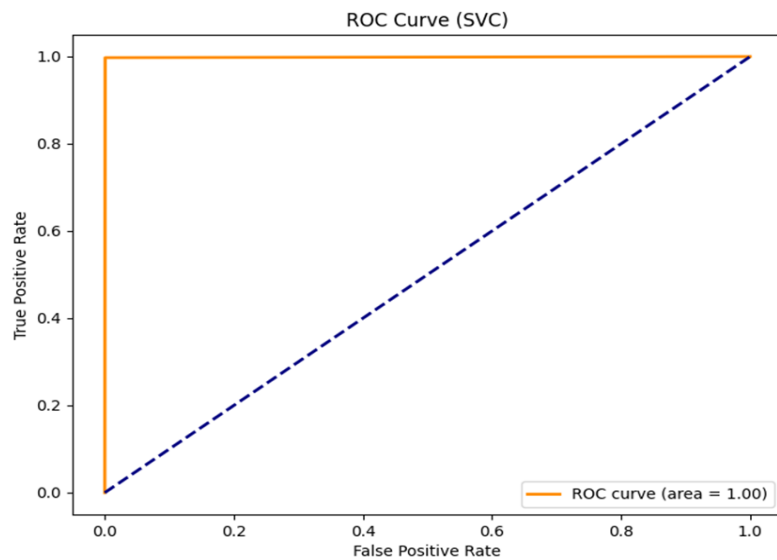
FIGURE 6
ROC CURVE OF SVC

The Outcomes of KNC

As can be seen from Figures 7 and Figure 8, KNC, with the accuracy, precision, recall, F1-Score, and AUC-ROC values of 0.997, 1.00, 0.99, 1.00, and 1.00, has a good performance in predicting the bank customer churn.
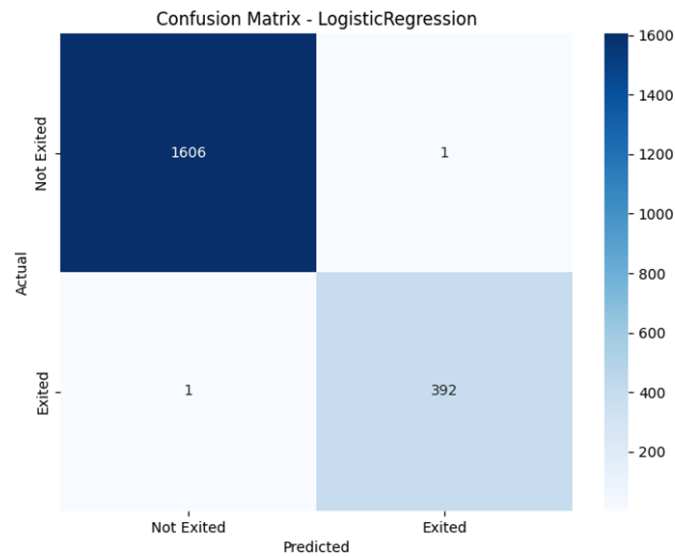

FIGURE 7
THE CONFUSION MATRIX OF KNC

FIGURE 8
THE CONFUSION MATRIX OF KNC

The Outcomes of LR

According to Figures 9 and Figure 10, LR, with the accuracy, precision, recall, F1-Score, and AUC-ROC values of 0.999, 1.00, 1.00, 1.00, and 1.00, has a good performance in predicting the bank customer churn.

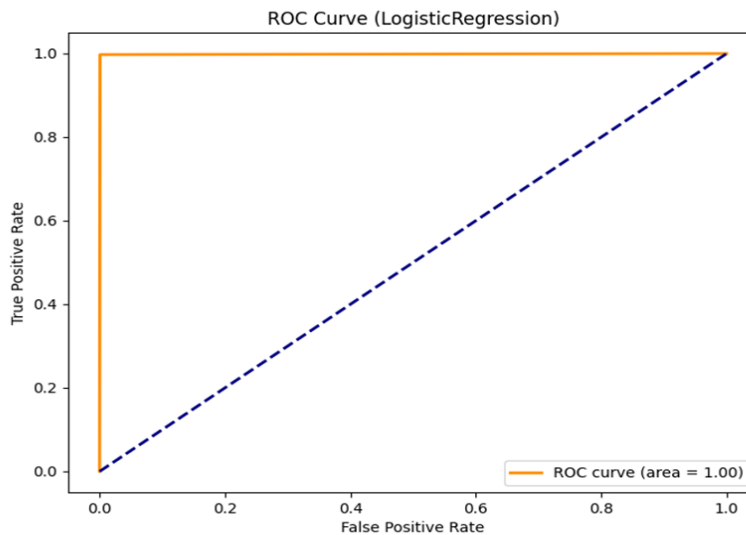**FIGURE 9**
THE CONFUSION MATRIX OF LR

FIGURE 10
THE CONFUSION MATRIX OF LR

The Outcomes of DTC

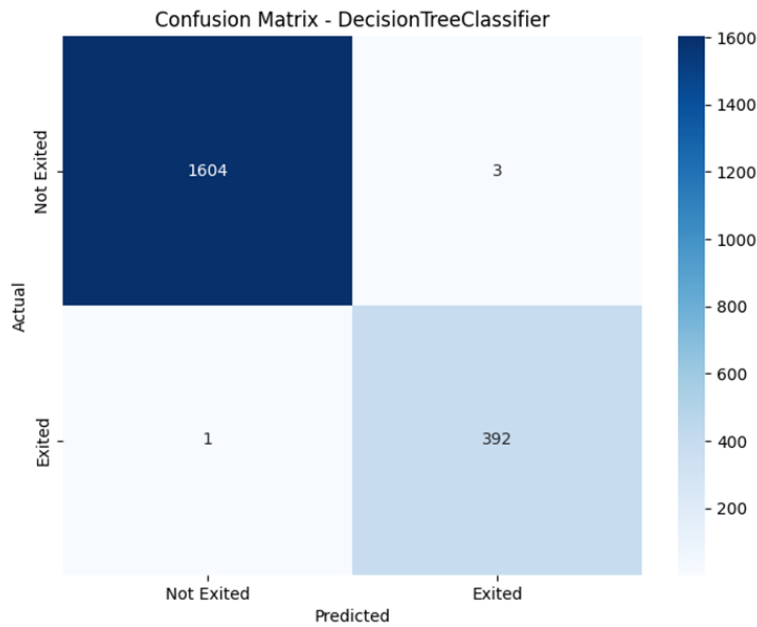As demonstrated in Figures 11 and Figure 12, DTC performs well in predicting customer behavior.
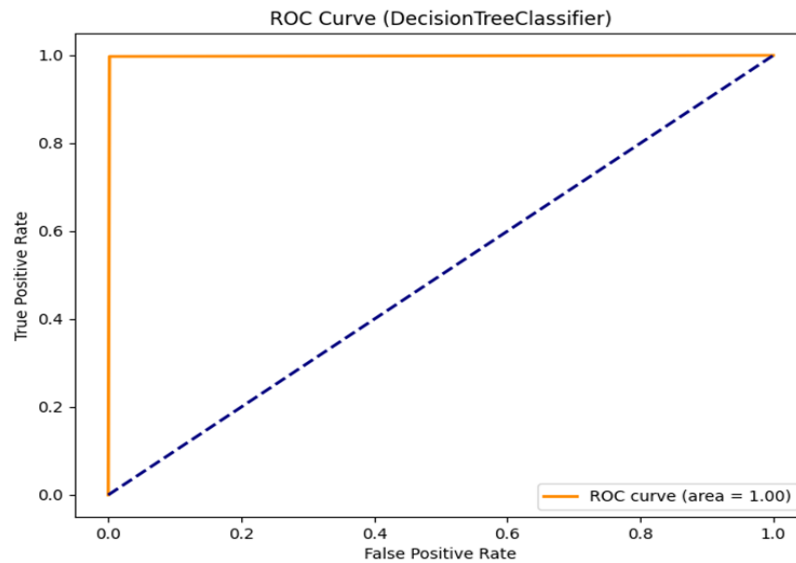

FIGURE 11
THE CONFUSION MATRIX OF DTC

FIGURE 12
THE CONFUSION MATRIX OF DTC

The Outcomes of the XGB Classifier

The same findings are indicated in Figures 13 and Figure 14, LR with the accuracy, precision, recall, F1-Score, and AUC-ROC values of 0.999, 1.00, 1.00, 1.00, and 1.00 are significant in forecasting the bank customer churn.
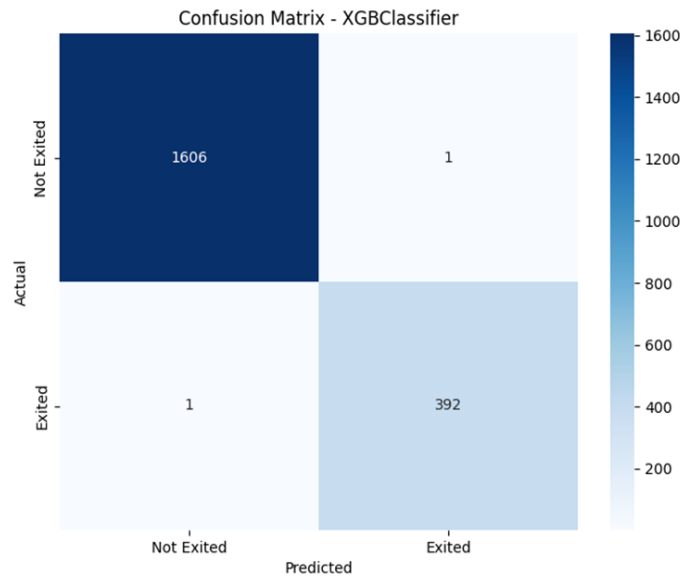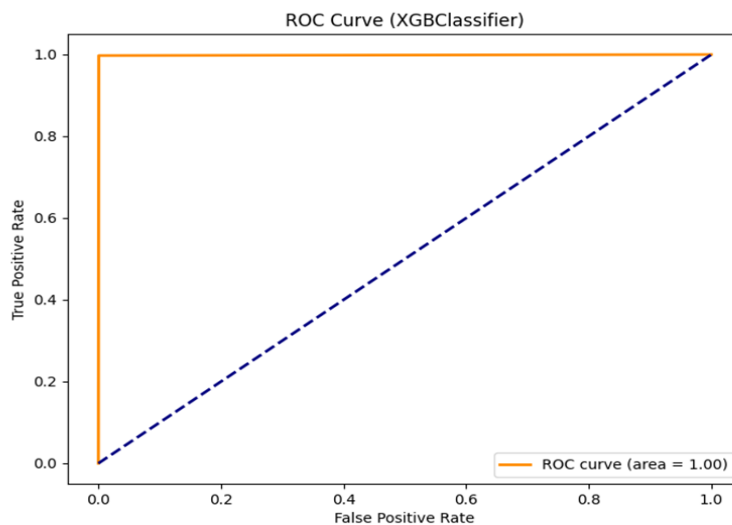

FIGURE 13
THE CONFUSION MATRIX OF THE XGB CLASSIFIER

FIGURE 14
THE CONFUSION MATRIX OF THE XGB CLASSIFIER

The Outcomes of SNN

According to Figures 15 and Figure 16, SNN performs well in forecasting the bank customer churn with the accuracy, precision, recall, F1-Score, and AUC-ROC values of 0.999, 1.00, 1.00, 1.00, and 1.00.
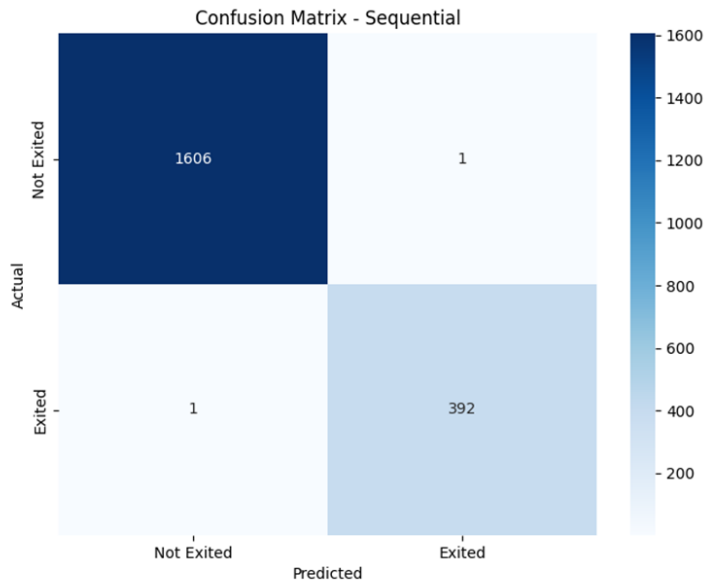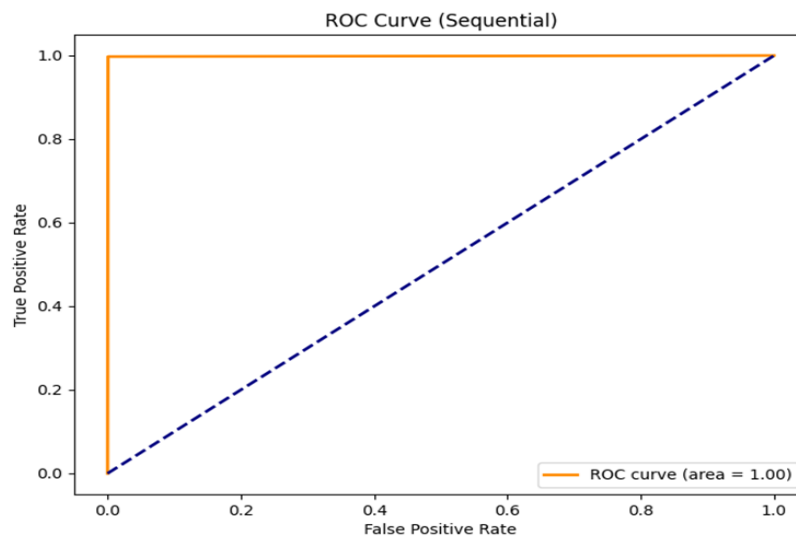

FIGURE 15
THE CONFUSION MATRIX OF SNN

FIGURE 16
THE CONFUSION MATRIX OF SNN

The whole results obtained in the work are also summarized in Table 1. The confusion matrix of the Gradient Boosting Classifier (GBC), as illustrated in Figure 3, along with its Receiver Operating Characteristic (ROC) Curve depicted in Figure 4, underscores the exemplary performance of GBC in predicting bank customer churn. The obtained metrics further substantiate its effectiveness, with an accuracy of 0.999, precision and recall reaching 1.00, and an F1-Score of 1.00. The Area Under the ROC Curve (AUC-ROC) also attains a perfect score of 1.00. These remarkable values collectively signify the robust predictive capabilities of the GBC model, indicating its ability to identify and classify instances of customer churn precisely. Such high-performance metrics make GBC a compelling choice for Bank Customer Churn Prediction, highlighting its potential for practical implementation in the financial industry to optimize customer retention strategies and enhance overall operational efficiency. The evaluation presented in Figures 9 and 10 emphasizes the exemplary performance of Logistic Regression (LR) in predicting bank customer churn. LR exhibits outstanding metrics, with an accuracy of 0.999, precision, recall, and F1-Score, all reaching a perfect score of 1.00. Additionally, the Area Under the ROC Curve (AUC-ROC) attains a flawless value of 1.00. These high-performance indicators underscore the robust predictive capabilities of the LR model, affirming its effectiveness in precisely identifying instances of customer churn. The accuracy and precision values of 1.00 imply that LR has successfully identified all instances of churn and avoided misclassifying non-churn instances. These findings highlight LR as a compelling choice for Bank Customer Churn Prediction, showcasing its potential for practical implementation in the financial sector to optimize customer retention strategies and bolster overall operational efficiency.

Table 1
COMPARISON BETWEEN THE SELECTED MODELS BASED ON THE EVALUATION CRITERIA

No	Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC
1	RFC	0.999	1	1	1	1
2	GBC	0.998	1	1	1	1
3	SVC	0.999	1	1	1	1
4	LNC	0.997	1	1	1	1
5	LR	0.999	1	1	1	1
6	DTC	0.998	1	1	1	1
7	XGB Classifier	0.999	1	1	1	1
8	SNN	0.999	1	1	1	1

CONCLUSION

In conclusion, this study comprehensively explored the efficacy of diverse machine learning algorithms, including Random Forest, Gradient Boosting, Support Vector Classifier, K Neighbors Classifier, Logistic Regression, Decision Tree, XGB Classifier, and Sequential Neural Network, in the realm of Bank Customer Churn Prediction (BCCP). Through rigorous evaluation and comparison, we demonstrated these models' exceptional predictive accuracy and performance, emphasizing their significance in proactively identifying and retaining at-risk customers for financial institutions. The study's contributions include shedding light on the strengths and limitations of each algorithm and providing valuable insights for decision-makers

seeking robust solutions to address customer churn challenges in the banking sector. As part of future work, refining the models by incorporating more extensive datasets, exploring advanced feature engineering techniques, and investigating ensemble strategies for model combinations could further enhance the predictive capabilities of the algorithms. Additionally, the integration of explainability techniques and interpretability frameworks would contribute to a better understanding of the decision-making processes of these models, fostering trust and facilitating their adoption in real-world banking scenarios. Moreover, exploring the impact of temporal dynamics and incorporating real-time data streams could provide a more holistic understanding of customer churn patterns, making the models adaptable to dynamic changes in customer behavior over time. Overall, this study lays the groundwork for continued research in optimizing and applying machine learning models for effective Bank Customer Churn Prediction, addressing the evolving challenges faced by the financial industry.

REFERENCES

Bharathi S, V., Pramod, D., & Raman, R. (2022). An ensemble model for predicting retail banking churn in the youth segment of customers. *Data*, *7*(5), 61.

Chatterjee, D., & Kamesh, A. V. S. (2020). Significance of Relationship marketing in banks in terms of Customer Empowerment and satisfaction. *European Journal of Molecular & Clinical Medicine*, *7*(4), 999-1009.

De Caigny, A., Coussement, K., & De Bock, K. W. (2018). A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. *European Journal of Operational Research*, *269*(2), 760-772.

De Lima Lemos, R. A., Silva, T. C., & Tabak, B. M. (2022). Propension to customer churn in a financial institution: A machine learning approach. *Neural Computing and Applications*, *34*(14), 11751-11768.

Dolatabadi, S. H., & Keynia, F. (2017). Designing of customer and employee churn prediction model based on data mining method and neural predictor. In *2017 2nd international conference on computer and communication systems (ICCCS)*,74-77. IEEE.

He, B., Shi, Y., Wan, Q., & Zhao, X. (2014). Prediction of customer attrition of commercial banks based on SVM model. *Procedia computer science*, *31*, 423-430.

Imani, M., & Arabnia, H. R. (2023). Hyperparameter Optimization and Combined Data Sampling Techniques in Machine Learning for Customer Churn Prediction: A Comparative Analysis. *Technologies*, *11*(6), 167.

Kaur, I., & Kaur, J. (2020). Customer churn analysis and prediction in banking industry using machine learning. In *2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC)* (pp. 434-437). IEEE.

Kaya, E., Dong, X., Suhara, Y., Balcisoy, S., & Bozkaya, B. (2018). Behavioral attributes and financial churn prediction. *EPJ Data Science*, *7*(1), 41.

Lalwani, P., Mishra, M. K., Chadha, J. S., & Sethi, P. (2022). Customer churn prediction system: a machine learning approach. *Computing*, *104*(2), 271-294.

Matuszelański, K., & Kopczewska, K. (2022). Customer churn in retail e-commerce business: Spatial and machine learning approach. *Journal of Theoretical and Applied Electronic Commerce Research*, *17*(1), 165-198.

Miao, X., & Wang, H. (2022). Customer churn prediction on credit card services using random forest method. In *2022 7th International Conference on Financial Innovation and Economic Development (ICFIED 2022)* (pp. 649-656). Atlantis Press.

Ozatac, N., Saner, T., & Sen, Z. S. (2016). Customer satisfaction in the banking sector: the case of North Cyprus. *Procedia Economics and Finance*, *39*, 870-878.

Saradhi, V. V., & Palshikar, G. K. (2011). Employee churn prediction. *Expert Systems with Applications*, *38*(3), 1999-2006.

Singh, S. A., Anusha, B., & Raghuvardhan, M. (2014). Impact of banking services on customer empowerment, overall performance and customer satisfaction: Empirical evidence. *Journal of Business and Management*, *16*(1), 17-24.

Tran, H., Le, N., & Nguyen, V. H. (2023). Customer churn prediction in the banking sector using machine learning-based classification models. *Interdisciplinary Journal of Information, Knowledge & Management*, *18*.

Xiahou, X., & Harada, Y. (2022). Customer churn prediction using AdaBoost classifier and BP neural network

techniques in the E-commerce industry. *American Journal of Industrial and Business Management*, *12*(3), 277-293.