



Propension to customer churn in a financial institution: a machine learning approach

Renato Alexandre de Lima Lemos¹ · Thiago Christiano Silva¹ · Benjamin Miranda Tabak² 

Received: 7 December 2021 / Accepted: 5 February 2022 / Published online: 6 March 2022
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2022

Abstract

This paper examines churn prediction of customers in the banking sector using a unique customer-level dataset from a large Brazilian bank. Our main contribution is in exploring this rich dataset, which contains prior client behavior traits that enable us to document new insights into the main determinants predicting future client churn. We conduct a horseshoe of many supervised machine learning algorithms under the same cross-validation and evaluation setup, enabling a fair comparison across algorithms. We find that the random forests technique outperforms decision trees, k -nearest neighbors, elastic net, logistic regression, and support vector machines models in several metrics. Our investigation reveals that customers with a stronger relationship with the institution, who have more products and services, who borrow more from the bank, are less likely to close their checking accounts. Using a back-of-the-envelope estimation, we find that our model has the potential to forecast potential losses of up to 10% of the operating result reported by the largest Brazilian banks in 2019, suggesting the model has a significant economic impact. Our results corroborate the importance of investing in cross-selling and upselling strategies focused on their current customers. These strategies can have positive side effects on customer retention.

Keywords Churn · Churn prediction · Machine learning · Random forests · Financial services

1 Introduction

Bank executives worldwide have already recognized the importance of increasing customer satisfaction. It is a fact that as customers adopt new technologies in other areas of their lives, their expectations and levels of demand for banking services increase as well. According to the World

Retail Banking Report 2019 [14], 66.8% of current banking customers have already used or intend to use a bank account from a non-traditional company (big tech or fintech) in the next three years. According to [43], 55% of bank executives see these non-traditional competitors in the financial sector as a threat to traditional banks. As a result of this differentiated competition scenario, retaining today's customer base becomes increasingly difficult for traditional banks.

Customer churn, a move in which a particular customer abandons his current company to join a competing company's services, has become increasingly common [58]. Numerous studies show that preventing customer churn saves money, as acquiring new customers can cost up to five times as much as satisfying and retaining existing customers [49, 61]. As a result, it is becoming increasingly critical for businesses to invest in managing their client relationships in order to avoid churn. Thus, the need to preserve their revenues has prompted companies to understand and analyze their clients' behavior to identify clients who are more prone to churn in advance. In this

All authors contributed equally to this work.

✉ Benjamin Miranda Tabak
benjamin.tabak@fgv.br

Renato Alexandre de Lima Lemos
renato.l.lemos@caixa.gov.br

Thiago Christiano Silva
thiago.csilva@p.uceb.br

¹ Universidade Católica de Brasília, QS 07 – Lote 01, EPCT - Taguatinga, Brasília - DF, Brasília, Distrito Federal 71966-700, Brazil

² School of Public Policy and Government, Fundação Getúlio Vargas, SGAN 602, Asa Norte, Brasília, Distrito Federal 70830-020, Brazil

way, businesses can act proactively to retain customers and increase profits.

Detecting churn specifically in the banking sector has additional challenges. First, large banks typically have tens of millions of customers in their portfolio. Strategies that attempt to reduce churn that involve human interventions do not scale up well. Second, they are incapable of adapting quickly enough to changes in customer needs. Third, even though banks segment clients across local managers, it is still difficult to detect customer patterns manually, especially if they manage a large number of customers. These features create the need of automated methods that are able to detect the non-trivial patterns of customer behavior that may suggest potential churn in these massive data sets in advance. These characteristics motivate the use of machine learning techniques, which provide supervised learning methods that have proved to learn non-trivial patterns in the data (without human intervention) and generalize well to previously unseen data.

This paper investigates the behavior of a representative dataset of 500,000 clients of a Brazilian financial institution, aiming to generate a churn predictive model of account holders through machine learning, capable of identifying the variables with a more significant predictive potential of a client's propensity to churn. We aim to develop a model that identifies clients who will most likely churn with the necessary advance. In this way, organizations would have sufficient time to run retention actions to hold these clients.

This paper contributes to the empirical literature on customer churn prediction in several ways. First, it comprehensively analyzes numerous well-known supervised learning classification algorithms via a horserace. In contrast, the empirical literature typically uses specific algorithms to deal with the problem, such as decision trees [8, 37], k -nearest neighbors [18, 63], elastic net [42], logistic regression [34, 38], Support Vector Machines (SVM) [19, 64], and random forests [39, 57]. We instead opt to test all these algorithms in a common empirical setup, allowing for a fair comparison of the classifiers.¹ Second, our dataset is unique and representative of a large Brazilian bank. Most empirical studies either use artificial datasets or limited data from a specific bank, compromising the empirical conclusions. This empirical constraint often comes from customer-level bank data being private and legally protected. Third, we leverage the availability of a large number of attributes in the dataset not only to obtain accurate predictions of customer churn but also to

understand which attributes have the highest predictive power when determining the likelihood of a potential churn. This analysis can provide insightful information on customer behavior that may be used to develop policies to mitigate customer churn.

Brazil is now the world's eighth largest economy, and its banking system, while still concentrated, is regarded as one of the most solid in the world. Despite this concentration, the Banking Report from the Central Bank of Brazil indicates that the fintech ecosystem is thriving, with a high rate of growth and a high volume of new constitution claims received and under analysis [5]. Additionally, recent referrals from regulators demonstrate a willingness to encourage competition further so that customers have an increased degree of freedom and comfort in selecting the institution that will provide the best financial services. Other examples of this direction include Open Banking and the emergence of rules that facilitate the portability of applications, salary credit, and loans between financial institutions. This context of competition promotion highlights the importance of promoting studies on customer churn, particularly among large traditional Brazilian banks, which have financial intermediation as their primary source of revenues. This makes these financial institutions vulnerable to customer loss and emphasizes the importance of understanding the main drivers of customer churn.

Preventing churn has become one of the most vital objectives for organizations given the increased competition for customers and the difficulty of replacing the loss of revenue caused by the exit of profitable customers [27, 28]. However, as already discussed, retaining existing customers is now one of the biggest challenges of financial institutions in a saturated and competitive market where customers are increasingly able to move to other service providers [33]. In this context, the development of precise and high-performance statistical models allows the identification in advance of customers that tending to churn becomes an essential condition for preserving these companies' competitiveness.

Besides, the advances in technology, globalization, and the emergence of fintech have raised competition in the market to sell financial products and services to levels never seen before. Likewise, the proliferation of mobile technologies and social networks, as well as the resulting expansion of access to information by consumers, has shortened distances, aided in the improvement in customers' financial literacy, expanded their ability to communicate with other actors located anywhere in the world, and consequently made them more receptive to changes. By becoming more active participants in their relationships with companies providing products and services situated anywhere globally, these same consumers have increased their expectations for the quality of products and services

¹ Model's stability is an important characteristic in empirical works. We refer the reader to [44, 45] for theoretical research in the context of neural networks. An application of a related mathematical modeling for COVID-19 spreading is studied in [1].

they consume and decreased their loyalty to companies with which they currently transact [50]. Therefore, it is a scenario that imposes a risk to the long-standing dominance imposed of the leading companies in the financial sector, the large traditional banks, particularly those that fail to adapt to this revolution [22]. This change in clients' behavior, who are less passive economic agents, combined with the increased likelihood of losing customers to rival companies, produces a genuine “war” in the dispute for clients between financial institutions.

2 Literature review

We begin by examining the scientific community's interest in customer churn. We conducted bibliographical research on the Scopus dataset on May 30, 2020, using the logical expressions (“machine learning OR “data mining” OR “knowledge discovery”) AND “bank*” AND (“churn*” OR “evasion” OR “dropout”) AND (“customer” OR “client”) applied over titles, abstracts or keywords. We recovered 491 references between articles and reviews published in journals and conferences from 2003 to 2019. Figure 1 displays the evolution of the number of publications over time. There is a growing interest in the subject, which is likely influenced by the increased availability of data as a result of companies' increased investments in big data solutions, as well as by the awareness and concern among these same companies about avoiding customer loss as a result of increased competition due to the entry of fintechs and big techs.

Churn is the abandonment of the company by a given client. This action usually is accompanied by a customer migration to a competitor. [58] conclude that customer

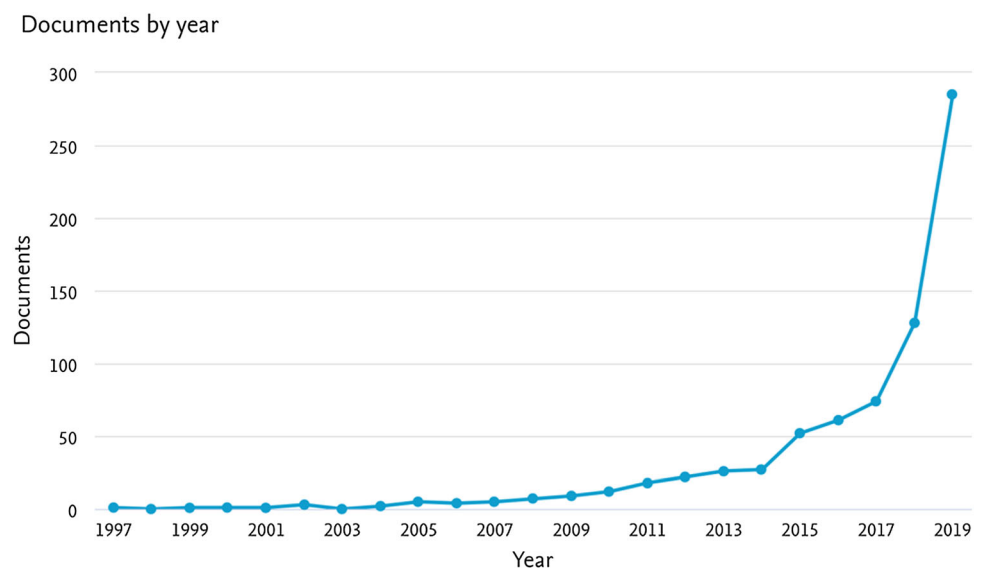
churn can occur actively or passively, according to the factor that motivated the movement (voluntary or involuntary). When developing a churn predictive model, the goal is to guide actions that could reverse active churn, i.e., the one in which the client voluntarily took the initiative to leave the relationship with the organization.

[7] and [19] explore specific aspects of organizations that favor churn. These include a lack of differentiation in the face of competition, competitors offering cutting-edge technology, employees lacking empathy, uncompetitive interest rates, low quality, and a lack of variety in services. [22] note that technological advancement, globalization, and consequently, competition facilitated competing companies to exploit these vulnerabilities to attract dissatisfied customers from other organizations. Additionally, the literature shows that replacing churned customers with new customers is economically disadvantageous [36, 47]. As a result, we believe banks should develop strategies for retaining existing customers in addition to the traditional strategy of acquiring new customers. Therefore, developing accurate churn predictive models and effective strategies is essential to prevent losing customers.

[35] summarize in five main points the importance of actions involving the reversal of churn and the retention of clients:

- (1) Customer retention reduces the need to prospect for new customers, allowing organizations to focus on strengthening relationships with existing customers;
- (2) Older customers, who are more familiar with the company, tend to purchase more and, when satisfied, can practice referral marketing;
- (3) Serving and maintaining long-term customers is less expensive due to the increased knowledge acquired during their consumption life cycle;

Fig. 1 Number of publications in journals and conferences from 2003 to 2019 with the co-occurrence of the terms machine learning and churn either in the title, abstract or keyword list in the scopus dataset



- (4) Long-term customers are typically less receptive to competitive marketing efforts; and
- (5) Customer loss is a cost of opportunity because it reduces sales and necessitates the acquisition of new customers to offset losses.

It is essential to analyze the temporal dynamics of the decision to interrupt the relationship. The customer often decides to interrupt months before deciding to do so effectively. Such dynamics are explored in several academic works [4, 6]. According to [4], this “slow” dynamic allows banks to use less strict time constraints than other business contexts. One example is mobile telephony, where customers generally switch from one operator to another in a short period, making forecasting a challenging task. In general, small time windows reduce the institution’s reaction time to reverse the churn. On the other hand, although significant time windows increase the time allowed for the bank’s reaction, they can also easily lead to inconsistent results due to a possible change in the environment over this period. [6] conclude that relevant changes in the economy, disruptions in business models, or even a political or financial crisis can influence customers’ prone to leave the bank. All of this suggests that it is necessary to find an optimal balance between the accuracy of the predictions and the allowed reaction time. For this reason, it is essential to define how long in advance we want, and we can know if a customer tends to churn. This answer depends on the bank’s needs and is also a significant challenge.

Another critical feature in churn studies is that records related to customers leaving the company, in general, are much lower than those related to customers who remain [3, 4, 12]. This effect can make prediction difficult because the available sample may not guarantee sufficient positive churn records for the analysis and lead to biased results. Empirically, unbalanced classes are one of the main problems with which machine learning methods must deal. There are several methods to mitigate unbalanced class problems. There is some emphasis on two of them, pre-processing and adaptation of learning algorithms, making them cost-sensitive. These procedures allow the assignment of distinct weights for hit/error during training involving the minority and majority classes. The pre-processing approaches include sample treatment methods that aim to balance the training set through data resampling mechanisms in the entry space, including minority class sampling, majority class subsampling, or the combination of both techniques. The alternative based on the adaptation of existing learning algorithms aims to improve, at the same time, the number of correct positive classifications and the general accuracy of the classifier [25, 61]. Therefore, efficient treatment of the class imbalance issue is also

vital for predicting customer evasion and dealing with our work.

Several techniques have been applied in the academic literature in the customer’s churn prediction, with emphasis on decision trees [8, 37], k -nearest neighbors [18, 63], elastic net [42], logistic regression [34, 38], SVMs [19, 64], and random forests [39, 57]. Despite presenting good predictive results, most of these studies focused on applying a single statistical model and, in some cases, used artificial and relatively reduced bases in their experiment. Our work distinguishes us from the existing research because we conduct a horserace of several classifiers and use a representative customer-level data set.

Decision tree algorithms are widely used to solve classification problems in machine learning, statistics, and other disciplines [11]. Using decision trees are appropriate when the purpose of data mining is the classification of data or prediction of outputs [51, 54]. Additionally, it is the optimal choice when the objective is to generate rules that are easily comprehended, explicated, and translated into natural language. In decision trees, the first node contains the most critical attribute, while subsequent nodes contain the less critical attributes. Decision trees assist users in determining which attributes have the greatest influence on their prediction tasks.

[23] define k -nearest neighbors as a simple and effective nonparametric classification method. To classify a data record, we retrieve its k closest neighbors from its neighborhood according to some similarity or distance metric. Then, we classify the record according to some function (e.g., majority) based on these nearest neighbors. Therefore, the k -nearest neighbor is sensitive to the choice of the k parameter, which should be defined using a cross-validation procedure.

[9] introduced SVMs, which are kernel-based supervised learning models. Given a labeled training set, the SVM uses a kernel transformation to represent observations as points in a higher multidimensional space. It then tries to identify the best separation hyperplanes (margins) between instances of different classes in this higher dimensional space. In churn prediction, SVM techniques have been extensively investigated and often show high predictive performance [16, 17, 48].

Logistic regression is an extension of the linear regression model adapted to classification problems. The intuition behind logistic regression is quite simple. Because we need a binary result, we perform the following steps [4, 53]: (i) map the linear regression predictions to $[0, 1]$ using a nonlinear function, such as sigmoid; (ii) interpret the new result as the probability of having one as a result; and (iii) predict one if the probability is higher than a chosen threshold (which is often 0.5); otherwise, predict 0.

[65] proposed the elastic net, which is built upon the traditional regression or logistic regression but also incorporates a convex combination of L1 (Lasso) and L2 (Ridge) penalties into the loss function. By introducing these terms, overfitting problems are mitigated and predictive algorithms' generalization power can be significantly increased.

The random forests method, introduced by [10], has performed well compared to many other classifiers. The strategy of the random forests technique is to select random subsets of attributes to cultivate trees so that each tree is grown in a sample of the training set [30]. According to [4], this approach's main disadvantage is computational time, which increases proportionally to the number of trees. Besides, the number of trees required for good performance is directly proportional to the number of predictors [31].

Also noteworthy are ensemble methods, which consist of using several combined learning algorithms to achieve better predictive performance than could be obtained from any learning algorithms in isolation [29].² According to [21], ensemble algorithms have become a popular solution method for unbalanced class problems. The academic literature highlights two of these methods: bagging and boosting. [25] explain that bagging involves having each model in the final decision set with equal decision weight to improve the variance of the model. The technique uses withdrawals from random subsets of the training set. As an example, already mentioned in this work, the random forests algorithm combines decision trees. [56], on the other hand, explore the boosting algorithm. The main goal of boosting is to improve classification performance by combining various classification models, called weak classifiers. This combination produces a new, more accurate classifier in the training set. The most popular boosting algorithm is AdaBoost, where weak classifiers are decision trees [20].

Several of the problems handled through machine learning methods make use of bases that involve many attributes. However, [62] and [32] recommend paying close attention to this theme because not only can many of these attributes be redundant or even irrelevant, but they can also reduce the performance of the models used. The selection of attributes should address this issue by identifying a small subset of relevant attributes from the original set. By removing irrelevant and redundant attributes, we reduce the data dimensionality, thereby accelerating and simplifying the learning process [24, 52]. Several studies in the literature examine the application of attribute selection techniques and classify them into two categories based on their evaluation criterion: filter approaches and wrapper approaches. Wrapper approaches generally achieve better

performance in classification tasks [13, 62]. According to [13], classification problems are typically addressed through a supervised attribute selection approach that uses the correlation between attributes and the class label as its fundamental principle.

3 Data and methodology

This section describes our data set and the data preparation steps, conducts an exploratory data analysis, and details the machine learning methodology used to select models and estimate their performance. Figure 2 depicts a high-level overview of the research steps used in our empirical investigation on customer churn prediction. We follow the phases described in the CRISP-DM process model: business understanding, data understanding, data preparation, modeling, evaluation, and application [15].

In classification problems, records are often well defined. However, there is a peculiarity in churn prediction because sometimes customers do not close the relationship directly with the bank but simply abandon their accounts. In this case, we must define specific criteria to distinguish this type of customer from active clients. It is also common to consider churned customers who do not make transactions or move enough money for a long time [26]. In the banking sector, [40] considers a client as churned when he is inactive for at least six months. In our research, we follow this convention.

We conduct a horserace of supervised learning algorithms to confer robustness to our results [59, 60]. We use the following classifiers for churn prediction: decision trees, k -nearest neighbors, elastic net, logistic regression, SVMs, and random forests. We also compare the performance of an ensemble method composed of the classifiers above. We use the k -fold cross-validation technique for model selection. Finally, we also run feature importance routines embedded in these algorithms to understand those attributes that better predict a potential customer churn.

As mentioned, the experiment's objective is to investigate the effectiveness of different statistical models in predicting the churn of clients of a financial institution. To do this, we built a sample composed of a set of 35 attributes, predominantly related to transactions carried out and persisted based on transactional systems, of 500 thousand customers.

3.1 Universe and study sample

The data used in this study come from a large Brazilian financial institution that reserves the right not to be identified. A sample of anonymized and representative data of

² For interesting applications see also [2] and [55].

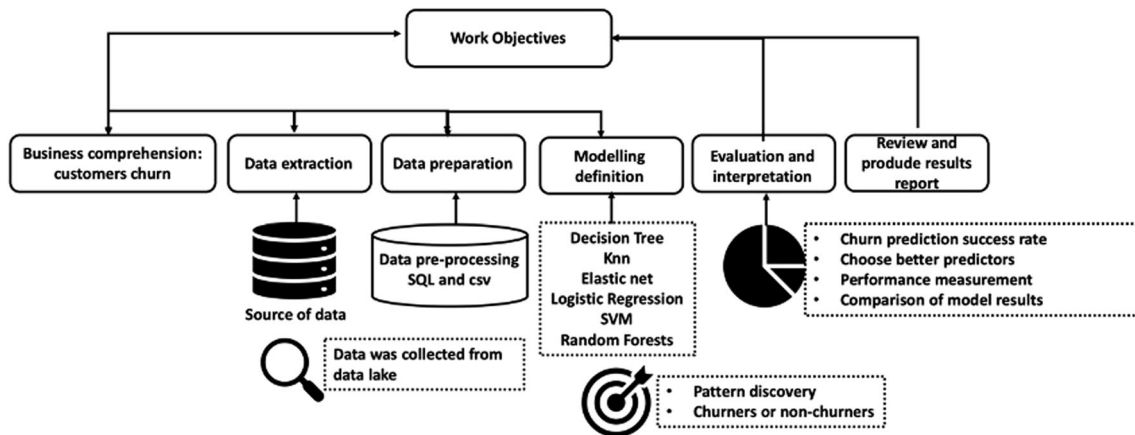


Fig. 2 Schematic of the research steps used in this empirical investigation on customer churn prediction

500,000 customers was used in compliance with bank secrecy standards.

3.2 Data preprocessing and attributes

As with any other bank, customer data is captured and stored throughout service provision. We can describe each customer's behavior through the multiple records spread across various legacy and transactional systems databases that track their operations over time. However, the existence of a data lake that feeds a CRM solution of the financial institution and centralizes the information facilitated the extraction of the necessary data.

As the volume of records is substantial and challenging to manage, it was necessary to prepare the data to reduce the computational time required to analyze it. To build a model and make predictions with practical relevance, we need a set of data representing the population we want to explore (customers). For this, we prepared a dataset at the customer level. Attributes in our dataset represent customers' attributes that, in the business view, summarize the characteristics that could influence their tendency to churn.

When selecting the attributes, we used expert knowledge in the field of banking business to choose a broad set of potential variables whose behavior could be related to a customer's decision to continue or leave the bank. Generally, when data are collected, we do not know which attributes will be significant and irrelevant. Therefore, we opted to select a broad set of potential variables with economic sense to the churn prediction problem and let the classifier output their relevance. After iterating in model definition and performance, we remove irrelevant attributes to improve the model's performance.

We also perform feature engineering to leverage the predictive power of our algorithms. We create features by using summary functions, such as the difference or percentage of other features. We calculate these functions over

6-month periods to track customers' historical movement over time, maintaining a single record per customer. After understanding the business, extraction, and preparation of the data, we established the main set of attributes that we use in our machine learning model. Table 1 reports these selected variables.

In the end, the generated dataset has 500,000 cases of current account clients observed over 12 months of relationship with the institution, accompanied by the final position, whether the client has churned or remained a client, thus enabling a supervised learning process. Each record has 35 attributes related to the client. Besides, we have forcibly balanced the base (subsampling of the majority class) so that, of the 500,000 customers included there, 250,000 churned, and 250,000 did not churn.

In predicting the churn of bank customers, there are different types of outliers. Sometimes, a customer leaves for external reasons outside the bank's control. [46] exemplify death or moving to a different region as events that can result in customer churn. Another unusual scenario is represented by customers who open a new account only for a specific purpose and close it as soon as they achieve their goal. These clients do not provide additional helpful information for churn behavior, so we should remove them ideally. Because these clients are indistinguishable from others, empirical methods are often applied to remove them from the dataset. [41] also suggest ignoring customers whose relationship lasts less than six months and customers who have carried out less than fifty transactions to solve this problem. In our case, we adopt the option to replace only deceased customers and customers with less than twelve months of a relationship or who have not made changes in their accounts in the last six months, starting from the collection date. We do not filter customers who have carried out less than fifty transactions, as we observed that it is widespread for customers to carry out less than

Table 1 Target variable (first row) and selected attributes (remaining rows) used in our customer churn prediction supervised task

Class/Attribute	Data type	Description
Churned	Binary (Yes or No)	Customer closed their current accounts or stopped moving them for six months (churned)
Segment	Nominal (4 segments)	Customer segment (Basic income, middle class, high income, and very high income)
Automatic_Debt	Binary (Yes or No)	Use of the direct debit service - at least once in the last 60 days
Salary_Credit	Binary (Yes or No)	Receipt of salary - at least once in the last 60 days
Accreditation	Binary (Yes or No)	Membership to the accreditation service/card make-up
Insurance	Binary (Yes or No)	Ownership of insurance product consortium, capitalization or pension plan
Portability_Request	Binary (Yes or No)	Request for salary credit portability to another financial institution
Complaint_Request	Binary (Yes or No)	A registered complaint in channels managed by OUVID (Ombudsman, SAC, Procon, BACEN)
Automatic_Debt_DIFF	Real value in $[-1, 1]$	Evolution of the use of the automatic debit service - at least once in the last 60 days
Salary_Credit_DIFF	Real value in $[-1, 1]$	Evolution of salary receipt - at least once in the last 60 days
Insurance_DIFF	Real value in $[-1, 1]$	Evolution of the insurance company's product ownership - insurance, consortium, capitalization or pension
Qualified_Products	Integer	Number of products that the customer owns, and that is indicated for the if segment
Qualified_Products_Previous	Integer	Quantity of customer products, and which is indicated for the Position segment: 6 months before
Qualified_Products_DIFF	Integer	Number of products that the customer owns, and that is indicated for the if segment - Absolute variation between 6 months
Qualified_Products_PERC	Percentage	Number of products that the customer owns, and that is indicated for the if segment - Percentage change between 6 months
Products	Integer	Number of products the customer owns
Products_Previous	Integer	Number of products the customer owns - Position: 6 months before
Products_DIFF	Integer	Number of products that the customer has - Absolute change between 6 months
Products_PERC	Percentage	Number of products that the customer has - Percentage change between 6 months
Transactions	Value in R\$	Number of spontaneous movements carried out in the current account
Transactions_Previous	Value in R\$	Number of spontaneous movements carried out in current account - Position: 6 months before
Transactions_DIFF	Value in R\$	Number of spontaneous movements performed in the current account - Absolute variation between 6 months
Transactions_PERC	Percentage	Number of spontaneous movements performed in the current account - Percentage change between 6 months
Investment	Value in R\$	Volume invested in investments, savings or deposit account
Investment_Previous	Value in R\$	Volume invested in investments, savings or deposit account - Position: 6 months before
Investment_DIFF	Value in R\$	Volume invested in investments, savings or deposit account - Absolute change between 6 months
Investment_PERC	Percentage	Volume invested in investments, savings or deposit account Percentage variation between 6 months
Credit	Value in R\$	The volume of commercial and housing loans active
Credit_Previous	Value in R\$	The volume of commercial and active housing credit - Position: 6 months before
Credit_DIFF	Value in R\$	The volume of commercial and active housing credit - Absolute change between 6 months
Credit_PERC	Percentage	The volume of commercial and active housing credit - Percentage change between 6 months
Profitability	Value in R\$	Profitability (financial return indicator) of the client, accumulated 12 months
Profitability_Previous	Value in R\$	Profitability (financial return indicator) of the client, accumulated 12 months - Position: 6 months before

Table 1 (continued)

Class/Attribute	Data type	Description
Profitability_DIFF	Value in R\$	Profitability (financial return indicator) of the client, accumulated 12 months - Absolute change between 6 months
Profitability_PERC	Percentage	Profitability (financial return indicator) of the client, accumulated 12 months - Percentage change between 6 months

two spontaneous operations per month. Such a filter would eliminate a large portion of the sample.

Concerning eventually missing values, given the abundance of data, we used an automated imputation method. We had less than 6% of the dataset with at least one attribute missing. We replaced the records of customers whose data were not complete with those that had complete records using a k -nearest neighbor procedure with $k = 3$. In this approach, the imputation algorithm only considers other observations with attributes that have information to match the incomplete observations. After the imputation methodology, our sample of 500,000 observations becomes complete.

We finalize the preprocessing routine by apply the steps: removal of near-zero variance attributes and standardization of all numeric variables. The execution of near-variance preprocessing step resulted in the elimination of the following attributes, whose variation of values can be considered negligible: “Segment_FX”, “Accreditation”, “Portability_Request”, “Complaint_Request”, “Automatic_Debt_DIFF”, “Salary_Credit_DIFF” and “Insurance_DIFF”. We then execute the established models using the remaining 28 attributes.

3.3 Representativeness of the data sample

This section provides statistical evidence that our data sample is representative of the entire population in terms of observable characteristics, customer age (birth generation), and geographic dispersion.

We obtained the sample at random from a total population of 9,713,861 customers who moved their current accounts over the 180 days before collection. We calculated some basic statistics about the sample content and compared them to statistics evaluated over the entire population to determine the representativeness of our sample. This analysis is critical to ensuring that our data sample accurately reflects the general behavior of our customers. Tables 2 and 3 show the calculated statistics (mean and median, respectively). In general, our sample’s mean and median are comparable to those found in the population.

Besides checking the representativeness of our sample concerning observable attributes in the dataset, we also

look at whether our data sample matches the population in terms of customers’ age in terms of their generation (baby boomers, generation X, Y, Z, and alpha). This segmentation is a highly relevant aspect to be observed, as customer behavior can vary from generation to generation. Table 4 shows the proportion of customers in the sample and population (in terms of the total number in sample/population). Again, the data sample closely matches the population’s age distribution of customers.

Brazil is a country with continental dimensions, and for this reason, it is essential that the sample faithfully reproduces the actual dispersion of customers in the country. Concerning geographical dispersion of customers, Table 5 shows that the data sample also accurately reflects the distribution of customers across the country’s states.

3.4 Model selection

We present the logic used for modeling and analysis of the sample in Fig. 3. The total period consists of 12 months. Due to the temporal nature of our data, we must use historical data to forecast future behavior: the first six months are used to construct predictors (attributes), and the last six months are used to define the target variable. Therefore, our attributes are composed of customers’ financial traits extracted from August 2018 to January 2019 (red color). Our target is to determine whether the client churned during the subsequent six months, i.e., from February to July 2019 (blue color).

We first divide our entire dataset into two disjoint but complete subsets (holdout): the training set and the test set. Since we have many observations, we use 90% of our sample to train the model (training set) and the remaining 10% to test its performance on unseen data (test set). This division is important so that our performance indicators do not become overoptimistic. To perform model selection, we apply a standard k -fold cross-validation procedure only using data from the training set ($k = 10$). To further reduce variance in our model selection, we independently repeat the cross-validation procedure ten times and take the average across these runs. Since we are collapsing our customer-level data into a single point (each customer is a data point in the dataset), it is reasonable to assume that our

Table 2 Comparison of the means of the data sample and the entire population

Attribute	Population (churned=0)	Population (churned=1)	Sample (churned=0)	Sample (churned=1)
Number of customers	8.879.145	834.716	250.000	250.000
Automatic_Debt	0.17	0.03	0.17	0.03
Salary_Credit	0,57	0.29	0.57	0.29
Qualified_Products	5.74	4.32	5.74	4.32
Products	6.75	5.10	6.75	5.10
Transactions	8.84	2.26	8.85	2.26
Investment	24,481.11	2,450.42	23,953.49	21,785.40
Credit	38,051.83	16,690.32	38,165.85	16,591.60

Table 3 Comparison of the medians of the data sample and the entire population

Attribute	Population (churned=0)	Population (churned=1)	Sample (churned=0)	Sample (churned=1)
Number of customers	8.879.145	834.716	250.000	250.000
Automatic_Debt	0	0	0	0
Salary_Credit	1	0	1	0
Qualified_Products	6	4	6	4
Products	6	5	6	5
Transactions	4.67	1	4.67	1
Investment	755.14	57.00	755.01	56.78
Credit	5,066.19	50.43	5,047.06	46.65

Table 4 Comparison of sample x population generations

Generations	% sample	% population
Baby boomers (until 1960)	21.9	23.6
Generation X (1961–1980)	41.3	42.3
Generation Y (1981–1997)	35.5	33.1
Generation Z (1998–2009)	0.9	0.6
Generation alpha (since 2010)	0	0

observations are iid, in such a way that a standard k -fold cross-validation is a valid model selection procedure.

3.5 Exploratory data analysis

This section provides an exploratory and visual analysis of the data to try to anticipate the understanding of which attributes could be most useful to separate the classes effectively. Figures 4, 5, and 6 show boxplots of some attributes with good predictive power. From the visual inspection, we can see that the Transactions attribute (Fig. 4) stands out as a potential good predictor of evasion for all segments of the institution's customers. To a lesser extent, the attributes Qualified_Products (Fig. 5) and

volume of credit (Fig. 6) also appear to have the potential for contribution.

4 Discussions and results

In this section, we report the main empirical results.

4.1 Horserace results

Table 6 reports the classifiers used in this paper to compare churn prediction along with the alias used in this section. As discussed earlier, we use a repeated k -fold cross-validation procedure with $k = 10$ (ten repeats) using data only from the training set (90% of the entire dataset) for selecting the best set of hyperparameters for each model (model selection procedure). We define the optimizing performance metric as the AUC-ROC, because it is a robust measure that is independent of the threshold used to determine the target class of the instances.

After the model selection procedure, we retrain each model with the optimal hyperparameters using the full training set. Then, we test the models' performance against the test set. Table 7 and Fig. 7 (box-and-whisker graph) present these results on the test set (10% of the entire data

Table 5 Comparison of state of residence sample x population

Region	State	% sample	% population
North	Acre	0.4	0.4
	Amazonas	0.8	0.7
	Amapá	0.3	0.3
	Pará	1.9	1.7
	Rondônia	0.6	0.6
	Roraima	0.2	0.2
	Tocantins	0.5	0.5
	North	4.7	4.4
Northeast	Alagoas	1.5	1.9
	Bahia	4.5	4.4
	Ceará	2.9	2.8
	Maranhão	1.3	1.2
	Paraíba	1.2	1.2
	Pernambuco	2.9	2.8
	Piauí	1.1	1.1
	Rio Grande do Norte	1.2	1.3
	Sergipe	0.9	0.9
	Northeast	17.5	17.6
Midwest	Distrito Federal	1.7	1.8
	Goiás	4.9	5.5
	Mato Grosso do Sul	1.3	1.3
	Mato Grosso	1.3	1.2
	Midwest	9.2	9.8
Southeast	Espírito Santo	1.8	1.9
	Minas Gerais	11.5	11.6
	Rio de Janeiro	7.6	7.0
	São Paulo	24.2	22.9
	Southeast	45.1	43.4
South	Paraná	7.8	8.6
	Rio Grande do Sul	8.4	8.6
	Santa Catarina	5.8	6.1
	South	22.0	23.3

set) for each of the six classifiers used. The analysis in Figure 7 allows us to conclude that, on average, the random forests model resulted in a higher ROC value. The results presented in Table 7 show the random forests superiority in three metrics: Accuracy, Precision, and F-measure (Fig. 8).

We also use an ensemble composed of the six classifiers used before. The following models were combined: decision trees, *k*-nearest neighbors, elastic net, logistic regression, SVM, and random forests. To train this model, we again apply a model selection procedure using only data from the training set. We optimize the weight of each of the constituent classifiers in the overall voting scheme of the ensemble. Here we follow the literature and fix the hyperparameters of each constituent classifier with the optimal hyperparameter values discovered in their individual model selection procedures and tune only the weights of each classifier using the ensemble’s voting scheme. After tuning, we obtain the following optimal combination: Ensemble = 3.2727 + 0.054 · Decision Tree – 0.266 · K-Nearest Neighbors – 0.0764 · ElasticNet – 4.7868 · Logistic Regression + 4.604 · SVM – 5.9724 · Random Forests. Table 8 and Fig. 9 show the results. The ensemble’s ROC (0.9018) was not statistically superior to the Random Forests’ classifier alone (0.9015). The ensemble did not obtain superior results because the classifiers are not weakly correlated.

4.2 Identification of attributes with high predictive power

Figure 9 shows the trained decision tree to analyze the churn propensity of customers in the next six months. As in the decision trees, the most critical attribute is the first node. The result confirms the expectation generated already in the stage of exploration of the data (visualization of boxplots to anticipate attributes with good predictor potential). This occurs since the attribute we choose as the

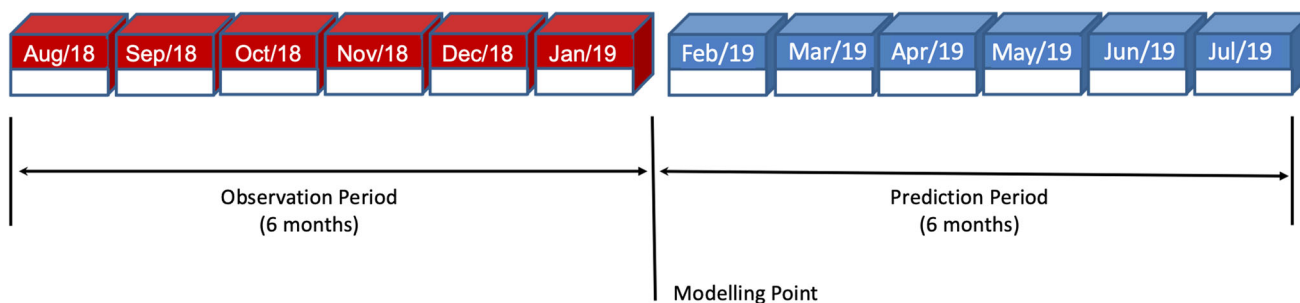


Fig. 3 Modeling strategy to construct our churn prediction model. Due to the temporal nature of our data, we must use historical data to forecast future behavior. Therefore, our attributes are composed of

customer’s financial traits extracted during August 2018 to January 2019 (red color). Our target is whether the client churned in the following six months, i.e., February to July 2019 (blue color)

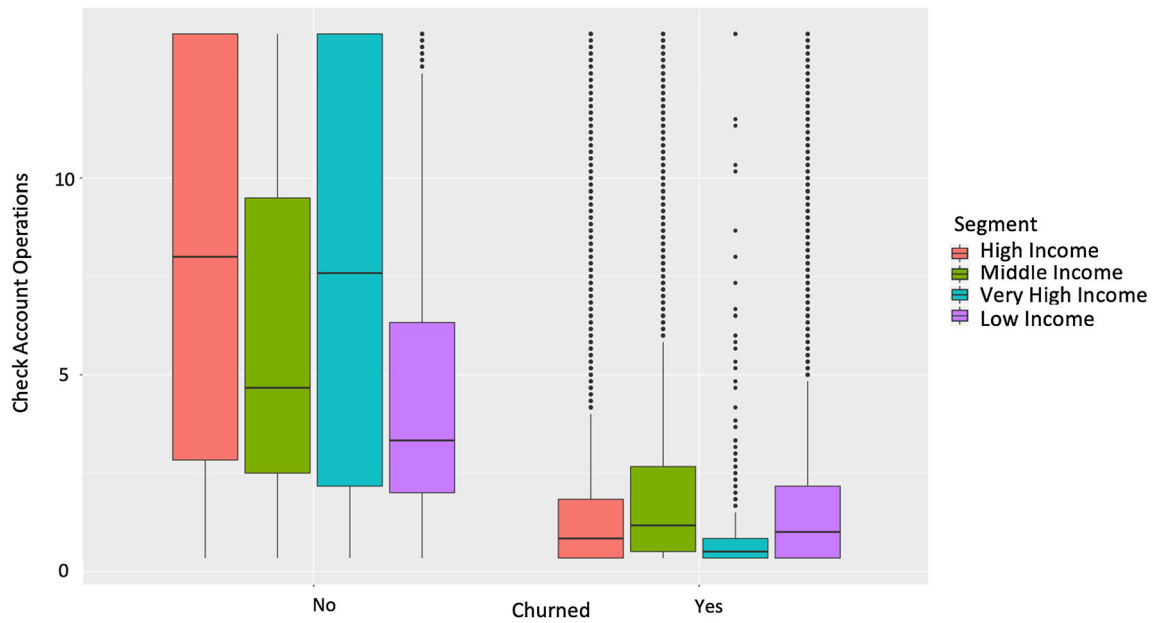


Fig. 4 Boxplot of the attribute check account transactions/operations (vertical axis) versus the target binary variable indicating whether the client churned in the next six months (horizontal axis). We categorize the plots according to the customer segment (very high, high, middle, low income)

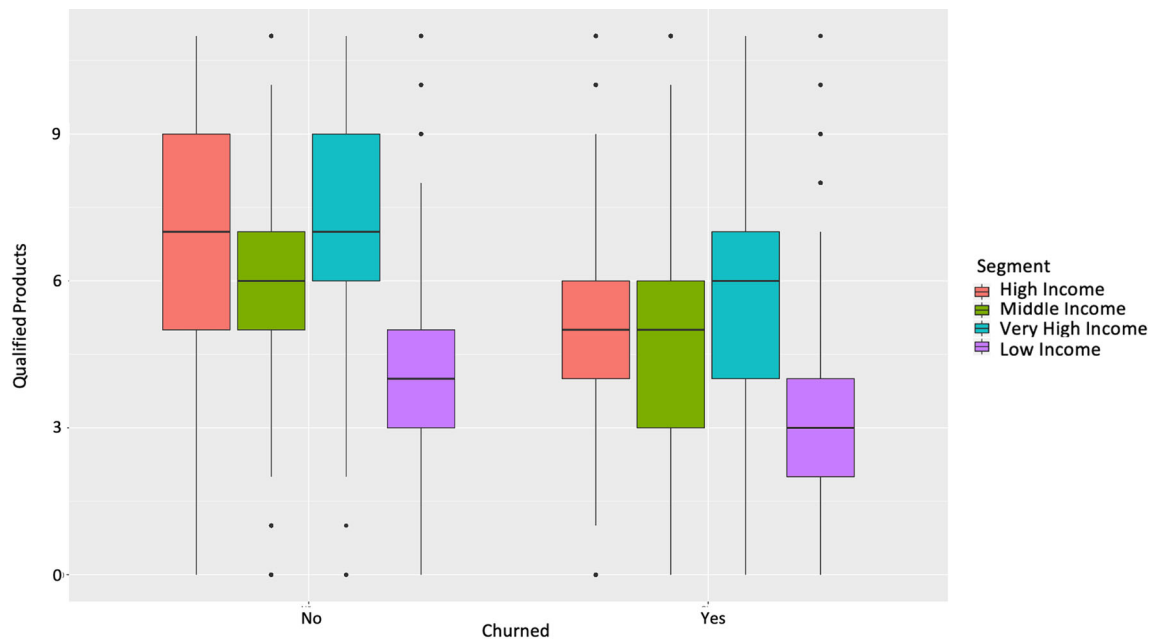


Fig. 5 Boxplot of the attribute number of qualified products (vertical axis) versus the target binary variable indicating whether the client churned in the next six months (horizontal axis). We categorize the plots according to the customer segment (very high, high, middle, low income)

first node of the tree was the column “Transactions,” and then the attributes “Investment” and “Credit.” The confirmation of this result indicates that, in the churn prediction, the observation of the financial flow (number of transactions carried out) often has a predictive potential higher than the observation of the variation of the amounts (balances) involved in investments and loans.

We now analyze the average importance of attributes in the different methods that we used in our horserace. A ranking of the most relevant attributes was then generated based on the average classification of the decision tree algorithms, logistic regression, and elastic net. We do not use all methods because some of them do not have an internal mechanism to quantify the attribute’s importance.

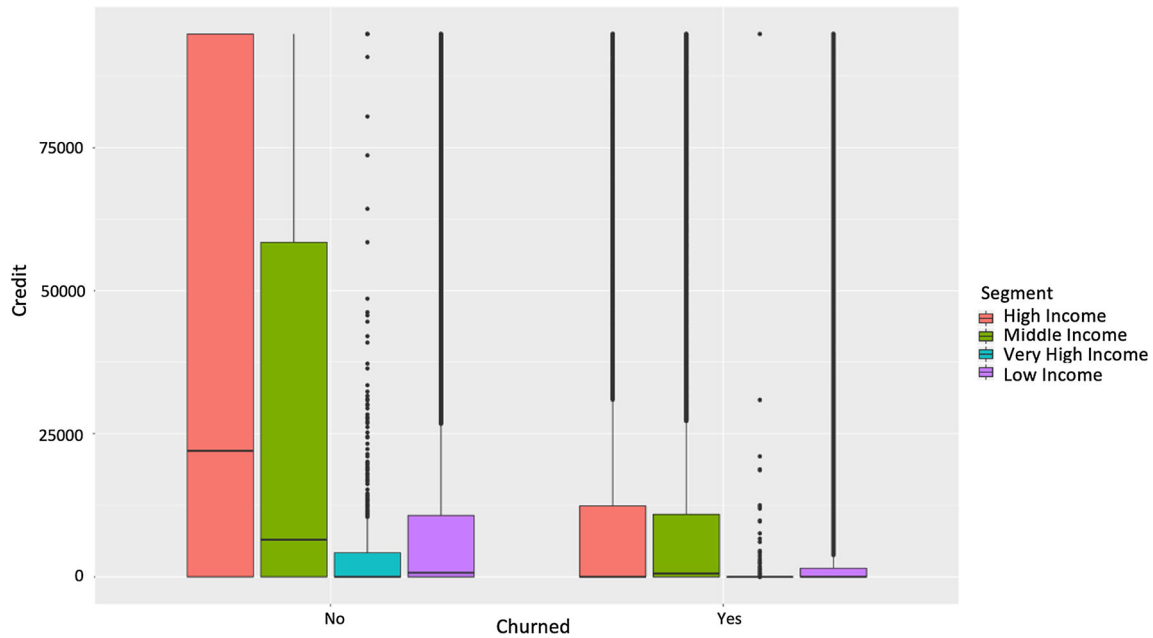


Fig. 6 Boxplot of the attribute volume of credit (vertical axis) versus the target binary variable indicating whether the client churned in the next six months (horizontal axis). We categorize the plots according to the customer segment (very high, high, middle, low income)

Table 6 Models used for classification (horserace)

Classifier	Method’s alias	Description
Decision trees	Rpart	Recursive partitioning and regression trees
<i>k</i> -nearest neighbors	knn	<i>k</i> -nearest neighbors
Logistic regression	glm (Family = binomial)	Generalized linear model
Elastic net	elasticnet	Logistic regression regularized with lasso and ridge
Support vector machines	svm	SVM with radial kernel function
Random forests	rf	Random forests

Table 7 Performance of the optimized models on the test set. Test data was heldout during the entire process (10% of the dataset)

	True Positive	True Negative	False Positive	False Negative	Accuracy	Precision	F-measure
Decision tree	38.8	39.4	10.6	11.2	78.2	78.5	78.05
Knn	37.8	40.1	9.9	12.2	77.9	79.2	77.36
Elastic net	40.5	35.7	14.3	9.5	76.2	73.9	77.29
Logistic regression	40.4	35.8	14.2	9.6	76.2	74.0	77.26
Svm	39.6	40.7	9.3	10.4	80.3	81.0	80.09
Random forests	40.1	42.6	7.4	9.9	82.8	84.4	82.25

We used training data (90% of the dataset) for model selection: we apply a repeated *k*-fold cross validation (ten independent times) with *k* = 10 to optimize the hyperparameters of each model. We used ROC as the optimizing metric in the training process. After selecting the best hyperparameters, we retrain each model with the entire training set (because we hold out one fold each time in a cross-validation procedure) with these optimized values.

Figure 10 ranks the 27 attributes in order of relevance for predicting the bank customer’s intention of churn. We normalize the coefficients in terms of the most crucial

attribute. The attribute “Credit” is presented as the most powerful predictor of customer churn, followed by “Profitability” and “Transactions.” There is a considerable

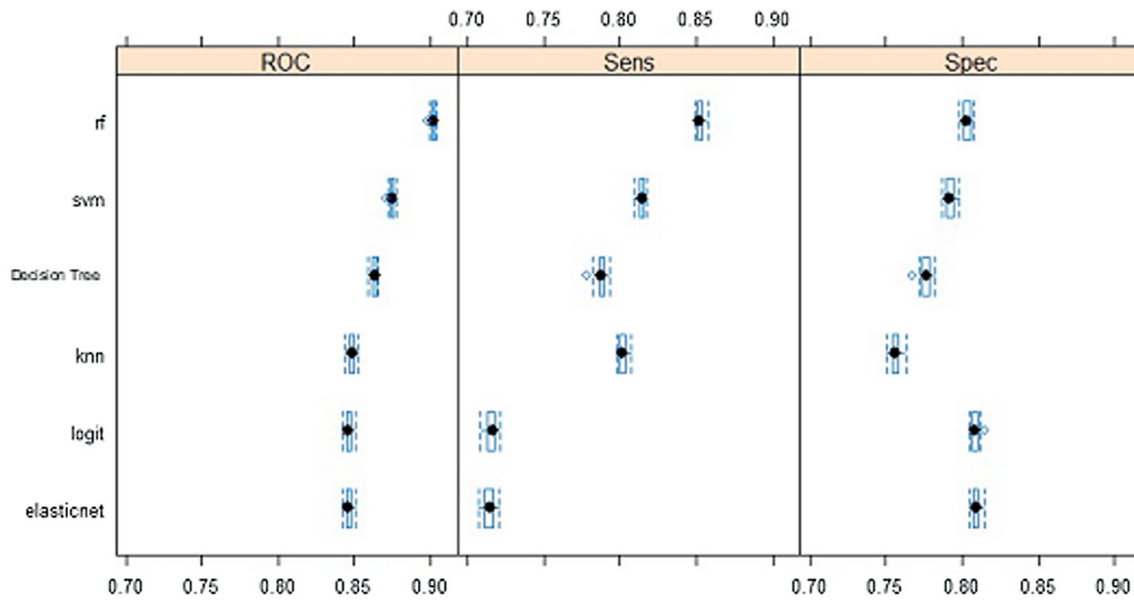


Fig. 7 Performance metrics on the test set of the six employed classifiers using a box-and-whisker graph. The black dot represents the median performance (color figure online)

Fig. 8 Performance of the optimized models for the metric ROC in the test set. The dot represents the average performance and the vertical bars, the standard error. The horizontal red-dashed line is the ensemble’s performance

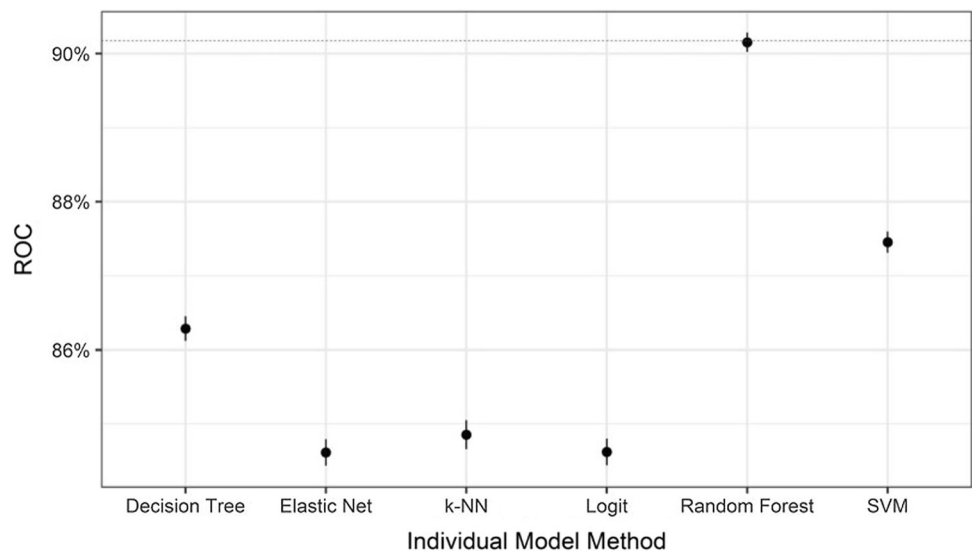


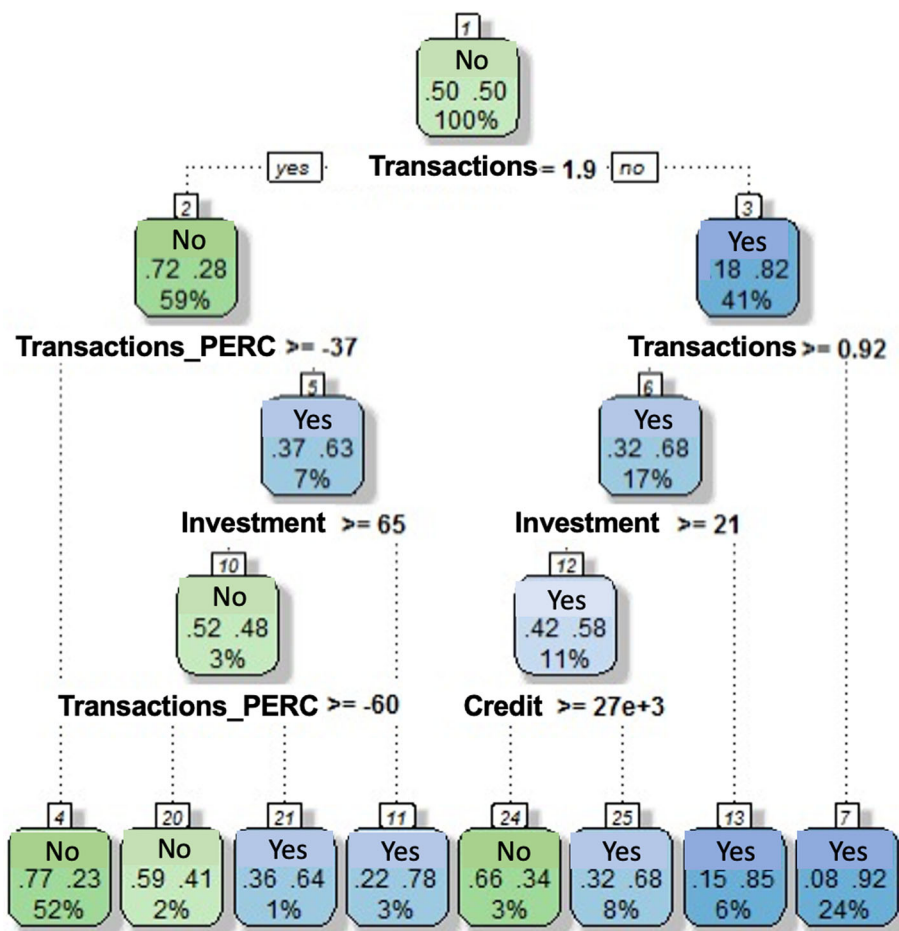
Table 8 Result (ROC) of each model and the associated standard deviation on the test set, including the ensemble

Method	ROC	Std. Dev. ROC
Ensemble of classifiers below	0.9018	
Decision tree	0.8628	0.0017
Knn	0.8485	0.0020
Elastic net	0.8461	0.0018
Logistic regression	0.8462	0.0018
Svm	0.8746	0.0014
Random forests	0.9015	0.0013

decrease in attributes’ importance, suggesting that few attributes could be sufficient for an efficient prediction.

Certain business perceptions are possible as a result of the ranking described above. The “Credit” attribute has the highest statistical power to predict customer churn. This attribute represents the volume of commercial or residential loans customers hold with the institution. As a result of this finding, it is reasonable to conclude that clients who take more expressive credit with the institution, such as real estate credit, are more likely to keep their accounts active throughout the commitment term. Although this condition has not historically demonstrated a guarantee of these customers’ engagement with the bank, maintaining

Fig. 9 Trained decision tree to predict customer churn in the next six months. Within each node, the first row shows the predicted class if one stops traversing the tree at that node. The second row shows the proportion of clients that do not churn and churn for the subset of data that falls in that tree node. The third row shows the support: the fraction of data that falls within that tree as a share of the total number of observations (in percent)



active accounts has been a requirement of these institutions to offer more favorable rates, which may explain such behavior.

The second position appears the variable “Profitability,” which represents the client’s financial return to the institution. This result is natural given that, in general, this financial return is directly proportional to the volume of credit extended or the number of products and services consumed by the client from the bank.

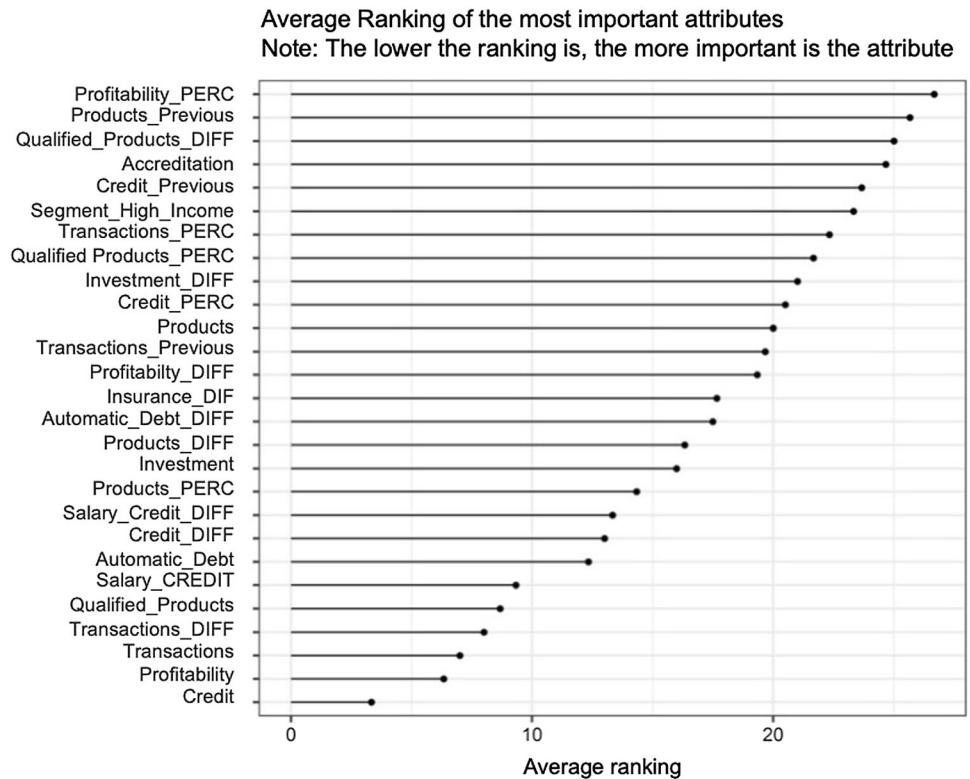
The third and fourth positions in the ranking correspond to the attribute “Transactions,” representing the average amount of current account transactions (credits and debits). This result confirms our expectation that we identified in our exploratory data analysis and in the decision tree generation. The result indicates that the lower the average number of transactions in the current account, the greater the likelihood of the account being closed. Intuitively, few account entries imply little activity and may indicate a lack of relationship or customer engagement with the bank. One possibility is that the customer has transferred his finances to another financial institution. One suggestion might be to keep an eye on these instances and take steps to strengthen the relationship. On the other hand, continuous monitoring

of a possible decrease in the volume of transactions should be undertaken, as this indicates a weakening of the relationship and an increase in the likelihood of churn.

The fifth position is “Qualified_Products,” which indicates the number of active banking products the customer has with the bank. The possible interpretation is that the more products the customer has, the more engaged with the bank he is. As the institution has a high value for the customer, the cost of leaving the institution may be higher from the customer’s viewpoint. This finding is another example of the importance of a strategy to strengthen the relationship between customers and institutions by selling additional products (cross-selling).

On the other hand, receiving the salary at the bank and having one or more bills with automatic debit (sixth and seventh positions) are banking services with recurring characteristics, suggesting a higher approximation of the client with the institution. A customer who directs their monthly salary to a bank account or even registers their monthly accounts to be debited automatically from their bank account balance is interested in a more robust and long-term relationship with the bank. Therefore, the customer exhibits a lower propensity to churn.

Fig. 10 Ranking of attributes. We take the average rank of each attribute across the methods used in the horserace (when applicable)



In general, the business intuition generated from the research results suggests that customers with a stronger relationship with the institution have a lower likelihood of closing their current accounts.³ Thus, cross-selling and up-selling strategies may be beneficial: strengthening the relationship through increased quantity and use of products and services can improve customer satisfaction and increase the cost of change, thereby contributing to customer retention. After the model selection procedure, we retrain each model with the optimal hyperparameters using the training set. Then, we test their performance against the test set. Table 7 and Fig. 7 (box-and-whisker graph) present these results on the test set (10% of the entire data set) for each of the six classifiers used. The analysis in Fig. 7 allows us to conclude that, on average, the random forests model resulted in a higher ROC value. The results presented in Table 7 show the random forests model’s superiority in the three comparison metrics: Accuracy, Precision, and F-measure.

³ That is, those who move their accounts frequently, have a greater variety of products and services, and obtain conforming bank loans.

5 Conclusions

This article aimed to evaluate supervised classifiers typically used in banking to predict customer churn using a unique dataset from a large Brazilian bank. Our paper contributes to the existing empirical literature on customer churn in several ways. First, we conduct a horserace of a set of supervised learning classification algorithms under the same validation and evaluation methodology to determine the algorithm that is best suited for our dataset. Second, we compile a unique and representative dataset of a large Brazilian bank at the customer level over time. Most empirical studies either use artificial datasets or aggregate data from a specific bank, which could compromise the empirical conclusions. This data limitation occurs because customer-level bank data is private and legally protected. Third, we leverage the availability of a large number of attributes in our dataset not only to obtain accurate predictions of customer churn but also to understand which attributes have the highest predictive power when determining the likelihood of a potential churn in the next semester.

We employed the following supervised classifiers in the horserace: decision trees, logistic regression, *k*-nearest neighbors, elastic net, SVMs, and random forests. We applied a repeated *k*-fold cross-validation to select the best hyperparameters for each model. Then, we evaluated the model’s performance using holdout test data. The random

forests model achieved the best results, even compared to an ensemble model composed of the above classifiers. Both random forests and the ensemble could be used in the banking environment to direct CRM efforts to promote customer retention and maintenance and lasting relationships between these institutions and their customer base.

Another important finding of the study was identifying attributes with the highest predictive power of a potential customer churn. We found that the frequency with which customers used financial services, the volume of credit extended to them (concessions), and their possession of products had higher predictive power than attributes related to transacted volumes (balances). Thus, strengthening the relationship with customers through the sale of products can be an effective strategy for customer retention and churn mitigation.

Finally, we conclude that predicting customer churn is a challenging task due to its temporal nature, which increases the overall complexity of data analysis. However, our results highlight that machine learning can help banks understand their customers' behavior in an automated way, thereby enabling them to act proactively and in advance to reverse a potential customer churn and mitigate revenue losses.

The random forests model yielded consistent models with superior results in our experiments. Using this technique, we were able to identify 80.2% of the customers who would churn in the following months (recall). On the other hand, 14.8% of customers who did not churn were classified as prone to churn (1 - specificity), which is a reasonable proportion given that the repercussions do not always result in a problem. On the contrary, it may even provide customers with increased satisfaction due to the extra attention received. As long as the data is distributed and treated correctly, the typical predictive model we discovered in our study can begin adding value to banks on the first day. Customer retention teams would then approach these customers with offers capable of reversing the churn in the most effective manner possible.

The customer's profitability (contribution margin) is calculated by subtracting the revenue stream from the operation's maintenance costs. Considering the customer's average margin of the institution and the accuracy of 80.2% achieved by the random forests model in detecting potential churn over a year,⁴ we conclude that the model's application has the potential to forecast annual losses of up to R\$ 2.12 billion at the customer level. This number accounts for up to 10% of the largest Brazilian banks'

operating results in 2019, highlighting the practical importance of the trained model.

Even the most conservative and linear percentage reversal projections can be attractive. For instance, if we applied a linear approach to all clients identified as potential churners by the model and the campaign achieved a 20% success rate, this action would preserve approximately R\$ 290 million in annual revenue.

We applied a simple calculation above based on the average margin generated by a customer of the institution. However, we know that the margin varies according to the customer's profile. Assuming that 20% of customers are responsible for 80% of the bank's results when we apply an approach strategy based on the return provided individually by each customer, the retention action's efficiency based on the evasion prediction model can achieve even better results in cost/benefit terms.

Acknowledgements Thiago C. Silva (Grant no. 308171/2019-5, 408546/2018-2) and Benjamin M. Tabak (Grants no. 310541/2018-2, 425123/2018-9) have received financial support from the CNPq foundation.

Declarations

Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Agarwal P, Nieto JJ, Ruzhansky M, Torres DF (2021) Analysis of infectious disease problems (Covid-19) and their global impact. Springer, New York
2. Ahmed M, Afzal H, Siddiqi I, Amjad M, Khurshid K (2020) Exploring nested ensemble learners using overproduction and choose approach for churn prediction in telecom industry. *Neural Comput Appl* 32:3237–3251
3. Au T, Ma G, Li S (2003) Applying and evaluating models to predict customer attrition using data mining techniques. *J Comp Int Manag* 6(1):10–22
4. Avon V (2016) Machine learning techniques for customer churn prediction in banking environments. Doctorate Thesis. Università degli Studi di Padova, Italy
5. BACEN (2018) Relatório de Economia Bancária (Banking Report). Banco Central do Brasil. https://www.bcb.gov.br/content/publicacoes/relatorioeconomibancaria/reb_2018.pdf
6. Ballings M, Van den Poel D (2012) Customer event history for churn prediction: how long is long enough? *Expert Syst Appl* 39(18):13517–13522
7. Berry MJ, Linoff GS (2004) Data mining techniques: for marketing, sales, and customer relationship management. Wiley, USA
8. Bin L, Peiji S, Juan L (2007) Customer churn prediction based on the decision tree in personal handyphone system service In 2007 International Conference on Service Systems and Service Management, pp 1–5 IEEE
9. Boser BE, Guyon IM, Vapnik VN (1992) A training algorithm for optimal margin classifiers In Proceedings of the fifth annual workshop on Computational learning theory, pp 144–152

⁴ We can obtain the results over a year by extrapolating the sample numbers for the year (834,716 customers dropped out within a semester multiplied by two).

10. Breiman L (2001) Random forests. *Mach Learn* 45(1):5–32
11. Breslow LA, Aha DW (1997) Simplifying decision trees: a survey. *Knowl Eng Rev* 12(1):1–40
12. Burez J, Van den Poel D (2009) Handling class imbalance in customer churn prediction. *Expert Syst Appl* 36(3):4626–4636
13. Cai J, Luo J, Wang S, Yang S (2018) Feature selection in machine learning: a new perspective. *Neurocomputing* 300:70–79
14. Capgemini E (2019) World retail banking report (last accessed on 03/28/2020)
15. Chapman P, Clinton J, Kerber R, Khabaza T, Reinartz T, Shearer C, Wirth R et al (2000) CRISP-DM 1.0: step-by-step data mining guide, vol 9. SPSS inc., p 13
16. Coussement K, Van den Poel D (2008) Churn prediction in subscription services: an application of support vector machines while comparing two parameter-selection techniques. *Expert Syst Appl* 34(1):313–327
17. Dehghan A, Trafalis T (2012) Examining churn and loyalty using support vector machine. *Bus Manag Res* 1(4):153
18. Eastwood M, Gabrys B (2009) A non-sequential representation of sequential data for churn prediction. In *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*, pp 209–218 Springer
19. Farquad MAH, Ravi V, Raju SB (2014) Churn prediction using comprehensible support vector machine: an analytical CRM application. *Appl Soft Comput* 19:31–40
20. Freund Y, Schapire RE (1997) A decision-theoretic generalization of on-line learning and an application to boosting. *J Computer Syst Sci* 55(1):119–139
21. Galar M, Fernandez A, Barrenechea E, Bustince H, Herrera F (2011) A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches. *IEEE Transactions Syst Man Cybern Part C (Appl Rev)* 42(4):463–484
22. Gomber P, Kauffman RJ, Parker C, Weber BW (2018) On the fintech revolution: interpreting the forces of innovation, disruption, and transformation in financial services. *J Manag Information Syst* 35(1):220–265
23. Guo G, Wang H, Bell D, Bi Y, Greer K (2003) Knn model-based approach in classification In *OTM Confederated International Conferences*” On the Move to Meaningful Internet Systems”, pp 986–996 Springer
24. Guyon I, Elisseeff A (2003) An introduction to variable and feature selection. *J Mach Learn Res* 3(Mar):1157–1182
25. Haixiang G, Yijing L, Shang J, Mingyun G, Yuanyue H, Bing G (2017) Learning from class-imbalanced data: review of methods and applications. *Expert Syst Appl* 73:220–239
26. Hiziroglu A, Seymen OF (2014) Modelling customer churn using segmentation and data mining. *Front Artif Intell Appl* 270:259–271
27. Idris A, Khan A (2012) Customer churn prediction for telecommunication: employing various various features selection techniques and tree based ensemble classifiers In *2012 15th International Multitopic Conference (INMIC)*, pp 23–27 IEEE
28. Kaur M, Singh K, Sharma N (2013) Data mining as a tool to predict the churn behaviour among Indian bank customers. *Int J Recent Innov Trends Comput Commun* 1(9):720–725
29. Krawczyk B, Schaefer G (2013) An improved ensemble approach for imbalanced classification problems In *2013 IEEE 8th international symposium on applied computational intelligence and informatics (SACI)*, pp 423–426 IEEE
30. Larivière B, Van den Poel D (2005) Predicting customer retention and profitability by using random forests and regression forests techniques. *Expert Syst Appl* 29(2):472–484
31. Liaw A, Wiener M et al (2002) Classification and regression by randomForest. *R news* 2(3):18–22
32. Maldonado S, Weber R, Famili F (2014) Feature selection for high-dimensional class-imbalanced data sets using support vector machines. *Information Sci* 286:228–246
33. Miguéis VL, Van den Poel D, Camanho AS, e Cunha JF (2012) Modeling partial customer churn: on the value of first product-category purchase sequences. *Expert Syst Appl* 39(12):11250–11256
34. Mutanen T, Ahola J, Nousiainen S (2006) Customer churn prediction—a case study in retail banking In *Proc of ECML/PKDD Workshop on Practical Data Mining*, pp 13–19
35. Neslin SA, Gupta S, Kamakura W, Lu J, Mason CH (2006) Defection detection: measuring and understanding the predictive accuracy of customer churn models. *J Market Res* 43(2):204–211
36. Nguyen EHX (2011) Customer churn prediction for the Icelandic mobile telephony market Ph. D. thesis, Faculty of Industrial Engineering, Mechanical Engineering and Computer Science, University of Iceland
37. Nie G, Rowe W, Zhang L, Tian Y, Shi Y (2011) Credit card churn forecasting by logistic regression and decision tree. *Expert Syst Appl* 38(12):15273–15285
38. Nie G, Wang G, Zhang P, Tian Y, Shi Y (2009) Finding the hidden pattern of credit card holder’s churn: a case of China In *International Conference on Computational Science*, pp 561–569 Springer
39. Patil AP, Deepshika M, Mittal S, Shetty S, Hiremath SS, Patil YE (2017) Customer churn prediction for retail business In *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*, pp 845–851 IEEE
40. Popović D, Bašić BD (2009) Churn prediction model in retail banking using fuzzy C-means algorithm. *Informatika* 33:2
41. Prasad UD, Madhavi S (2012) Prediction of churn behavior of bank customers using data mining tools. *Bus Intell J* 5(1):96–101
42. Prashanth R, Deepak K, Meher AK (2017) High accuracy predictive modelling for customer churn prediction in telecom industry In *International Conference on Machine Learning and Data Mining in Pattern Recognition*, pp 391–402 Springer
43. PwC (2014) Retail banking 2020 evolution or revolution? (last accessed on 03/28/2020)
44. Rajchakit G, Agarwal P, Ramalingam S (2021) Stability analysis of neural networks. Springer, New York
45. Rajchakit G, Sriraman R, Boonsatit N, Hammachukiattikul P, Lim CP, Agarwal P (2021) Exponential stability in the Lagrange sense for Clifford-valued recurrent neural networks with time delays. *Adv Diff Equ* 2021:256
46. Rajeswari M, Devi T (2015) Design of modified ripper algorithm to predict customer churn. *Int J Eng Technol* 4(2):408
47. Sabbeh SF (2018) Machine-learning techniques for customer retention: a comparative study *Int J Adv Computer Sci Appl* 9(2):273–281
48. Shaaban E, Helmy Y, Khedr A, Nasr M (2012) A proposed churn prediction model. *Int J Eng Res Appl* 2(4):693–697
49. Sharma A, Panigrahi D, Kumar P (2013) A neural network based approach for predicting customer churn in cellular network services *arXiv preprint arXiv:1309.3945*
50. Sia SK, Soh C, Weill P (2016) How DBS bank pursued a digital business strategy. *MIS Q Executive* 15(2):105–121
51. Silva TC, Zhao L (2012) Network-based high level data classification. *IEEE Transactions Neural Netw Learn Syst* 23(6):954–970
52. Silva TC, Zhao L (2012) Network-based stochastic semisupervised learning. *IEEE Transactions Neural Netw Learn Syst* 23(3):451–466
53. Silva TC, Zhao L (2012) Stochastic competitive learning in complex networks. *IEEE Transactions Neural Netw Learn Syst* 23(3):385–398

54. Silva TC, Zhao L (2016) Machine learning in complex networks, vol 1. Springer, New York
55. Sivasankar E, Vijaya J (2019) Hybrid PPFCM-ANN model: an efficient system for customer churn prediction through probabilistic possibilistic fuzzy clustering and artificial neural network. *Neural Comput Appl* 31:7181–7200
56. Vafeiadis T, Diamantaras KI, Sarigiannidis G, Chatzisavvas KC (2015) A comparison of machine learning techniques for customer churn prediction. *Simul Model Pract Theory* 55:1–9
57. Wang G, Liu L, Peng Y, Nie G, Kou G, Shi Y (2010) Predicting credit card holder churn in banks of China using data mining and MCDM In 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, Volume 3, pp 215–218 IEEE
58. Wen Z, Yan J, Zhou L, Liu Y, Zhu K, Guo Z, Li Y, Zhang F (2018) Customer churn warning with machine learning In The Euro-China Conference on Intelligent Data Analysis and Applications, pp 343–350 Springer
59. Wolpert DH (1996) The lack of a priori distinctions between learning algorithms. *Neural Comput* 8(7):1341–1390
60. Wolpert DH, Macready WG (1997) No free lunch theorems for optimization. *IEEE Transactions Evolut Comput* 1(1):67–82
61. Xiao J, Xiao Y, Huang A, Liu D, Wang S (2015) Feature-selection-based dynamic transfer ensemble model for customer churn prediction. *Knowl Information Syst* 43(1):29–51
62. Xue B, Zhang M, Browne WN, Yao X (2015) A survey on evolutionary computation approaches to feature selection. *IEEE Transactions Evolut Comput* 20(4):606–626
63. Zhang Y, Qi J, Shu H, Cao J (2007) A hybrid KNN-LR classifier and its application in customer churn prediction In 2007 IEEE International Conference on Systems, Man and Cybernetics, pp 3265–3269 IEEE
64. Zhao Y, Li B, Li X, Liu W, Ren S (2005) Customer churn prediction using improved one-class support vector machine In International Conference on Advanced Data Mining and Applications, pp 300–306 Springer
65. Zou H, Hastie T (2005) Regularization and variable selection via the elastic net. *J R Statistical Soc series B (Statistical Methodol)* 67(2):301–320

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.