JAME
E-ISSN:2791-7401
Volume : 1, Issue 2, 2021
URL: https://journals.gen.tr/jame

Received: 30.11.2021
Acccepted: 13.12.2021

RESEARCH ARTICLE

Customer churn analysis in banking sector: Evidence from explainable machine learning models

Hasraddin Guliyev[1] **Ferda Yerdelen Tatoğlu**[2]

1 The Economic Research Center of Turkish World, Azerbaijan State Economic University, Azerbaijan.
Ph.D. Student at Department of Econometrics, Istanbul University, Istanbul, e-mail: hasradding@unec.edu.az

2 Istanbul University, Faculty of Economics, Department of Econometrics, Istanbul/Turkey, e-mail: yerdelen@istanbul.edu.tr

Abstract

Although large companies try to gain new customers, they also want to retain their old customers. Therefore, customer churn analysis is important for identifying old customers without loss and developing new products and making new strategic decisions for retaining customers. This study focuses on the customer churn analysis, that is a significant topic in banks customer relationship management. Identifying customer churn in banks will helps the management to classification who are likely to churn early and target customers using promotions, as well as provide insight into which factors should be considered when retaining customers. Although different models are used for customer churn analysis in the literature, this study focuses on especially explainable Machine Learning models and uses SHapely Additive exPlanations (SHAP) values to support the machine learning model evaluation and interpretability for customer churn analysis. The goal of the research is to estimate the explainable machine learning model using real data from banking and to evaluate many machine learning models using test data. According to the results, the XgBoost model outperformed other machine learning methods in classifying churn customers.

Keywords: Customer Loyalty, Customer Retention, Customer Churn Analysis, Machine Learning Models, Tree-Based Predictive Models

JEL Codes: C53, C55, M31

Citation: GULIYEV, H. & YERDELEN TATOGLU, F., (2021). Customer churn analysis in banking sector: Evidence from explainable machine learning models. *Journal of Applied Microeconometrics (JAME).* 1(2), 85-99, DOI: 10.53753/jame.1.2.03

Corresponding Author:
Hasraddin Guliyev
E-mail: hasradding@unec.edu.az

1. INTRODUCTION

Customer churn is a serious issue in an era of increasingly crowded markets and increased competition amongst businesses (Colgate et al. 1996). According to much research, the cost of acquiring new consumers is 1/5 the cost of keeping existing ones (Athanassopoulos 2000). For this reason, firms prefer to retain existing customers than add new ones and apply policies in this direction. One of the most valuable qualities in strategies designed to decrease or prevent customer churn is customer behavior data in the current customer base (Ganesh et al. 2000). As a result, as part of a consumer strategic plan targeted at decreasing customer churn, the discovery, and exploration for customers with a strong desire to leave the organization, or customer churn prediction, is essential (Blattberg et al. 2008).

Customer churn is a well-known issue in most sectors (Saradhi and Palshikar 2011), hence it's critical to develop a perfect predictive model designed in support of customer churn that could be used to formulate customer retention strategy. This topic is much more important in markets where competition is high and acquiring new customers is more difficult than retaining existing customers. Banking is one of the sectors where analyzing customer behavior and estimating customer churn based on these behaviors is an essential topic of research. Customer churn analysis results have a large impact on the bank's policy. Because the results of churn analysis allow banks to develop new customer strategies or improve existing ones. In addition, banks are critical to a country's financial growth and development, so the banking sector is an essential factor to the country's and people's financial stability. Because it is not always possible to get new customers in the competitive banking market, banks' primary goal is to ensure that existing customers are retention. Because banks, like all companies in the service sector, are customer-oriented, customer relationships with banks are a priority to their long-term business achievement. Studies conducted for the banking sector of various countries have revealed that, due to the competitive and dynamic nature of the banking sector, ensuring customer satisfaction is an important policy to prevent customer churn. Developing strong relationships with their customers have an advantage to high customer satisfaction and thus customer loyalty and turns into the most important factor for the stability, growth, and profitability of businesses.

In customer churn prediction in banking or other sectors, for example, a scoring approach supports the calculation of a potential churn probability per customer is based on their past data. The need to retain existing customers to maintain market share has led to an increased need for the development of various machine learning techniques for churn customer analysis. In customer churn prediction logistic regression (LR) is a very widespread paradigm to predict a churn probability since of great comprehensibility (Verbeke et al. 2012). The LR model, however, has a weak classification performance. On the other hand, despite the high prediction performance of the machine learning models explanation is difficult. In this study, we suggest an explainable machine learning model that combines the comprehensibility of logistic regression with the high classification performance of machine learning models.

2. LITERATURE REVIEW

With the advancement of techniques in the last 10-15 years, customer churn analysis studies using machine learning have grown in popularity. In the literature, it's clear that customer churn analysis has been used in a variety of industries (Kawale et al. 2009). While most studies have been done for the telecommunications and communication sector, customer churn analysis has also found applications in a wide range of fields, including e-commerce, banking, insurance, retail trade, energy, games and entertainment, and the medical (see Ahn et al. 2006; Bose and Chen 2009; Khan et al. 2010; Soeini and Rodpysh 2012; Buettgens et al. 2012; Long et al. 2012; Abbasimehr et al. 2013). Because this study is a customer churn analysis application in banking, it will be focused on this sector. Although the factors affecting customer satisfaction and loyalty differ for all countries, there are common points in the studies carried out.

According to Chakiso (2015), to obtain valuable customers, satisfy customers, and ensure customer loyalty, relational marketing (such as trust, bonding, communication, and reciprocity) must be used in the banking sector. Customers are encouraged through strong customer relationships to be always more satisfied and loyalty of bank, according to many studies. Conforming to Ozatac et al. (2016) the most significant determinants of customer satisfaction in banking are the accuracy of the information, the responsiveness of employees, access to all services, ability of employees, reliability, security of financial transactions, personalization, and consistency. As determined by Singh et al. (2013) lead to customer satisfaction are punctuality, effective communication, direct and acceptable information, efficient employee services in banking. Pasha and Waleed (2016) supervised a study in Pakistan to determine the factors that control customer loyalty. Customer satisfaction, brand loyalty, pricing policy, and service quality were found to be the most important determinants of customer loyalty in their study. The development of relational marketing dimensions such as quality service, personalized products, reliability, personalized communication, problem management, customer education, customer engagement, and the use of new technology, have been improved customer empowerment, according to Chatterjee and Kamesh (2020).

Besides relationship marketing, studies show that more objective reasons play a significant role in customer loyalty for banking. Many customers are served by banks through a variety of channels, including ATMs, mobile applications, and internet banking. Customers who have become more aware of service quality could be moving their financial services from one bank to another for a variety of reasons, including technological advancements, customer-friendly service, low-interest rates, geographic closeness, and a variety of services offered. When customers' options for service expand, a competitive market emerges. As a result, the competition among banks improves bank reliability and service quality significantly however it also increases the risk of churn to customers. In banking, as in many other sectors, developing a model which predicts customer is churn based on demographic, psychological, and transaction data is critical and in machine learning models are possible to predict who is churn customer and why. These predictive models have the advantage to lead to the design of personalized service and products and encourage customer loyalty with resulting in increased customer high satisfaction.

Mutanen (2006) offered a logistics regression-based customer churn study of the retail banking sector. Naveen et al. (2009) conducted detailed research with data mining techniques for churn customers that use credit cards. Bilal (2016) used gender, age, average monthly income, consumer status (retired, student, employed, unemployed), and whether the customer uses two or more bank products as control variables in the neural network model. According to Bilal customers that use multiple banking products are less probability of churn. Keramati et al. (2016) used the decision tree (DT) model to investigate churn customers in electronic banking (internet bank, telephone bank, mobile bank, ATM). They discovered the customers' dissatisfaction (duration of customer engagement, number of customer complaints), service usage (total number of uses and transactional amounts), and demographic variables (age, gender, employment status, education level) are effective on customer churn. Brânduşoiu et al. (2016) used a big dataset that includes 21 control variables for an advanced data mining model that predict prepaid customer churn. He et al. (2009) utilized a prediction model based on the Artificial Neural Network(ANN) algorithm for the complication of customer churn in a big Chinese telecom corporation with roughly 5.2 million consumers. The overall accuracy rate for prediction was 91.1 percent in the study. Nie et al. (2011) applied LR and DT models to predict churn customers using credit cards belonging to a Chinese bank. They discovered that the LR model outperformed from DT model to predict churn customers in a large dataset containing financial data from 135 variables for 60 million customers. Rajamohamed and Manokaran (2018) compared different classification models such as the k-nearest neighbor, Support Vectors Machine, Random Forest, Decision Tree, and Naive Bayes to predict customer churn in banking and discovered the Support Vectors Machine model was the most accurate, followed by the Random Forest model. Lopez-Diaz et al. (2017) compared 7 classification models for their predicting customer churn in a Spanish bank with 823,985 customers and observed that logistic regression was the greatest performance used for customer churn prediction.

In this study, in parallel with the literature, the effects of various factors such as age, income, gender, credit card status, and discount opportunities offered by banks on customer churn were examined with LR, DT, RF and Xgboost classification models.

3. METHODOLOGY

3.1. Logistic Regression Model (LR)

Logistic regression models discover the relationship among qualitative and other variables. In most models established with logistic regression, dependent variable has only two results. Usually, the emphasized event that is being realized is indicated by 1 and the one which is not realized by 0. The scientific society in the domains of economics, financial sector, and other social and environmental sciences gets now incorporated these models (Jabeur 2017; Zheng et al. 2020). The LR model is used to estimate the likelihood of an occurring event based on a set of predictors. The following is the predicted output of the logistic regression:

$$L_i = ln\left(\frac{P_i}{1-P_i}\right) = Z_i = \beta X_i + u_i \quad (1)$$

In the above expression, is a linear representation of the input variables and takes a value between -∞ and +∞, while takes a value between 0 and 1. LR has several statistical flaws. Multicollinearity and decreased performance accuracy are two of them.

3.2. Decision Tree

Ross Quinlan developed the C4.5 Decision Tree (DT) Classification Method as an expansion of the ID3 algorithm, which he previously created. These classifiers use the data samples to build a decision tree as a machine learning technique. The edge-based segmentation strategy is used to build decision tree models with an information gain metric used to select an appropriate input variable from among all the tree's input variables. The study selects a test drive through n outcomes that splits the data set *N*, as well as, training data set into subsets ($N_1, N_2, N_3,..,N_k$). *(Ci,P)* is the total number of samples in *P* that belong to C_i, and |P| is the total number of samples in *P*. The entropy of the set P is given by;

$$info(P) = -\sum_{i=0}^{k} \frac{freq(C_i,P)}{|P|} log_2\left(\frac{freq(C_i,P)}{|P|}\right) \quad (2)$$

The overall knowledge subject of *N* may be calculated after *N* is split with regard to the outcomes of a given characteristic, about *z*. N's information content may be calculated using Info *(N)*. The entire information content of *N* is equal to the weighted sum of each subset's entropies.

$$info_z(N) = \sum_{i=0}^{n} \frac{|N_i|}{|N|} info(N_i) \quad (3)$$

The gain is given by:

$$Gain(z) = info(N) \quad - \quad info_z(N) \quad (4)$$

It divides *N* about the test on *z* to give information. This is completed to choose the attribute *z* with the greatest knowledge gain. (1) Condition all the samples in a dataset fit to the same class, the decision tree generates a leaf node to select that class. (2) Otherwise, any input variable provides any information gain, a decision node leading the tree with the class's expected value is produced. (3) If an unknown instance's class is confronted, a decision node is constructed leading the tree together with the class's expected value.

3.3. Random Forest

Random Forest (RF) is a famous ML model used for data classification (Çağlayan et al. 2020). This algorithm is frequently utilized in sectors such as investing (Jabeur 2017), customer management and marketing (Salminen et al. 2019). A group of trees underpins the RF. It is complemented with an aggregate of the prediction's mean value, which is produced at the conclusion of each of the trees, reducing the lack of robustness of a single tree. Each of the trees is created using a subset of input variables that are picked at random. The following is an expression for the estimated model:

$$\hat{y} = \frac{1}{n}\sum_{i=1}^{n} g_k(x) \quad (5)$$

The vector of input features is *x*, and *g(x)* is a collection of the k[th] learner random trees. The RF final estimate is the mean of the outcome of the whole trees. As a result, with such weights, each individual tree has an impact on the RF estimation. Corresponding to Yeşilkanat (2020), the Random Forest model is superior to other machine learning methods. This is due to the former' stability in the direction of acquiring training data from subsets automatically and shaping trees using random techniques. Furthermore, because the Random Forest model achieves training by applying bootstrapping on a randomly chosen independent subset of datasets, the overfitting quantity is preserved.

3.4. eXtreme Gradient Boosting

The eXtreme Gradient Boosting (XgBoost) is the model that implements Chen and Guestrin's (2016) gradient boosting technique. It is a widely utilized flexible tool on the way to tree boosting algorithm achieves cutting-edge classification and effectiveness (Mai et al. 2020). The result is generated by the XgBoost, which is a collection of regression trees. The following equation is used to arrive at the final score:

$$\hat{y} = \sum_{h=1}^{H} g_h(x) \tag{6}$$

The number of trees in this equation is H, and the score for each tree's leaf is K. Multicollinearity has no effect on the XgBoost, which is an additional benefit. In order to maximize the model performance, XgBoost involves the selection of certain parameters. Parameter tuning is essential for the XGBoost to get around overfitting and too much confusion of the model. But, because the XgBoost utilizes multiple settings, this can be difficult. On the way to maximize the hyper-parameter values, we applied the grid search method with cross-validation.

3.5. The Performance Metrics of Classification Models

In order to determine which of the applied machine learning classification models are more successful both individually and among themselves, some performance metrics must be examined (Çağlayan 2020). These metrics are used to assess the effectiveness of the classification method in use and to compare classification models. Multiple metrics of models should be considered because evaluating these values as a single success criterion would be incorrect. All observation in the test data set is replaced in the model created with the training data set in the classification models, and classification prediction scores are achieved. The results of comparing the predicted values with the actual values are used to determine how well this model predicts, as well as its success and performance. The confusion matrix summarizes the results of the model's accuracy in making a prediction, as well as the conclusions of the performance evaluation of the machine learning classification model.

Figure 1. Confusion Matrix

Confusion Matrix		**Actual Values**	
		0	**1**
Predicted Values	**0**	True Positive TP	False Positive FP
	1	False Negative FN	True Negative TN

Figure 1 shows the confusion matrix is explained as follows for a two-category classification model:

True Positive (TP); indicates that observations with a true class value of 1 are correctly predicted as 1.

True Negative (TN); indicates the situation where observations with a true class value of 0 are correctly predicted as 0.

False Negative (FN); shows that observations with a true class value of 1 are incorrectly evaluated as 0 as a result of the prediction.

False Positive (FP); shows that observations with a true class value of 0 are incorrectly evaluated as 1 as a result of the prediction (Deng et al. 2020).

The accuracy rate (ACC) is calculated by taking the ratio of the number of classified observations () to the total number of samples (). This enables the evaluation of the estimation of the value of the estimation result made with the classification model as 1 when the true value of a class is 1, and the case that the estimated value of the class is 0 when the true value of the class is 0. ACC can be calculated using the following formula:

$$ACC = \frac{TP + TN}{FN + FP + TP + TN} \tag{7}$$

With a confusion matrix, we can also calculate sensitivity and specificity rates. The sensitivity is the ratio of correctly classified (TP) positive input values to the total true positive values (TP + FN):

$$Sensitivity = \frac{TP}{TP + FN} \tag{8}$$

The specificity is the ratio of the correctly classified (TN) to the total positive values (TN+FP) of the number of observations:

$$Specificity = \frac{TN}{TN + FP} \tag{9}$$

The sensitivity, specificity rate and the ACC metrics have values ranging from 0 to 1, and if they are close to 1, the model's performance is exceptionally good. Furthermore, sensitivity and specificity are inversely related, which means that as sensitivity grows, specificity decreases and conversely (Lambert and Lipkovich 2018).

The other metric AUC is a measurement of the entire area under the Receiver Operating Characteristic (ROC) curve, and it is one of the metrics used to evaluate model performance, together with the ROC curve (Jabeur et al 2021). The AUC value ranges from 0 to 1, with a value near 1 indicating a more accurate model. The distributions of TN and TP do not intersect when the area under the ROC curve is large, indicating that the classes have been successfully separated (Mai et al 2019).

3.6. Imbalanced Classification Problems

For customer churn analysis, studies have identified an imbalanced class distribution on customer data sets. Because the sample size of churn customers is substantially less than that of non-churn customers, the following scenario might occur; the accuracy of the classification is high, while churn customer prediction accuracy is low. So, the problem with unbalanced datasets is that typical classification learning techniques are typically biased towards the majority classes (referred to as "negative"), resulting in a greater misclassification rate in minority class occurrences (referred to as "positive" class) (Chawla 2009). The most common approach to this problem is to use a resampling technique to balance the class distribution of the training set before training a classification model. Random oversampling (ROS) and random undersampling (RUS) are two approaches for resampling (RUS). ROS, which consists of decreasing the data by deleting instances belonging to the majority class with the goal of equalizing the number of examples of each class; and RUS, which intends to reproduce or generate new positive examples in order to acquire importance (Batista et al. 2004). The main disadvantage of random undersampling is that it might lose potentially relevant data that could be significant in the induction process. The elimination of data is an important decision to make, hence many undersampling proposals include heuristics to overcome the limits of non-heuristic decisions. Random oversampling, on the other hand, may increase the likelihood of overfitting since it duplicates the minority class instances exactly. In this manner, a symbolic classifier, for example, may generate rules that appear to be accurate but only cover one reproduced case.

Ensemble learning (tree-based) models are another option for improving the performance of a single classifier by training multiple separate classifiers and integrating their outputs to produce the final choice (Kuncheva 2004). Cost-sensitive ensembles, on the other hand, use the ensemble learning algorithm to lead cost reduction rather than altering the underlying classifier in order to accept costs in the learning process. Ensemble learning models include Random Forest, AdaBoost, and XgBoost. Ensemble Learning models are well-known in data mining and machine learning for their good performance in a wide range of applications, and it may be the better alternative for the class imbalance problem (Wozniak 2014). For example, Ahmad et. al (2019) discovered that tree-based models performed better from undersampling for unbalanced classification in their customer churn analysis in the telecommunications sector.

4. DATA AND ANALYSIS

4.1. Data and Variables

This paper aims on application of machine learning models for predicting the churn customers. The research is based on real data from a bank. Before customer churn analysis, we need to determine churn status of customers. Customers who close individual loans and do not apply for new loans despite 9 months after the close date of the loan are included in the churn customer category and take 1 value for the dependent variable of classification model, however, it takes 0 value when customer applying to loan during the 9 months from the close date of loan. According to the calculations, 91% of the customers generally applied for the second loan within 9 months, so we used the 9-month criterion to determine customer churn status.

The database of a bank was used in the data collecting process, and 274,542 observations were analyzed once all the pre-elimination processes are completed. The most appropriate input variables which according to the local market conditions are selected for predicting the customer's churn status and summarized in Table 1.

Table 1. Definition of Variables

Variable	Assigned Short Name	Definition
churn status	churn_status	shows customer's churn category. If the customer is churn, it takes a value of 1, if is non-churn, a value of 0.
Input Variables		
customer's age	age	Shows customer's age
average income	salary	shows the average income of the customer for the last 12 months.
gender	gender	if the customer is male, it takes a value of 1, if is female, a value of 0.
loan amount	amount	amount of customer's last used credit
interest rate	interest_rate	nominal interest rate calculated to the customer's last used loan
credit term	duration	duration of customer's last used credit (months)
credit closing and early payme nt (days)	closed	It gets negative values daily if the customer has paid the loan before the expiry date; positive values if the loan has been delayed.
interest rate discount	rate_discount	if the customer is offered an interest rate discount in last loan compared to previous loan, it gets a value of 1 and in other case a value of 0.
amount increase	amountup	if the customer is offered an increase in the loan amount in last loan compared to previous loan, it gets a value of 1 and in other case a value of 0.
competition region	competition	if customers live in competition region where more branch of other banks exist 1, if not, a value of 0.
credit card status	creditcard	if the customer has a credit card, it takes a value of 1, if not, a value of 0.
salary card	card_status	if the customer's salary card belongs to bank, it takes the value 0 and if it belongs to another bank, it takes the value of 1.
credit count	creditcount	it shows how many individual loans the customer has drawn to determine the relationship with the bank. since it is a categorical variable, customers with 1 individual loan were taken as a base and 3 dummy variables were

4.2. Describing of Variables and Correlation Analysis

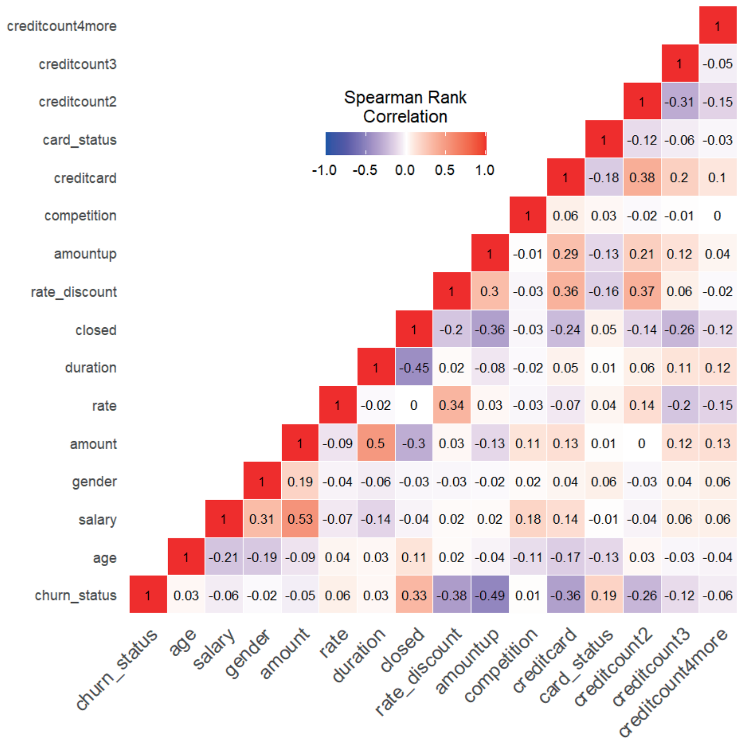
Table 2 summarizes the descriptive statistics for all variables examined in this study. When looking at the proportion of outcome variable that is churn status, we see that 88.5 percent of customers are no churner and 11.6 percent of customers are churner. The average age of the customers that input variable is 49.5, with a standard deviation of 12.8, and a range of 19 to 70. The mean annual salary is 309, with a low of 50 and a high of 11850. When we look at the categorical input variable like gender, we notice that males account for 52.3 percent of the customer while females account for 47.7%. In other input variables, it can be interpreted the same way.

Table 2. Descriptive Statistics of Variables

Variables	Statistics / Frequency	Variables	Statistics / Frequency	Variables	Statistics / Frequency
churn status	No-churn (88.5%) Churn (11.5%)	interest rate	Mean (sd) : 28.4 (2.2) min ≤ med ≤ max: 14 ≤ 28.2 ≤ 39	competition	No (73.0%) Yes (27.0%)
age	Mean (sd): 49.5 (12.8) min ≤ med ≤ max: 19 ≤ 52 ≤ 70	duration	Mean (sd): 29.2 (10.5) min ≤ med ≤ max: 3 ≤ 36 ≤ 156	credit card	No (31.5%) Yes (68.5%)
salary	Mean (sd): 309 (250.9) min ≤ med ≤ max: 50 ≤ 199.6 ≤ 11850	closed	Mean (sd): -370.6 (378.9) min ≤ med ≤ max: -1820 ≤ -264 ≤ 120	card status	No (92.7%) Yes (7.3%)
gender	Male (52.3%) Female (47.7%)	rate discount	No (39.7%) Yes (60.3%)	credit count	1 (41.2%) 2 (46.0%) 3 (10.3%) 4+ (2.5%)
credit amount	Mean (sd): 2558.4 (2132.8) min ≤ med ≤ max: 300 ≤ 2000 ≤ 20000	amount up	No (31.1%) Yes (68.9%)		

We have calculated the correlation coefficients before using machine-learning techniques to ensure that the input variable selection is accurate. Figure 2 shows the pairwise Spearman`s rank correlation among the variables in our study:

Figure 2. Spearman's Correlation Heatmap



Source: Authors 'own calculation from the dataset.

Spearman's correlation evaluates monotonic relationships, and it is a reliable tool for big datasets with outliers. Calculated Spearman's correlation coefficients indicate customer churn status has a moderate negative correlation with rate discount (ρ=-0.38) and the amount up (ρ=-0.49) in the next credit approval. Furthermore, customer churn has a moderate negative correlation with credit card status (ρ=-0.38) and credit counts variables (creditcount2, creditcount3 and creditcount4+), however a positive correlation with credit closed (ρ=0.33) and card status (ρ=0.19) that indicating the customer's salary card belongs to another bank. Many variables' correlation coefficients were calculated at a low level. As a result, using nonlinear machine learning models to better explain the relationship with churn status may be advantageous because of low-level correlation coefficients.

4.3. Machine Learning Models' Overall Performance and Variable Importance

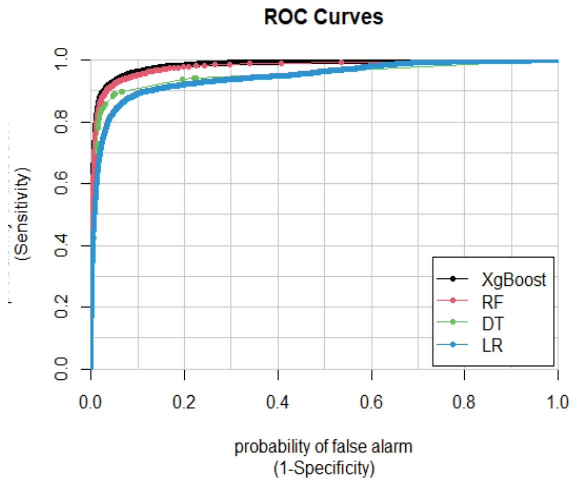
The performance of the indicated machine-learning models is compared in this section. The data from 274,542 customers was split into two parts; 80% of the dataset (219,634 customers) was used to train machine learning models, while 20% of the dataset (54,908 customers) was used to test them. The effectiveness of each individual model was determined using some performance metrics. We have explored the results of the proposed machine learning models and evaluated these results through the model performance metrics (i.e., sensitivity, specificity, accuracy, and Receiver Operating Characteristic- ROC curve).

Table 3. The Performance of Machine Learning Models in Testing Data

Performance Metrics	Machine Learning Models			
	LR	DT	RF	Xgboost
Sensitivity	0.9896	0.9819	0.9843	0.9854
Specificity	0.5693	0.8206	0.8387	0.8504
ACC	0.9410	0.9632	0.9675	0.9697
AUC	0.9464	0.9510	0.9797	0.9850

Table 3 expresses the results computed with the advanced machine-learning models such as Logistic regression (LR), Decision Tree (DT), Random Forest (RF), and XgBoost. We used the *caret* package in R for model estimates in the training dataset and parameter turning to deal with the overfitting problem and model performance boosting. To calculate hyperparameters, we used 5-cross validation and the grid search method. First, we used the Logistic regression model; while its sensitivity (0.9896) is high, its specificity (0.5693) is very low, so we cannot conclude that the LR model is very accuracy model in classification. For the Decision Tree model, the complexity parameter is estimated to be 0.0012. Although the DT model's sensitivity (0.9819), specificity (0.8296), ACC (0.9410) and AUC (0.9510) are acceptable, it cannot be considered the best model. The maximum depth parameter, which is the number of variables randomly sampled as candidates at each split, is calculated as 4 for the Random Forest model. Although the performance of the RF model is very good, it still cannot be said to be the best model. We used the XgBoost model and determined that the optimal number of tree sizes was 130. Because of the values of sensitivity (0.9854), specificity (0.8504), accuracy (0.9697), and AUC (0.9850), as well as the closeness of these metrics to 1, the XgBoost model had a higher predictive performance in the test data set when all models were compared. To emphasize, we would like to point out that a more accurate estimation of the churned customer (positive class) is more important for customer churn analysis, so having a higher specificity rate is more advantageous for us. Consequently, the area under of Receiver Operating Characteristic (ROC) curve plotted for XgBoost is higher compared to other machine learning models.

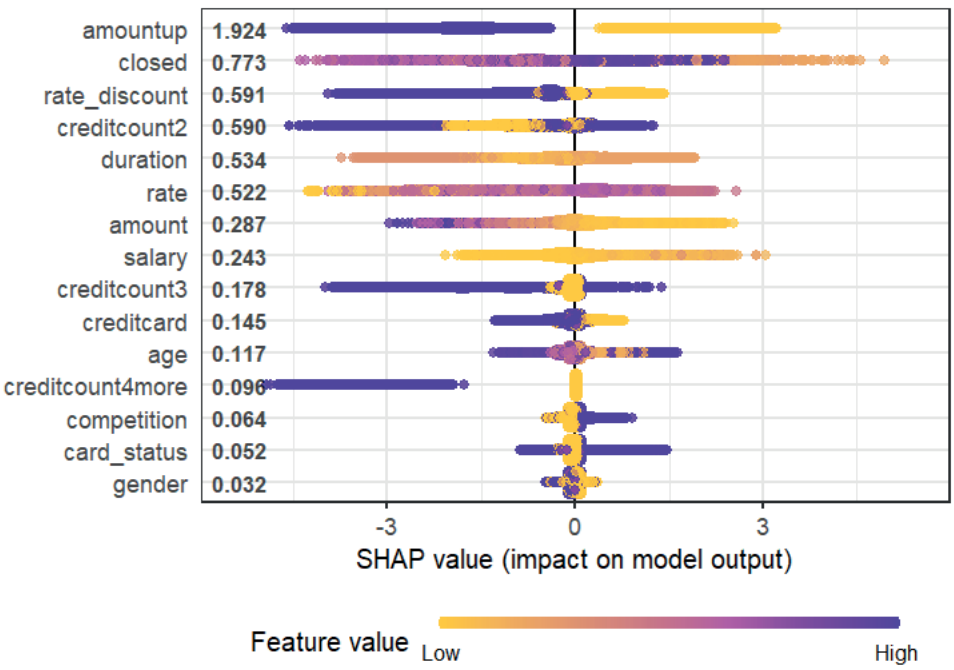
Figure 3. Compare of Receiver Operating Characteristic (ROC) Curves



Source: Authors 'own calculation from the dataset.

It is useful to know the proportional contributions of all factors on the final forecast outcome when predicting the churner. Lundberg et al. (2018) recently suggested the SHAP to assess the significance of specific characteristics. This can benefit in balancing the accuracy and interpretability of black-box machine-learning models. The importance of variables is shown in Figure 4.

Figure 4. The Shapley values of XgBoost Model.



Source: Authors 'own calculation from XgBoost Model.

Figure 3 shows the importance of the variables with the effects of the variables. A Shapley value for a feature and an instance is represented by each point on the summary plot. The feature is demonstrated on the y-axis position, while the Shapley value is demonstrated on the x-axis position. The value of the feature is represented by the color, which ranges from low to high. To get a meaning of the Shapley value per input variable, the overlapping points are jittered in the y-axis direction. The features are ranked from the most important to the least important one. The 5 most important variables to explain customer churn status are the customer is offered a higher loan amount compared to the previous loan (amount up), early payment of credit or delaying credit (closed), the customer is offered a lower interest rate compared to previously loan (rate discount), the number of credits (credit count 2) and the duration of customer's last used credit (duration). Shapley values show that when an interest rate discount (rate discount) and more loan amount (amount up) is determined is applied to a new loan, the probability of churn decreases. In addition, if the number of credits (credit number 2, credit number 3, credit number 4 +) increases, the probability of churn decreases. Interestingly, a decrease in the duration of the previous loan and having a credit card (card status) reduces the probability of customer churn. However, the probability of churn increases as the customer's age rises and another bank provides a monthly salary to customers with their own bank card (this is due to the change in the customer's workplace). Furthermore, the interest rate of credit and the customer to live in the competitive region increases the probability of churn.

5. CONCLUSION

In this paper, we proposed ML methods for predicting customer churn. The machine learning predictive models need to achieve high AUC values. Firstly, to test and train the model, the sample dataset is divided into 80% for training and 20% for testing. For validation and hyperparameter tuning, we selected to use 5-fold cross-validation. In addition, we contended with another problem: the data was not balanced and only about 11.5% of the dataset had included churn customers. To solve this problem, undersampling methods or tree base algorithms are suggested, so we applied the tree-based models such as Decision Tree, Random Forest, and XgBoost. The XgBoost outperformed for each metric, with an AUC of 96.97 percent and the Random Forest model comes in second.

Shapley values present the most important variable that explains the churn status of the customers and it indicated positive or negative effect of input variables for XgBoost model. In general, the important reasons that increase the risk of churn of the customer are the fact that the salary card belongs in another bank in the next period and that the customer inhabits in the competitive region where there are other alternatives to take credit. The most important reason that increases the probability of churn is that the applied credit interest is high, which in turn reduces the customer's ability to repay the loan and shifts the customer towards the other bank that offers more favorable interest rates for subsequent loans. The customer's relationship with the bank is a critical component in minimizing the risk of customer churn. If the customer has a longer-term relationship with the bank, then the customer will benefit from the advantages offered by the bank's loyalty program and will maintain the relationship with the bank for a long time. In addition, applying the interest rate discount and upping the amount of credit to the customer decreases the risk of churner.

Our results suggest that building a model that can accurately anticipate customer retention might have some management and financial consequences for banking in order to reduce the probability of churn. Firstly, correctly classifying a customer as a churner or non-churner helps decrease the expenses associated with misclassification. Second, our findings show that academics and practitioners do not have to rely exclusively on conventional methodology as logistic regression for predicting customer churn. Finally, our findings suggest management recommendations for improving the decision-making process in the context that customer churn prediction. Banks and financial institutions may use XgBoost models to correctly identify clients who are at risk of churn, focus their efforts on them, and potentially get profit. Companies should more focus on customer retention policies rather than concentrating on new target markets, which are generally difficult to gain. So, the findings of the machine-learning techniques of this research could have a variety of policy implications for customer relationship management and the marketing strategy of the company. In the future, more explainable machine learning methods should be used, and models with higher performance should be suggested for predicting customer churn.

Funding

The authors declare that this study has no financial support.

Conflict of Interest

The authors declare that they have no conflicts of interest.

REFERENCES

- ABBASIMEHR, H., SETAK, M., & SOROOR, J. (2013). A framework for identification of high-value customers by including social network-based variables for churn prediction using neuro-fuzzy techniques. *International Journal of Production Research*, *51*(4), 1279-1294.
- AHMAD, A. K., JAFAR, A., & ALJOUMAA, K. (2019). Customer churn prediction in telecom using machine learning in big data platform. *Journal of Big Data*, *6*(1), 1-24.
- AHN, J. H., HAN, S. P., & LEE, Y.S. (2006). Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry. *Telecommunications policy*, *30*(10-11), 552-568.
- AKAY, E.C., SOYDAN, N.T.Y. & GACAR, B.K. (2020). MAKİNE ÖĞRENMESİ VE EKONOMİ: BİBLİYOMETRİK ANALİZ. PressAcademia Procedia, 12 (1), 104-105.
- ATHANASSOPOULOS, A.D. (2000). Customer satisfaction cues to support market segmentation and explain switching behavior. *Journal of business research*, *47*(3), 191-207.
- BATISTA, G.E., PRATI, R.C., & MONARD, M.C. (2004). A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD explorations newsletter*, *6*(1), 20-29.
- BILAL ZORIĆ, A. (2016). Predicting customer churn in banking industry using neural networks. *Interdisciplinary Description of Complex Systems: INDECS*, *14*(2), 116-124.
- BLATTBERG, R. C., KIM, B.D., & NESLIN, S.A. (2008). Churn Management. In *Database Marketing* (pp. 607-633). Springer, New York, NY.
- BOSE, I., & CHEN, X. (2009). Quantitative models for direct marketing: A review from systems perspective. *European Journal of Operational Research*, *195*(1), 1-16.
- BRÂNDUŞOIU, I., TODEREAN, G., & BELEIU, H. (2016). Methods for churn prediction in the pre-paid mobile telecommunications industry. In *2016 International conference on communications (COMM)* (pp. 97-100). IEEE.
- BUETTGENS, M., NICHOLS, A., & DORN, S. (2012). Churning under the ACA and state policy options for mitigation. *Prepared for Robert Wood Johnson Foundation, Timely Analysis of Immediate Health Policy Issues, http://www.urban. org/UploadedPDF/412587-Churning-Under-the-ACA-and-State-Policy-Options-for-Mitigation. pdf.*
- ÇAĞLAYAN AKAY, E. (2018). Ekonometride Yeni Bir Ufuk: Büyük Veri ve Makine Öğrenmesi. Social Sciences Research Journal, 7(2): 41-53.
- ÇAĞLAYAN AKAY, E. (2020). Ekonometride Büyük Veri ve Makine Öğrenmesi: Temel Kavramlar, Der Yayınları, İstanbul.
- CHAKISO, C.B. (2015). The effect of relationship marketing on customers' loyalty (Evidence from Zemen Bank). *EMAJ: Emerging Markets Journal*, *5*(2), 58-70.
- 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.
- CHATTERJEE, D., SEKHAR, S.C., & BABU, M.K. (2021). Customer Empowerment-A Way to Administer Customer Satisfaction in Indian Banking Sector. *NVEO-NATURAL VOLATILES & ESSENTIAL OILS Journal| NVEO*, 1621-1629.
- CHAWLA, N.V. (2009). Data mining for imbalanced datasets: An overview. *Data mining and knowledge discovery handbook*, 875-886.

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

- COLGATE, M., STEWART, K., & KINSELLA, R. (1996). Customer defection: a study of the student market in Ireland. *International journal of bank marketing*.

- DENG, X., LIU, Q., DENG, Y., & MAHADEVAN, S. (2016). An improved method to construct basic probability assignment based on the confusion matrix for classification problem. *Information Sciences*, *340*, 250-261.

- GANESH, J., ARNOLD, M.J., & REYNOLDS, K.E. (2000). Understanding the customer base of service providers: an examination of the differences between switchers and stayers. *Journal of marketing*, *64*(3), 65-87.

- HE, Y., HE, Z., & ZHANG, D. (2009). A study on prediction of customer churn in fixed communication network based on data mining. In *2009 sixth international conference on fuzzy systems and knowledge discovery* (Vol. 1, pp. 92-94). IEEE.

- JABEUR, S.B. (2017). Bankruptcy prediction using partial least squares logistic regression. *Journal of Retailing and Consumer Services*, *36*, 197-202.

- JABEUR, S. B., GHARIB, C., MEFTEH-WALI, S., & ARFI, W.B. (2021). CatBoost model and artificial intelligence techniques for corporate failure prediction. *Technological Forecasting and Social Change*, *166*, 120658.

- KAWALE, J., PAL, A., & SRIVASTAVA, J. (2009). Churn prediction in MMORPGs: A social influence based approach. In 2009 international conference on computational science and engineering (Vol. 4, pp. 423-428). IEEE.

- KERAMATI, A., GHANEEI, H., & MIRMOHAMMADI, S. M. (2016). Developing a prediction model for customer churn from electronic banking services using data mining. *Financial Innovation*, *2*(1), 1-13.

- KHAN, A.A., JAMWAL, S., & SEPEHRI, M.M. (2010). Applying data mining to customer churn prediction in an internet service provider. *International Journal of Computer Applications*, *9*(7), 8-14.

- KUNCHEVA, L. (2004). Combining pattern classifiers methods and algorithms. john wiley&sons. *Inc. Publication, Hoboken*.

- LAMBERT, J., & LIPKOVICH, I. (2008). A macro for getting more out of your ROC curve. In *SAS Global forum, paper* (Vol. 231).

- LONG, X., YIN, W., AN, L., NI, H., HUANG, L., LUO, Q., & CHEN, Y. (2012, March). Churn analysis of online social network users using data mining techniques. In *Proceedings of the international MultiConference of Engineers and Conputer Scientists* (Vol. 1).

- LÓPEZ-DÍAZ, M. C., LÓPEZ-DÍAZ, M., & MARTÍNEZ-FERNÁNDEZ, S. (2017). A stochastic comparison of customer classifiers with an application to customer attrition in commercial banking. *Scandinavian Actuarial Journal*, *2017*(7), 606-627.

- LUNDBERG, S. M., ERION, G. G., & LEE, S. I. (2018). Consistent individualized feature attribution for tree ensembles. *arXiv preprint arXiv:1802.03888*.

- MAI, F., TIAN, S., LEE, C., & MA, L. (2019). Deep learning models for bankruptcy prediction using textual disclosures. *European journal of operational research*, *274*(2), 743-758.

- MUTANEN, T. (2006). Customer churn analysis–a case study. *Journal of Product and Brand Management*, *14*(1), 4-13.

- Naveen, N., Ravi, V., & Kumar, D. A. (2009). Application of fuzzyARTMAP for churn prediction in bank credit cards. *International Journal of Information and Decision Sciences*, *1*(4), 428-444.

- NIE, G., ROWE, W., ZHANG, L., TIAN, Y., & SHI, Y. (2011). Credit card churn forecasting by logistic regression and decision tree. *Expert Systems with Applications*, *38*(12), 15273-15285.

- 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.
- RAJAMOHAMED, R., & MANOKARAN, J. (2018). Improved credit card churn prediction based on rough clustering and supervised learning techniques. *Cluster Computing*, *21*(1), 65-77.
- RISSELADA, H., VERHOEF, P.C., & BIJMOLT, T.H. (2010). Staying power of churn prediction models. *Journal of Interactive Marketing*, *24*(3), 198-208.
- SALMINEN, J., YOGANATHAN, V., CORPORAN, J., JANSEN, B.J., & JUNG, S.G. (2019). Machine learning approach to auto-tagging online content for content marketing efficiency: A comparative analysis between methods and content type. *Journal of Business Research*, *101*, 203-217.
- SARADHI, V.V., & PALSHIKAR, G.K. (2011). Employee churn prediction. *Expert Systems with Applications*, *38*(3), 1999-2006.
- SINGH, S., ANUSHA, B., & RAGHUVARDHAN, M. (2013). Impact of Banking Services on Customer Empowerment, Overall Performance and Customer Satisfaction: Empirical Evidence. *Journal of Business and Management (IOSR-JBM)*, *16*(1), 17-24.
- SOEINI, R. A., & RODPYSH, K. V. (2012). Applying data mining to insurance customer churn management. *International Proceedings of Computer Science and Information Technology*, *30*, 82-92.
- VERBEKE, W., DEJAEGER, K., MARTENS, D., HUR, J., & BAESENS, B. (2012). New insights into churn prediction in the telecommunication sector: A profit driven data mining approach. *European journal of operational research*, *218*(1), 211-229.
- WALEED, A., PASHA, A., & AKHTAR, A. (2016). Exploring the impact of liquidity on profitability: Evidence from banking sector of Pakistan. *Journal of Internet Banking and Commerce*, *21*(3).
- WOŹNIAK, M., GRANA, M., & CORCHADO, E. (2014). A survey of multiple classifier systems as hybrid systems. *Information Fusion*, *16*, 3-17.
- YEŞILKANAT, C. M. (2020). Spatio-temporal estimation of the daily cases of COVID-19 in worldwide using random forest machine learning algorithm. *Chaos, Solitons & Fractals*, *140*, 110210.
- ZHENG, K., ZHANG, Z., & SONG, B. (2020). E-commerce logistics distribution mode in big-data context: A case analysis of JD. COM. *Industrial Marketing Management*, *86*, 154-162.
- ZORIĆ, A.B. (2016). Predicting customer churn in banking industry using neural networks. *Interdisciplinary Description of Complex Systems: INDECS*, *14*(2), 116-124.