Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 13 NO: 1 PAGE: 335-345 YEAR: 2024
DOI:10.17798/bitlisfen.1408349

Using Machine Learning Algorithms to Analyze Customer Churn with Commissions Rate for Stocks in Brokerage Firms and Banks

Hakan KAYA*

Istanbul Okan University, Graduate School, Banking Doctorate (PhD) Program, 34959 Tuzla / ISTANBUL / Türkiye
(ORCID: 0000-0002-0812-4839)

Keywords: Customer churn, stock commission rates, brokerage firms and banks, machine learning

Abstract

Stock commission rates of banks and brokerage firms are a critical factor for investors. These rates affect the cost of stock investments. It's crucial to highlight the significance of stock commission rates in brokerage firms and banks, as well as the factors that influence their determination. This article aims to draw attention to the study's focus on customer churn and commission rates within the financial industry. Previous research has mainly focused on identifying the key variables affecting customer churn without considering its impact on forecast accuracy. This work has two primary research goals: first one is to investigate how commission rates affect the accuracy of customer churn prediction in brokerage firms and the banking sector using machine learning models, and second one is to compare and evaluate the most effective machine learning approaches for predicting customer churn. The customer churn management approach was enhanced through the analysis of a data set obtained from a bank and brokerage firm. This data set, comprised of 7816 entries and 14 columns, reflects the firm's transactions and was sourced from a publicly accessible database. The analysis employed Decision Tree, Random Forest, K-NN, Gaussian NB, and XGBoost algorithms to evaluate performance using three accuracy measures. Two approaches are included for model creation. According to the first analysis results, the Gaussian NB, for second approach the K-NN algorithms gave the best result.

1. Introduction

Institutions can gain valuable insights by analyzing customer data and behavior to identify potential churn risks. Utilizing predictive analytics and machine learning algorithms, institutions can proactively pinpoint at-risk customers and take necessary actions to prevent churn.

Churn analysis refers to the process of identifying customers who may leave a company and implementing appropriate marketing precautions.

Brokerage firms and banks typically earns a portion of the commissions sale of stocks that the makes for an investor. Anticipating and mitigating customer churn presents a significant opportunity for additional revenue generation for both brokerage firms and banks. In the banking and brokerage sectors, various machine learning algorithms are commonly used for analyzing customer churn.

Stock commission is a fee that investors incur when they buy or sell stocks, charged by brokerage firms or banks involved in stock transactions. Typically, stock commission is calculated as a percentage of the investment amount. It's worth noting that commission rates can vary among brokerage firms, banks, and countries. Moreover, these rates may differ based on factors such as the type and amount of the transaction, as well as the investor's client status. In addition to the

*Corresponding author: hkaya5@stu.okan.edu.tr
Received: 22.12.2023, Accepted: 28.02.2024

commission, exchange fees and a 5% Banking and Insurance Transactions Tax (BITT) are collected based on the commission amount. Commission rates are automatically deducted from the accounts of investors engaging in stock transactions.

What are some common machine learning algorithms used to analyze customer churn in the banking and brokerage industry? Can you provide examples of successful customer churn prevention strategies implemented by businesses using predictive analytics? answers to these questions were sought. Additionally, how businesses can use predictive analytics and machine learning algorithms to identify at-risk customers and prevent customer churn is discussed.

With the rapid advancements in machine learning, it has become increasingly valuable to develop predictive models that can determine whether a customer will leave a specific bank. In this study, we aim to perform data analysis and develop a churn prediction model to forecast which customers are likely to churn based on historical data and various features. Machine learning techniques such as Decision Trees, Random Forests, K-NN, Gaussian NB, and XGBoost algorithms are utilized to identify churn patterns [1]. These ML methods are not only employed in the banking sector but also applied across various industries including insurance [2], medical systems [3], cyberbullying [4], retail marketing [5], automobile and gaming industry [3], and telecommunication [6]. Therefore, the significance of this study can be summarized in three aspects: i) gathering churn customer data from approximately 8,000 institutions, ii) overcoming class imbalance issues by using a combination of SMOTE data sampling and the K-NN classifier during the initial stage of analysis, and iii) conducting a comprehensive assessment of five models to facilitate informed model selection and evaluation [1].

Many machine learning techniques can research customer churn analysis. In this study, banking was compared with brokerage firms. Here is a brief overview of the related work on the available methods: A recent study employed [7] on utilizing methods such as stochastic gradient boosting, random forest, logistic regression, and k-nearest neighbors to predict early customer churn and develop effective retention strategies. In another study [8] developed an approach for customer churn prediction utilizing three intelligent models: AdaBoost, Random Forest (RF), and Support Vector Machine (SVM). To address the issue of an unbalanced dataset, the team applied the Synthetic Minority Oversampling Technique (SMOTE) along with a combination of undersampling and oversampling. Another study conducted [9] the topic of customer churn prediction across various domains such as telecommunications, energy sectors, retail banking, e-banking, and insurance was explored. The modelling techniques utilized in this field include Logistic Regression, Neural Network Model, Random Forest, Decision Tree, Support Vector Machine, and Rough set approach. These methods have been implemented to detect churn among customers. Other work in the literature [10] SHapely Additive exPlanations (SHAP) values to improve the assessment and interpretability of the machine learning model. The research aimed to develop an interpretable machine learning model using authentic banking industry data and assess various machine learning models using test data. A similar study [11] explored the techniques such as k-means clustering for customer segmentation, as well as logistic regression, k-nearest neighbors, random forest, decision tree, and support vector machine algorithms to analyze the dataset. The literature [12] took a slightly different approach to customer churn management by analyzing a dataset obtained from a real-world telecommunication firm. To analyze the dataset, the researchers employed Artificial Neural Network (ANN), Support Vector Machines (SVM), and Naive Bayes (NB) algorithms. The performance of the analysis was evaluated using four accuracy measures. In the given study [13] emphasized the prediction of customer churn within the banking sector using a distinctive customer-level dataset acquired from a prominent Brazilian bank. The researchers compared various machine learning models including decision trees, elastic net, k-nearest neighbors, logistic regression, random forests and support vector machines. For example, in one study in the literature [6] provides offer readers a general understanding of the commonly used data mining methods, their results, and performance, while also highlighting the need for further research in the Telecommunication Industry. In contrast, another study [14] the analysis involved examining the customer behavior data of a real water purifier rental service within an electronics company in Korea to create and validate a churn prediction model. The model's performance was assessed using the F-measure and area under curve (AUC) metrics.

The rest of this paper is structured as follows: Section 2 outlines the research

methodology. The experimental findings are presented in Section 3. Finally, Section 4 concludes the paper and recommendations to the instutions to captures the development of a predictive model to pinpoint influential factors contributing to customer churn, encompassing the testing of different machine learning algorithms, the use of SMOTE to address class imbalance, and the subsequent repetition of the initial approach's steps.

2. Material and Method

This research focused on analyzing brokerage firms and banks in Turkey, specializing in stock commission rates. A dataset of 7816 observations was utilized. The study initially evaluated the weights of fourteen features and then proceeded to predict non-churn and churn customers using Decision Tree, Random Forest, K-NN, Gaussian NB, and XGBoost machine learning methods. The performance metrics considered in this study included Accuracy, Recall, Precision, and F1-Score.

The research was conducted using anonymous data from brokerage firms and banks, ensuring that only data from customers with permission to process data was anonymized and made available for analysis under the scope of Personal Data Protection Authority. Prior to creating and training the model, the dataset's features were extracted and formatted appropriately for model creation and prediction.

2.1. Used Algorithms

2.1.1.Decision Tree

A decision tree is a flowchart-like structure with nodes representing features, branches representing decision rules, and leaf nodes indicating the outcome or class label.

The algorithm recursively divides the data based on selected features, creating subsets with shared class labels. This process continues until a stopping condition is met, such as reaching a maximum depth or having a minimum number of instances in each leaf [18].

2.1.2.Random Forest

Random Forest is a popular machine learning algorithm that combines decision trees with an ensemble learning approach. It can handle classification and regression tasks effectively, offering high accuracy and robustness.

In Random Forest, multiple decision trees are built, each trained on a different subset of data and a random subset of features. This randomness helps prevent overfitting and improves the model's ability to generalize to new data. During training, each tree is grown similarly to traditional decision trees, but with only a random subset of features considered for splitting at each node [19]. To make predictions, the algorithm combines the individual tree predictions through voting (for classification) or averaging (for regression).

2.1.3.K-NN

K-Nearest Neighbors (K-NN) is a popular machine learning algorithm used for both classification and regression tasks. It makes predictions based on the similarity of data points.

In K-NN, the "K" represents the number of nearest neighbors considered for predictions. To determine proximity, K-NN uses distance measures like Euclidean or Manhattan distance. The algorithm calculates distances between the new data point and all others in the training set, selecting the K nearest neighbors. Choosing the right K value is crucial to avoid overfitting, where the model becomes too sensitive to noise in the data [20].

2.1.4.Gaussian NB

Gaussian Naive Bayes (GNB) is a popular machine learning algorithm for classification tasks. It assumes that features follow a Gaussian distribution and are independent of each other.

GNB estimates the mean and standard deviation of each feature for each class during training. To make predictions, it calculates the conditional probability of each class given the feature values using Bayes' theorem. GNB is computationally efficient, works well with high-dimensional data, and handles missing values effectively [21]. However, if features are strongly correlated, GNB may not perform as well as other algorithms.

2.1.5.XGBoost

XGBoost, short for Extreme Gradient Boosting, is a powerful and efficient machine learning algorithm that belongs to the gradient boosting framework. It combines weak prediction models, usually decision

trees, to create a robust and accurate model. The algorithm iteratively trains and adds new models to improve upon the errors made by the previous ones, using a technique called gradient boosting [17].

To enhance its speed and efficiency, XGBoost incorporates various optimizations. These include parallel computing, column block loading, and approximate algorithms for finding splits. These optimizations make XGBoost one of the fastest and most scalable gradient boosting frameworks available.

2.2.Used Performance Metrics

Here are the definitions of the four terms related to binary classification:

True negatives (TN): These are the instances that are correctly predicted as negatives (zeros).

True positives (TP): These are the instances that are correctly predicted as positives (ones).

False negatives (FN): These are the instances that are incorrectly predicted as negatives (zeros) when they are actually positives (ones).

False positives (FP): These are the instances that are incorrectly predicted as positives (ones) when they are actually negatives (zeros).

2.2.1.Accuracy

Accuracy can be defined as the proportion of correct predictions to the total number of predictions made by the system [15].

$$Accuracy = (TP+TN) / (TP+TN+FN+FP) \quad (1)$$

2.2.2.Precision

Precision is a measure of success when predicting positive outcomes. It is calculated as the ratio of true positives (TP) to the sum of true positives and false positives (TP + FP).

Precision represents the classifier's ability to correctly identify samples as positive and not falsely label negative samples as positive.

$$Precision = TP / (TP + FP) \quad (2)$$

2.2.3.Recall

Recall is a metric that measures the success of predicting positive outcomes [16]. It is calculated as the ratio of true positives (TP) to the sum of true positives and false negatives (TP + FN).

Recall represents the classifier's ability to identify all positive samples, capturing the proportion of positive instances correctly detected by the model.

$$Recall = TP / (TP + FN) \quad (3)$$

2.2.4.F1-Score

The F1-Score is a metric that combines the recall and precision into a single value. It is calculated as the harmonic mean of the recall and precision values.

The F1-Score provides a balanced measure of a classifier's performance, taking into account both the ability to correctly identify positive samples (recall) and the ability to avoid false positives (precision). The F1 score can be computed by using the formula: 2 * (Precision * Recall) / (Precision + Recall)

3. Results and Discussion

To gain a deeper understanding and improve our ability to predict customer attrition, our primary focus will be on the target variable "Exited." This variable serves as a vital indicator of churn and will guide our efforts in building a robust churn model.

3.1. Categorical Variables

3.1.1.Exited

The variable "Exited" serves as an important indicator, distinguishing between existing customers (0) and churned customers (1), and providing valuable insights into customer retention.

The dataset exhibits with around 80% of the data representing existing customers and the remaining 20% representing churned customers.

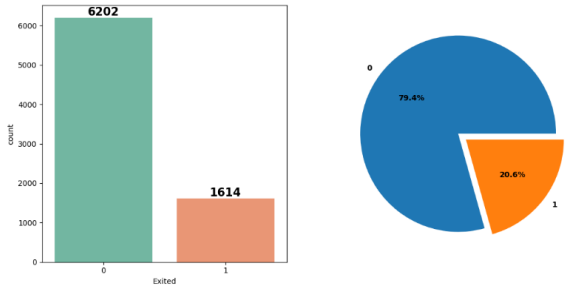

Figure 1. Customer Churned Disribution

3.1.2.Institutions

The dataset includes customer churn data from two types of institutions: Banks and Brokerage Firms, providing a comprehensive view of customer behavior within these industries.

In terms of customer distribution, banks represent 60% of the customer base, while Brokerage Firms make up the remaining 40% of institutions. This distribution allows for a well-rounded analysis and understanding of churn patterns in both banks and brokerage firms, enabling us to develop insights and strategies that cater to the specific characteristics of each type of institution.

Table 1. Distribution of Customer in Institutions

Banks	60.22262
Brokerage Firms	39.77738

Interestingly, the findings of both analyses are remarkably similar to one another on churn analysis.

Table 2. Distribution of Customer Churn in Institutions

Brokerage Firms	0	79.125121
	1	20.87487
Banks	0	79.498619
	1	20.501381

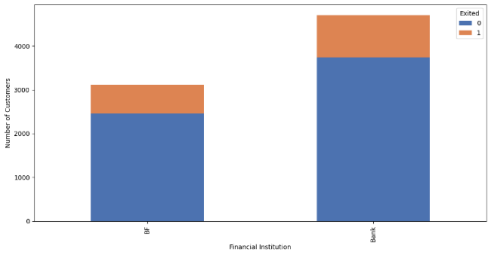

Figure 2. Financial Institutions and Customer Churn Distribution

3.1.3.Gender

The customer base is divided between male and female customers, with males representing 54% and females making up 46% of the total customer population.

Table 3. Distribution of Customer Gender

Male	54.183726
Female	45.816274

Churn rates by gender is intriguing to note that 26% of female customers have churned, indicating a higher attrition rate in this group. On the other hand, the churn rate for male customers is 16%, suggesting relatively lower customer churn within this segment.

Table 4. Distribution of Customer Churn Rates between Genders

Female	0	74.280927
	1	25.719073
Male	0	83.636364
	1	16.363636

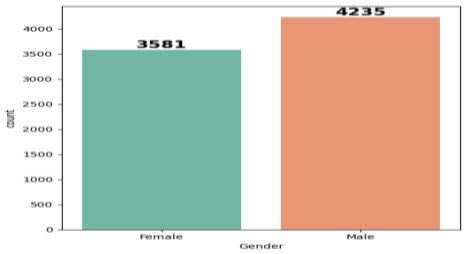

Figure 3. Gender and Customer Churn Distribution

3.1.4.Tenure

The duration of customer tenure ranges from 0 to 10 years, with an average tenure of 5 years. This suggests a wide spectrum of customer relationships established with the institutions over time.

Table 5. Distribution of Tenure

count	7816.000000
mean	5.013946
std	2.882681
min	0.000000
25%	3.000000
50%	5.000000
75%	7.000000
max	10.000000

The majority of customers, accounting for around 90%, maintain a tenure that spans from 1 year to 9 years. This concentration signifies a strong sense of customer loyalty within this timeframe, underscoring the importance of fostering long-term relationships with clients.

Table 6. Distribution of Concentrated Tenure Range

0	4.183726
1	10.273797
2	10.248209
3	10.158649
4	9.992323
5	10.145855
6	9.608495
7	10.427329
8	10.491300
9	9.825998
10	4.644319

3.1.5. Number of Products

Customers exhibit a diverse range of product ownership, typically ranging from 1 to 4 products. This indicates varying levels of engagement with the bank's offerings, showcasing the breadth of options available to cater to individual needs and preferences.

Table 7. Distribution of Number of Products

count	7816.000000	**1**	4002
		2	3551
mean	1.527764	**3**	215
std	0.584068	**4**	48
min	1.000000		
25%	1.000000		
50%	1.000000		
75%	2.000000		
max	4.000000		

No. Of Prod.	**Exited**	
1	0	71.91
	1	28.09
2	0	92.59
	1	7.41
3	0	83.26
	1	16.74
4	1	100.00

Churned customers are worth mentioning that all customers who had 4 card products have churned, suggesting a strong likelihood of attrition within this particular group. Furthermore, a substantial percentage (83%) of customers with 3 card products has also churned, underscoring the significance of conducting a thorough analysis of churn patterns specific to each product.

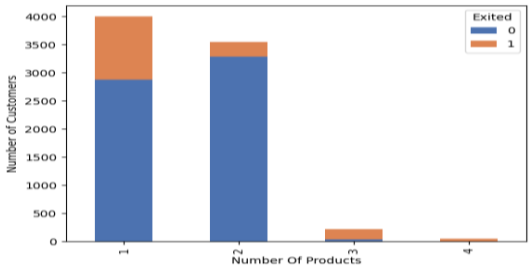

Figure 4. Number of Products and Customer Churn Distribution

3.1.6. Has Credit Cards

Around 70% of customers are found to possess a credit card, highlighting a substantial portion of the customer base actively utilizing this financial product. This demonstrates the widespread adoption and popularity of credit cards among the customer community.

Table 8. Distribution of Has Credit Cards

count	7816.000000	**0**	2323
		1	5493
mean	0.702789		
std	0.457059		
min	0.000000		
25%	0.000000		
50%	1.000000		
75%	1.000000		
max	1.000000		

Has Credit Cards	**Exited**	
0	0	78.82
	1	21.18
1	0	79.57
	1	20.43

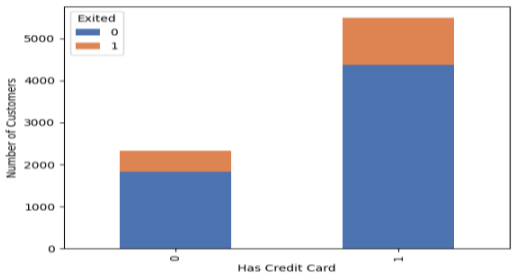

Figure 5. Has Credit Card and Customer Churn Distribution

3.1.7. Is Active Member

Roughly 52% of customers are actively enrolled as members, indicating a significant portion of the customer base actively participating in the bank's membership program.

Within the group of active members, approximately 14% have churned, while among non-active members, the churn rate stands at around 28%. The number of active and non-active customers is approximately equal. However, it is noteworthy that the churn rate among non-active customers is nearly double compared to active customers. This suggests that customers who are not actively engaged in the institutions's offerings are more likely to deactivate their banking facilities.

Table 9. Distribution of Is Active Member

count	7816.000000	**0**	3814
		1	4002
mean	0.512027		
std	0.499887		
min	0.000000		
25%	0.000000		
50%	1.000000		
75%	1.000000		
max	1.000000		

IsActive Member	**Exited**	
0	0	72.65
	1	27.35
1	0	85.73
	1	14.27

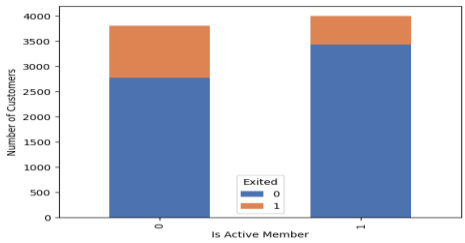

Figure 6. Is Active Member and Customer Churn Distribution

3.1.8. Credit Score

The credit scores of customers span a range from 350 to 850, indicating the diversity in their creditworthiness.

Table 10. Distribution of Credit Score

count	7816.000000
mean	650.099667
std	96.826809
min	350.000000
25%	583.000000
50%	652.000000
75%	717.000000
max	850.000000

A noteworthy observation is that all customers with credit scores below 300 have churned. This underscores the importance of credit scores as a potential factor contributing to customer attrition. It emphasizes the significance of addressing credit-related issues in order to enhance customer retention and mitigate churn risk.

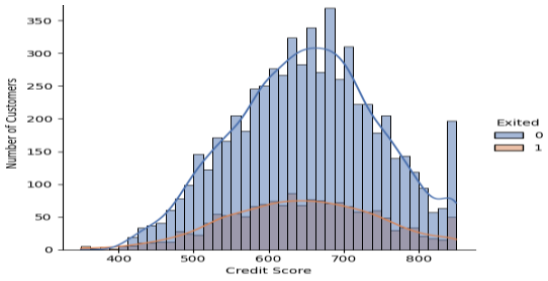

Figure 7. Credit Score and Customer Churn Distribution

3.1.9. Age

The age range of customers extends from 18 to 92, with an average age of approximately 38. This signifies a diverse customer base encompassing a wide range of age groups.

Table 11. Distribution of Age

count	7816.000000
mean	38.946392
std	10.517831
min	18.000000
25%	32.000000
50%	37.000000
75%	44.000000
max	92.000000

It has been observed that customers in the age range of 50 to 58 have shown a higher churn rate. This highlights the necessity of implementing targeted retention strategies specifically tailored for customers within this age group.

Additionally, the distribution of customer age is right-skewed (positively skewed), indicating that there are relatively fewer customers in older age groups. This skewness emphasizes the importance of considering age as a potential factor in predicting churn and underscores the need to implement suitable measures to retain customers across all age ranges.

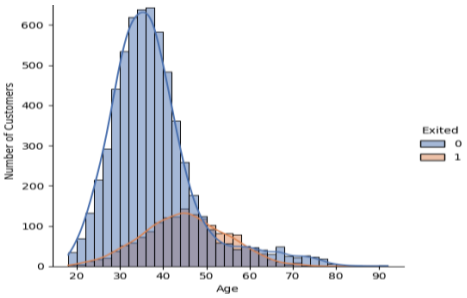

Figure 8. Age and Customer Churn Distribution

3.1.10. Balance

Currently, there are over 2300 customers who have a balance of zero in their accounts. When excluding these customers with zero balances, the distribution of account balances follows a normal distribution pattern. It is worth noting that customers with a zero balance are more prone to deactivating their accounts.

Table 12. Distribution of Balance

count	7816.000000
mean	76762.958515
std	62418.094005
min	0.000000
25%	0.000000
50%	97703.005000
75%	127811.165000
max	250898.090000

This implies that the remaining customers, who maintain a positive balance, exhibit a more typical distribution of account balances. This enables the bank to gain a clearer understanding of their financial status and effectively manage their accounts.

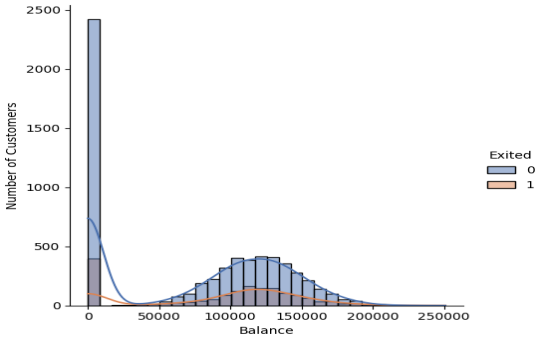

Figure 9. Balance and Customer Churn Distribution

3.1.11. Estimated Salary

The salary range of our customers is quite diverse, with estimates ranging from $11 to 199k. This indicates a significant variation in income levels among our customer base.

Table 13. Distribution of Estimated Salary

count	7816.000000
mean	100239.979496
std	57463.265346
min	11.580000
25%	51413.442500
50%	100405.680000
75%	149216.320000
max	199992.480000

The salary distribution among our customers shows a relatively equal spread across the entire range, indicating a balanced representation of customers across various income levels. Additionally, the distribution of customer churn is also evenly distributed, suggesting that salary alone is not a significant factor in customer churn.

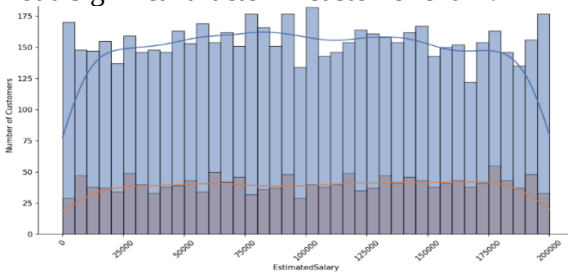

Figure 10. Estimated Salary and Customer Churn Distribution

3.2. Model Creation

This section covers two approaches, and as a result, a Predictive Model was developed to identify the key factors that have a significant impact on customer churn. The research method involves phases from data collection to data preparation and pre-processing to make it usable in the model. This includes commission rates to validate the hypothesis and addressing imbalanced data using SMOTE, a common technique for enhancing the sample size of the minority group when there is an imbalance.

Figure 11. Architecture of the Proposed Method

3.2.1 Approach 1

The dataset was split into 80% for training (6252) and 20% for testing (1564) purposes. This separation allowed us to create distinct sets for training and testing the models. To evaluate various models and determine the best performing one, cross-validation was employed.

The Gaussian NB models were utilized to generate predictions on the test data. Furthermore, it was noted that the dependent variable (Exited) exhibited an imbalance in its distribution.

Table 14. Tested Classifiers and Performance Metrics

Classifier	Average		
	Accuracy	Precision	Recall
Decision Tree	0.6731	0.5042	0.5049
Random Forest	0.6737	0.5042	0.5047
K-NN	0.7570	0.5066	0.5025
Gaussian NB	0.7935	0.3968	0.5000
XGBoost	0.7833	0.4986	0.5007

Table 15. Gaussian NB Models Evaluation using Different Metric Values

	precision	recall	f1-score	support
0	0.79	1.00	0.88	1241
1	0.00	0.00	0.00	323
accuracy			0.79	1564
macro avg	0.40	0.50	0.44	1564
weighted avg	0.63	0.79	0.70	1564

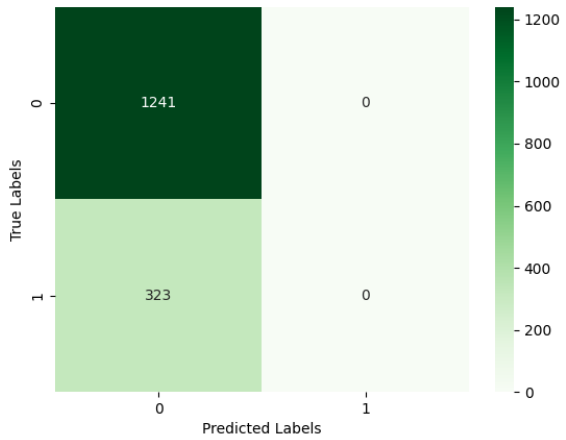

Figure 12. Model Evaluation using Confusion Matrix of Gaussian NB

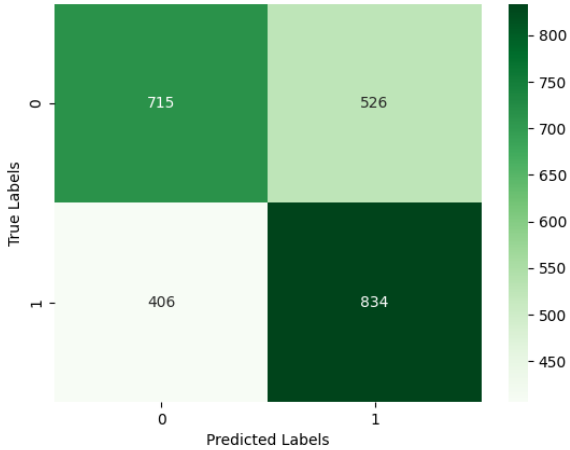

Figure 13. Model Evaluation using Confusion Matrix of K-NN

3.2.2. Approach 2

To address the issue of unbalanced datasets, the Synthetic Minority Oversampling Technique (SMOTE) was employed. SMOTE selects instances that are close in the feature space and creates synthetic samples along the line connecting those instances. By applying SMOTE, the class imbalance in the target variables was mitigated, resulting in 2481 balanced instances. Following this, the steps outlined in approach 1 were executed.

Table 16. Tested Classifiers and Performance Metrics Using the SMOTE Technique

Classifier	Average		
	Accuracy	**Precision**	**Recall**
Decision Tree	0.5557	0.5557	0.5557
Random Forest	0.5732	0.5760	0.5742
K-NN	0.6123	0.6130	0.6123
Gaussian NB	0.5060	0.5066	0.5060
XGBoost	0.6024	0.6029	0.6024

Table 17. K-NN Model Evaluation using Different Metric Values

	precision	**recall**	**f1-score**	**support**
0	0.64	0.58	0.61	1241
1	0.61	0.67	0.64	1240
accuracy			0.62	2481
macro avg	0.63	0.62	0.62	2481
weighted avg	0.63	0.62	0.62	2481

Several key factors have been identified as significant contributors to the deactivation of customers' facilities. These factors include institutions, gender, is active member, age and number of products.

The model used to predict customer deactivation achieved impressive results, with F1 score, recall, and precision values, all approximately 0.71. This suggests that the model has a acceptable ability to correctly identify positive cases while minimizing false positives and maximizing true positives.

Overall, the model exhibits strong performance across multiple evaluation metrics, showcasing its effectiveness in accurately predicting customer deactivation and capturing the desired outcomes.

4. Conclusion & Recommendations

The focus of the study is to draw attention to customer churn and commission rates in financial institutions. The analysis of the impact of stock commission rates on customer churn in financial institutions reveals vital insights into the dynamics of customer retention.

This article analyzes five different machine learning algorithms to identify key factors that influence customer deactivation in the brokerage firms and banks institutions. The Decision Tree, Random Forest, K-NN, Gaussian NB, and XGBoost algorithms are commonly used in churn analysis due to their ability to perform comprehensive analysis by combining different data types. While each algorithm operates differently, they excel at capturing complex relationships and interactions

within the data, making them valuable for churn analysis.

The predictive model aimed to identify the critical factors influencing customer churn. In the initial approach, five distinct machine learning algorithms were tested as classifiers using various performance metrics. The Gaussian NB model demonstrated a 0.7935% higher accuracy compared to the other models. Subsequently, in the second approach, 2481 balanced samples were generated by employing SMOTE to address the class imbalance in the target variables. The KNN model exhibited a 0.6123% higher accuracy compared to the alternative models. Following this, the steps outlined in the first approach were executed once again.

The study highlights that institutions, gender, active membership, age, and the number of products significantly impact churn rates. To address this, targeted initiatives and tailored retention strategies are recommended for specific segments, such as the 50-58 age group.

When examining the impact of stock commission rates on customer churn in financial institutions, some concepts to consider are: Proactive strategies can be implemented to identify and target customers at risk of churn. It can leverage predictive analytics and machine learning algorithms to predict customer behavior and intervene before churn occurs. To validate your findings and optimize your models, consider pilot testing your approach with a smaller group of customers before full implementation. Continuously evaluate the correlation between commission adjustments, customer retention, and revenue generation to strike an optimal balance.

Tailoring commission structures to individual customer needs is crucial for increasing satisfaction and loyalty. By offering customized pricing models that align with each customer's unique needs and behaviors, financial institution can significantly boost satisfaction and loyalty. Focus on enhancing overall customer satisfaction through personalized service, transparent communication, and value-added offerings. Invest in customer support resources and educational materials to empower customers and foster long-term relationships.

In this research presents a powerful approach to combating customer churn in financial institutions using a combination of proactively targeting, personalizing, and measuring. In research paves the way for a future-proof customer retention strategy in the financial sector.

Statement of Research and Publication Ethics

The study is complied with research and publication ethics.

References

[1] M. A. H. Farquad, V. Ravi, and S. B. Raju, "Data mining using rules extracted from SVM: An application to churn prediction in bank credit cards," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 5908 LNAI, pp. 390–397, 2009.

[2] N. A. Akbar, A. Sunyoto, M. Rudyanto Arief, and W. Caesarendra, "Improvement of decision tree classifier accuracy for healthcare insurance fraud prediction by using Extreme Gradient Boosting algorithm," in *Proceedings-2nd International Conference on Informatics, Multimedia, Cyber, and Information System, ICIMCIS 2020*, pp. 110–114, 2020, DOI: 10.1109/ICIMCIS51567.2020.9354286.

[3] S. M. Fati, A. Muneer, N. A. Akbar, and S. M. Taib, "A continuous cuffless blood pressure estimation using tree-based pipeline optimization tool," *Symmetry*, vol. 13, no. 4, 2021, DOI: 10.3390/sym13040686.

[4] A. Muneer and S. M. Fati, "A comparative analysis of machine learning techniques for cyberbullying detection on twitter," *Future Internet*, vol. 12, no. 11, pp. 1–21, 2020, DOI: 10.3390/fi12110187.

[5] M. Al-Ghobari, A. Muneer, and S. M. Fati, "Location-aware personalized traveler recommender system (lapta) using collaborative filtering knn," *Computers, Materials and Continua*, vol. 69, no. 2, pp. 1553–1570, 2021, DOI: 10.32604/cmc.2021.016348.

[6] F. Kayaalp, "Review of customer churn analysis studies in telecommunications industry," *Karaelmas Science & Engineering Journal*, vol. 7, no. 2, pp. 696-705 2017.

[7] B.Prabadevi, R. Shalini, and B. R. Kavitha, "Customer churning analysis using machine learning algorithms," *International Journal of Intelligent Networks*, vol. 4, pp. 145-154, 2023, DOI: https://doi.org/10.1016/j.ijin.2023.05.005.

[8] A. Muneer, R. F. Ali, A. Alghamdi, S. M Taib, A. Almaghthawi, and E. A Ghaleb, "Predicting customers churning in banking industry: A machine learning approach," *Indones. J. Electr. Eng. Comput. Sci*, vol.26, no.1, pp. 539-549, 2022, DOI: http://doi.org/10.11591/ijeecs.v26.i1.

[9] J.Britto, R.Gobinath, "A detailed review for marketing decision making support system ın a customer churn prediction", *Int. J. Sci. Technol. Res,*vol. 9, no. 4, pp. 3698-3702, 2020.

[10] H. Guliyev, T F. Y.atoğlu, "Customer churn analysis in banking sector: Evidence from explainable machine learning models," *Journal of Applied Microeconometrics,* vol. 1, no. 2, pp. 85-99, 2021, DOI: 10.53753/jame.1.2.03.

[11] H. Tran, N. Le, and V. H.Nguyen, "Customer churn prediction in the banking sector using machine learning-based classification models," *Interdisciplinary Journal of Information, Knowledge & Management,* vol. 18, pp. 87-105, 2023, DOI: https://doi.org/10.28945/5086.

[12] O. Kaynar, M. F. Tuna, Y. Görmez, , and M. A Deveci, "Makine öğrenmesi yöntemleriyle müşteri kaybı analizi," *Cumhuriyet Üniversitesi İktisadi ve İdari Bilimler Dergisi,* vol. 18, no. 1, pp. 1-14, 2017.

[13] R. A. de Lima Lemos, T. C. Silva, and B. M. Tabak, "Propension to customer churn in a financial institution: a machine learning approach," *Neural Comput. Appl*., vol. 34, no. 14, pp. 11751–11768, 2022, DOI: https://doi.org/10.1007/s00521-022-07067-x.

[14] Y. Suh, "Machine learning based customer churn prediction in home appliance rental business," *J. Big Data*, vol. 10, no. 1, 2023., DOI: https://doi.org/10.1186/s40537-023-00721-8.

[15] S. Naseer, S. M. Fati, A. Muneer, and R. F. Ali, "IAceS-deep: Sequence-based identification of acetyl Serine sites in proteins using PseAAC and deep neural representations," *IEEE Access*, vol. 10, pp. 12953–12965, 2022, DOI: https://doi.org/10.1109/access.2022.3144226

[16] A. Muneer and S. M. Fati, "Efficient and automated herbs classification approach based on shape and texture features using deep learning," *IEEE Access*, vol. 8, pp. 196747–196764, 2020, DOI: https://doi.org/10.1109/access.2020.3034033

[17] T.Chen, C. Guestrin, Xgboost: Reliable large-scale tree boosting system. In *Proceedings of the 22nd SIGKDD Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA* (pp. 13-17), 2015.

[18] V.G. Costa, C.E Pedreira,. "Recent advances in decision trees: an updated survey," *Artif Intell Rev* 56, pp. 4765–4800, 2023. DOI: https://doi.org/10.1007/s10462-022-10275-5

[19] L. Breiman, Random Forests. *Machine Learning*, vol. 45, no.1 pp. 5–32, 2001. DOI: https://doi.org/10.1023/A:1010933404324

[20] S. Kiliçarslan and E. Dönmez, "Improved multi-layer hybrid adaptive particle swarm optimization based artificial bee colony for optimizing feature selection and classification of microarray data," *Multimed. Tools Appl.*, 2023, DOI: https://doi.org/10.1007/s11042-023-17234-4

[21] N.Sahani, R. Zhu, J. H.Cho, and C. C Liu, "Machine Learning-based Intrusion Detection for Smart Grid Computing: A Survey," *ACM Transactions on Cyber-Physical Systems*, vol. 7, no. 2, pp. 1-31, 2023.