Research in International Business and Finance 65 (2023) 101913

Contents lists available at ScienceDirect

Research in International Business and Finance

journal homepage: www.elsevier.com/locate/ribaf

Modelling bank customer behaviour using feature engineering and classification techniques☆

Mohammad Zoynul Abedin [a,*], Petr Hajek [b], Taimur Sharif [c], Md. Shahriare Satu [d], Md. Imran Khan [e]

[a] Department of Finance, Performance & Marketing, Teesside University International Business School, Teesside University, Middlesbrough, TS1 3BX Tees Valley, United Kingdom
[b] Science and Research Centre, Faculty of Economics and Administration, University of Pardubice, Studentska 84, 532 10 Pardubice, Czech Republic
[c] Faculty of Arts, Society & Professional Studies, Newman University, Bartley Green, Gunners Lane, Birmingham B32 3NT, United Kingdom
[d] Department of Management Information Systems, Noakhali Science & Technology University, Bangladesh
[e] Department of Computer Science and Engineering, Gono Bishwabidyalay, Bangladesh

ARTICLE INFO

Keywords:
Customer behaviour
Data mining
Feature transformation
Feature selection
Classification techniques

ABSTRACT

This study investigates customer behaviour and activity in the banking sector and uses various feature transformation techniques to convert the behavioural data into different data structures. Feature selection is then performed to generate feature subsets from the transformed datasets. Several classification methods used in the literature are applied to the original and transformed feature subsets. The proposed combined knowledge mining model enable us to conduct a benchmark study on the prediction of bank customer behaviour. A real bank customer dataset, drawn from 24,000 active and inactive customers, is used for an experimental analysis, which sheds new light on the role of feature engineering in bank customer classification. This paper's detailed systematic analysis of the modelling of bank customer behaviour can help banking institutions take the right steps to increase their customers' activity.

1. Introduction

The banking industry is known for its inventive information systems and innovative technologies. Recent improvements in information technology and the rise of the Internet have helped banks collect and use vast amounts of data. As bank managements have taken on these tasks, they have gradually realized that information and data are not the same thing. Data need to be analyzed and extracted before they can be converted into meaningful information that can be used in making decisions.

One of the most important elements of bank customer classification models is feature engineering, which is the process of transforming raw data into new features and selecting the best features for improving the performance of classification methods. This often overlooked aspect is becoming increasingly important, especially in the contemporary world of financial uncertainty caused by the pandemic. In fact, previous classification models (Liu et al., 2022; Yuan et al., 2022; Aslam et al., 2022) tended to neglect the importance of feature engineering.

☆ This work has been supported by the scientific research project of the Czech Sciences Foundation Grant No. 19-15498S.

* Corresponding author.
E-mail addresses: m.abedin@tees.ac.uk (M.Z. Abedin), petr.hajek@upce.cz (P. Hajek), t.sharif@newman.ac.uk (T. Sharif), shahriar.setu@gmail.com (M.S. Satu), imrankhan770707@gmail.com (M.I. Khan).

https://doi.org/10.1016/j.ribaf.2023.101913
Received 28 January 2022; Received in revised form 13 October 2022; Accepted 27 February 2023
Available online 1 March 2023

Recently, various knowledge mining techniques have been used in industries to uncover hidden information about customer behaviour (Kalaivani and Sumathi, 2019; De Caigny et al., 2020; Jain et al., 2021; Chen et al., 2021; Alam et al., 2021). Indeed, knowledge discovery has helped in decision making in many research areas, such as finance, marketing, computer science, psychology, and medicine. Knowledge mining can influence banking services so that they become more acceptable and reliable for customers (Clerkin and Hanson, 2021; Berggrun et al., 2020). However, most businesses are still struggling to find the best data mining method that will lead to better services based on customer behaviour.

Academic experts are developing different machine learning-based classification methods for mining knowledge about bank customers so they can identify valuable customers and prevent customer churn in a timely manner. For example, Amin et al. (2019) used the Naïve Bayes (NB) method to predict customer churn while considering the uncertainty of data. Other machine learning methods, such as the support vector machine (SVM) and random forest (RF) methods, showed even more promising results in predicting bank customers' behaviour in the credit market (Liu et al., 2022). Recent empirical evidence suggested that feature selection is a critical pre-processing step when one is developing a bank customer classification model (Yuan et al., 2022; Aslam et al., 2022). However, far too little attention has been paid to feature engineering, one of the key steps in knowledge mining. This is surprising given that previous studies in finance have shown that feature engineering significantly affected the prediction accuracy of classification methods (Bahnsen et al., 2016; Long et al., 2019; Abedin et al., 2020; Zhang et al., 2021a,b). Therefore, the goal of this paper is to predict bank customers' activity. This is done by examining their operational behaviour at different stages of the data mining process and with different approaches to the process. Note that this study uses the concepts of knowledge mining and data mining as alternatives to each other. In contrast to existing studies that limited their data pre-processing step to feature selection (Keramati et al., 2016; Yuan et al., 2022), this study also considers feature transformation methods. More importantly, unlike previous researchers, the current study is not restricted to a single method of feature transformation or feature selection. Instead, this study uses an empirical comparison of multiple methods to provide banks with an effective combination of feature engineering approaches for processing customer behaviour data.

This study proposes a bank customer classification model enhanced with a feature engineering process. The importance of feature engineering is experimentally validated in predicting bank customer behaviour. The proposed model is divided into three stages, namely, feature transformation, feature selection, and machine learning-based classification. This study uses five approaches to feature transformation, namely, the sine, logarithm, min–max, Z-score, and cosine methods. This produces five transformed datasets. Then, five feature selection methods, namely, the correlation, chi-square, Gini index, reliefF and T-score methods, are applied to the primary and transformed datasets. Several classifiers are trained to evaluate the transformation and feature selection approaches. The machine learning approaches used are the decision tree (DT), extreme learning machine (ELM), gradient boosting (GB), k-nearest neighbour (k-NN), logistic regression (LR), multilayer perceptron (MLP), NB, RF, SVM, and xgboost classifier (XGB) methods. This study uses multiple classifiers in the decision support system because their performances vary depending on the data transformation method used and data dimensionality. To validate the proposed methodology, we use a benchmark customer behaviour dataset.

The contributions of this study relative to other studies are that several feature engineering approaches are compared, along with machine learning classifiers, allowing us to construct a knowledge mining system that combines appropriate feature transformations with state-of-the-art machine learning methods. Furthermore, this study contributes to the literature by performing a comparative analysis of different feature transformation and feature selection methods in modelling bank customer behaviour. Thus, our study sheds light on the role of feature engineering in bank customer classification. Finally, practical insights are provided for constructing an accurate yet easy-to-interpret model predicting bank customer activity; this may help bank managers to identify customers who are leaving and their counterparts who are not leaving their institutions.

The results of this study indicate that a knowledge mining system based on the combination of the Z-transformation, ReliefF feature selection, and RF methods offers an optimal decision support system for mining customer behaviour patterns. The findings of this study have managerial and academic implications. Firstly, stakeholders in the banking industry will be able to minimize the customer churn rate, and this will increase their profitability. In addition, the findings of this study will help bankers to ensure customer satisfaction and operational efficiency. Academic experts will find a comparative study of data mining methods that can help analyze customer behaviour data in various business areas.

This paper is structured as follows. Section 2 presents the review of related literature. Section 3 outlines the stages of modelling bank customer behaviour. It describes data transformation and feature selection techniques, along with classification methods. Section 4 highlights the experimental design. Section 5 presents the experimental results. Finally, Section 6 concludes the paper and presents future research directions.

2. Literature review

There is already a large amount of data stored in banks, and it is constantly growing at an enormous rate. The main obstacle encountered during this study was converting this huge amount of data into a learned information and competitive intelligence database. Several studies have used bank customer data to identify behavioural patterns using various knowledge mining techniques.

Mujica et al. (2002) proposed a case-based reasoning system for analyzing data on bank customer behaviour. In that system, cases were extracted using self-organizing feature maps trained in an unsupervised manner, and this led to a simple and efficient system. However, this approach can easily result in inaccuracies owing to the difficulty of adapting existing cases in the database to new customers. Wojnarski (2002) examined the same dataset using a local transfer function classifier to identify active and inactive bank customers. Reception fields in the hidden layer of the supervised neural network architecture were used to extract behavioural patterns from the database. However, this model was reportedly prone to overfitting, with a poor performance on one class of the

testing data. Baumann et al. (2007) used 1951 retail bank customer records and developed different models to predict the share of wallet (SOW) for deposits, debts, and loans. However, these models were based only on LR, which does not allow for the extraction of nonlinear relationships between the input and output attributes that characterize bank customers. Ngai et al. (2009) reviewed the categorization of previous studies based on data mining methods for different customer relationship management tasks. Clustering methods were predominantly used for customer identification, while classification methods were predominantly used for customer acquisition and retention.

Keramati et al. (2016) examined the customer churn dataset using data mining techniques. More precisely, a DT was used to identify customer churn in electronic banking services. The backward feature selection method was used to eliminate irrelevant input attributes. However, only the career attribute was found to be redundant in the dataset. Fejza et al. (2017) proposed a model of consumer behaviour in banking that incorporated consumer theory. Significant factors of bank selection were found by using a multivariate regression model, which provided a strong theoretical underpinning for the following studies.

A machine learning model was proposed by Rahman and Khan (2018) to detect customer behaviour by analyzing the University of California Irvine (UCI) bank marketing dataset. DT, MLP, and k-NN were used for that purpose, with k-NN outperforming the other two methods in terms of prediction accuracy. Kalaivani and Sumathi (2019) examined 1.5 million exiting customers using the DT, MLP, SVM, NB, and LR methods to take strategic steps to prevent customer churn. To reduce the data dimensionality, factor analysis and principle component analysis were used as traditional feature extraction techniques. However, compared with feature selection techniques, their approach led to a loss of potentially relevant information and a reduction in the interpretability of the features. To identify valued bank customers, Zhou et al. (2019) proposed a machine learning model where MLP and association rule mining methods were used to analyze the behaviour of bank customers. The neural network-based method was more accurate in terms of classification due to its capacity to model non-linear behavioural relationships in the data. Raju and Dhandayudam (2018) also analyzed the labelled bank dataset using the NB, DT, and MLP methods to investigate the behaviour patterns for effective customer relationship management. In this case, the DT performed best, and this can be attributed to its ability to deal effectively with categorical variables. A customer relationship management system was developed by Chen et al. (2021) to support bank decision making. Again, high system scalability and maintainability were achieved by employing the DT to exploit data on bank consumer habits. Abbasimehr and Shabani (2019) used time-series clustering to identify four categories of bank business customers, namely, high-valued customers, middle-valued customers, prone-to-churn customers, and churners. Additional profit was obtained and potentially relevant information was extracted from customer textual documents to prepare a successful customer retention campaign (De Caigny et al., 2020). Ho et al. (2019) proposed a machine learning model of customer behaviour to ensure data security for bank customers, and this suggested further opportunities for using data on the bank customer behaviour.

To summarize the above literature, the research to date has tended to focus on machine learning-based prediction methods rather than on how to effectively perform feature engineering tasks. Specifically, existing approaches rely heavily on the relevance and scales of input attributes in the original datasets. However, data characteristics can significantly affect the complexity of problems in prediction caused by overlaps in the feature values, the separability of classes' distributions, and the curse of dimensionality (Charte et al., 2022). The research methodology proposed in this study overcomes this problem by thoroughly examining the effects of feature engineering techniques, including feature transformation and feature selection, on the prediction performance of various machine learning classification methods. To achieve high accuracy, classification methods need appropriate representations for their input data, and data transformations make such data more adequate for the given task.

3. Model of bank customer behaviour

In this section, we briefly describe several feature transformation, feature selection, and classification techniques and their parameters that were used in the proposed integrated data mining model investigating the behaviour of bank customers in this study.

3.1. Feature transformation methods

Various feature transformation techniques (Dong and Liu, 2018; Abedin et al., 2020) were used for an analysis of customer behaviour in this study to increase the information value of the input attributes and thus improve the classification performance of machine learning methods. The feature transformation methods were applied to the original (primary) customer dataset, that is, before the feature selection and classification techniques were employed. Table 1 briefly describes the methods used in this study.

One of the main problems with most data mining tasks is that differences in the scales of the input features can increase the difficulties of the task being modelled. A model with different scales is likely to be unstable, and this means that it may suffer from a poor generalization performance. Normalization (Han et al., 2011) requires being able to precisely estimate the minimum and maximum values of the features, and this can be problematic if one uses only training data. Moreover, the shape of the feature distribution remains unchanged after normalization. One standard method of dealing with situations where one feature has much more variance than the others is to use standardization (Han et al., 2011). However, when the feature does not follow a linear distribution, a nonlinear transformation such as log transformation (Akter et al., 2019) is preferable because it minimizes skewness and maps the underlying distribution close to a normal distribution. Finally, sine and cosine transformations scale the data to unify their amplitudes (Vidal and Kristjanpoller, 2020).

Table 1
Definition of feature transformation techniques used in this study. The original and transformed features are represented by x and y, respectively, and $\mu(x)$ and $\sigma(x)$ are the average and standard deviation values of x, respectively.

Method	Description	Equation
Log transformation	A quantitative operation that converts feature values by applying natural logarithm to increase the likelihood and fit of machine learning methods and transforming skewed data to an approximately normal distribution (Akter et al., 2019).	$y = clog_b(1 + x)$
Min–max normalization	Scaled transformation prevents attributes with larger scales from biasing the values of the objective functions and speeds up the convergence of the gradient descent-based algorithms. However, noise in the data can be increased (Han et al., 2011).	$y = \frac{x - min(x)}{\max(x) - min(x)}$
Z-score normalization	A common scale for the attribute is produced, thus reducing the impact of outliers. For attributes with small variance, the problem of noise in the data can be exacerbated.	$Z - score = \frac{x - \mu(x)}{\sigma(x)}$
Sine transformation	Feature rescaled to the range [−1, 1]. The amplitude of the data is 1. Allows working with negative attribute values.	$y = \sin(x)$
Cosine transformation	The amplitude of the data can be analyzed as a function of the data frequency and faster convergence can be reached.	$y = \cos(x)$

Table 2
Summary of the used feature selection techniques.

Method	Description
Correlation-based filter	Both class relevance (correlation between the feature and class attribute) and feature inter-dependency (correlations between each feature pair) are considered (Hall, 2000).
Chi-square	Features are selected according to the values of the chi-squared statistic testing the dependency between an input attribute and target (class) attribute (Han et al., 2011).
Gini index	Evaluates the feature purity (discriminative power) with respect to the class attribute (Tan et al., 2016).
ReliefF	ReliefF (Urbanowicz et al., 2018) evaluates individual features by iterated sampling where a feature score is measured by selecting nearest instances from different classes. This method is noise-tolerant but is not able to discriminate redundant features.
T-score	T-score (Chandra and Gupta, 2011) is based on an evaluation (T-statistic) of each feature using the sample size, mean, and standard deviation values of the features with respect to each class.

3.2. Feature selection methods

Different features contributed at different levels in the predictive model, with some features being more relevant and less redundant (Cai et al., 2018). To identify the features with high predictive capacity and filter out irrelevant features, we used five traditional feature selection techniques. These feature selection techniques, which are briefly listed in Table 2.

The correlation-based filter (Hall, 2000) selects features that are relevant to the class attribute, and it discards redundant features using an entropy-based measure. The chi-square method (Han et al., 2011) tests whether the class feature is independent from the given independent feature. The Gini index (Tan et al., 2016) selects features based on their capacity to distinguish data instances from different classes. It is noteworthy that this method is often used to generate DT-based classifiers. The ReliefF method (Urbanowicz et al., 2018) utilizes the Relief algorithm to select features based on their feature scores, which are calculated using the nearest samples of the same and different classes. The T-score method (Chandra and Gupta, 2011) is based on the popular Student's paired T-test comparing the samples of two classes.

3.3. Classification techniques

After implementing feature transformation and feature selection, diverse classification methods were used to predict customer behavioural patterns. These methods were adopted due to their successful application in related bank customer classification problems (de Lima Lemos et al., 2022; Kinge et al., 2022; Papouskova and Hajek, 2019; Bhatore et al., 2020).

1. **Decision Tree (DT):** The DT (Song and Ying, 2015) is a flowchart structured like a tree in which each internal node denotes a condition and each branch denotes the outcome of that attribute.
2. **Gradient Boosting (GB):** Gradient boosting (Bentéjac et al., 2021) involves an iterative process whereby base classifiers target hard-to-predict instances to change their distribution, and this assigns a weight to each training instance and adjusts the weight at the end of each round of boosting by approximating the probabilistic distribution $P(x)$ as follows:

$$P(x) = \sum_{k=0}^{K} \beta_k w\left(x, a_k\right), \tag{1}$$

where $P(x)$ is a weighted sum of w functions representing K weak learners.
3. **Extreme Learning Machine (ELM):** The ELM (Sattar et al., 2019) is a feed-forward neural network with a single hidden layer. In contrast to gradient-based methods, ELM assigns random values to the weights between the neurons of the input and hidden layers, allowing the learning algorithm to converge faster. The output of the ELM can be defined as:

$$f(x) = \sum_{i=0}^{M} \beta_i G\left(a_i, b_i, x\right) \tag{2}$$

where $G(.)$ represents the activation function of the ith neuron in the hidden layer.
4. k-**Nearest Neighbour** (k-**NN**): k-NN (Han et al., 2011; Satu et al., 2017) is a simple, straightforward, and non-parametric classification algorithm in which a given vector is explored by using k surrounding training instances in the original data space. The euclidean distance between instances X_1 and X_2 is typically used to find the nearest neighbours:

$$\text{dist}\left(X_1, X_2\right) = \sqrt{\sum_{i=1}^{n} \left(x_{1i} - x_{2i}\right)^2}. \tag{3}$$

5. **Logistic Regression (LR):** LR (Abedin et al., 2019a,b) is a statistical method for investigating the relationship between multiple independent attributes and a dependent class attribute. A logistic function is used to model this relationship as follows:

$$g(x) = \ln\left[\frac{\pi(x)}{1-\pi(x)}\right] = \ln\left[\frac{P(y=1|x)}{P(y=0|x)}\right] = \beta + \sum_{i=1}^{p} \beta_i x_i. \tag{4}$$

6. **Multi-Layer Perceptron (MLP):** The MLP (Abedin et al., 2019a,b) is a finite acyclic graph in which the nodes represent neurons with sigmoid activation functions and the edges denote synapses that pass the signal from the input layer through the hidden layer to the output neurons. The output of a perceptron a_i is calculated as follows:

$$net_0 \leftarrow w + \sum_{j \in \text{Pred}(i)} w_{ij} a_j, \tag{5}$$

$$a_i \leftarrow f_{log}\left(net_0\right), \tag{6}$$

where net_0 is the input potential and w are synapse weights.
7. **Naïve Bayes (NB):** NB (Han et al., 2011) is a probabilistic classifier based on the Bayes theorem that assumes independence among input attributes (predictors). The probabilities of classes C_i given input attributes X are obtained as follows:

$$P\left(C_i|X\right) = \frac{P\left(X|C_i\right) P\left(C_i\right)}{P(X)}. \tag{7}$$

8. **Random Forest (RF):** The RF (Tan et al., 2016) is a class of ensemble methods that combines the predictions of multiple decision trees. Each decision tree uses a random vector of features with a fixed probability distribution. Base learners are trained on diverse training data, which is ensured by using the bootstrap aggregating (bagging) technique.
9. **Support Vector Machine (SVM):** The SVM (Moula et al., 2017; Abedin et al., 2019b; Zhang et al., 2021a,b) is a supervised learning technique that works well with high-dimensional data and avoids the curse of dimensionality by, at the same time, minimizing residuals and maximizing the margin hyperplane as follows:

$$Y(x) = \text{sign}\left(\sum_{i=0}^{n} y_i \alpha_i K\left(x_i, x_j\right) + b\right), \tag{8}$$

where K is a kernel function between two input vectors x_i and x_j.

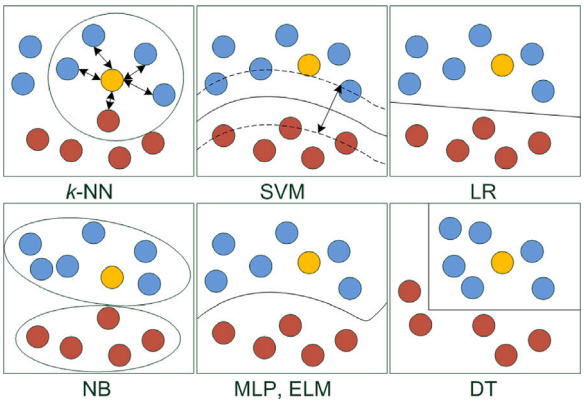

Fig. 1. Decision boundaries of the methods used (k-nearest neighbour (k-NN), support vector machine (SVM), logistic regression (LR), Naïve Bayes (NB), multi-layer perceptron (MLP), extreme learning machine (ELM), and decision tree (DT)). This figure illustrates the different decision boundaries of the classifiers used.

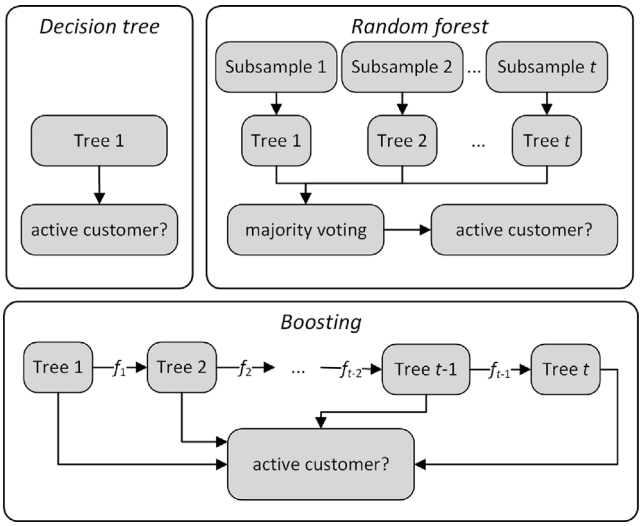

Fig. 2. Flowcharts of the single decision tree, random forest, and boosting methods for bank customer classification. For boosting, $f_1, f_2, \ldots, f_{t-1}$ are the models of additive decision trees. This figure shows how the different tree-based classifiers decide the final customer class.

10. **Extreme Gradient Boosting (XGB):** XGB (Chen and Guestrin, 2016) is an optimized, flexible, efficient, portable, and distributed gradient boosting library where boosted trees are generated in a stepwise fashion by using extra randomization and regularization to reduce their variance and improve their robustness to overfitting, respectively.

To facilitate the differences between the above methods, Fig. 1 illustrates their decision boundaries and Fig. 2 shows the differences between single decision trees and ensemble learning methods, that is, the RF and boosting methods (GB and XGB).

To train the classifiers, their parameters must be set. Using the settings recommended in the above references, the values of the training parameters were determined as shown in Table 3. To fix the high number of hyperparameters, we relied on trial-and-error optimization for most of the hyperparameters, and the grid search method was only used to fix the number of nodes in the ELM (from the range of 10 to 1000).

4. Experimental design

In this section, we explain the data collection process and demonstrate the analysis of the performance by using the confusion matrix and calculating the accuracy, precision, and F-measure.

4.1. Data collection

This study examined a dataset of bank customer behaviour. Our model used various feature transformations, feature selections, and classifiers from the scikit-learn machine learning library. We collected the dataset from the 2nd EUNITE Competition for

Table 3
Machine learning approaches used in this study and the settings of their hyperparameters. To achieve the optimal classification performance, the values of hyperparameters were determined by trial and error.

Method	Parameters
LR	cut-off point = 0.5
DT	C4.5 algorithm, min. no. of instances per leaf = 2, confidence factor for pruning = 2
NB	batch size = 100
MLP	one hidden layer with # features/2 nodes, learning rate = 0.3, iterations = 500
k-NN	k (nearest neighbours) = 5, Euclidean distance
ELM	one hidden layer with 80 nodes, activation function = inv_multiquadric
SVM	radial basis function kernel, complexity C = 1
GB	default booster, learning rate = 0.01, maximum depth = 0.5
XGB	learning rate = 0.1, Gamma = 0, depth of trees = 6, iterations = 200
RF	100 trees, splits features = 5, maximum tree depth = 25

modelling bank customer behaviour data (Mujica et al., 2002; Wojnarski, 2002). The participants of the competition were 143 people from 33 countries; they shared their results, with at least 15 participants submitting their results in a report. The competition was formally associated with the Tatra Banka banking house in Slovakia. Individuals in the banking industry need these kinds of applications to be able to do research on and identify significant factors that will allow them to predict whether a customer is active or not. If that customer is not active, several proactive measures need to be taken to retain the customer and not let the customer go to another bank. The dataset contained 12,000 training samples and 12,000 testing samples (bank customers), which were characterized using 36 attributes, the first 6 of which were nominal and the next 30 of which were numerical. The 37th attribute denoted the output class, where class 0 represented an inactive customer and class 1 denoted an active customer. It is also worth noting that the dataset was balanced, so it was not necessary to use a class-balancing algorithm in this study.

The objective of this study was to find an appropriate representation of bank customer behaviour data to detect with high accuracy the behavioural patterns leading to customer engagement in the bank's business (i.e., active customers with a non-zero account balance) and to distinguish them from the patterns of inactive customers (i.e., customers with a low account balance and who did not use other banking products). Recognizing the customer status can help a bank predict customer churn early and identify common patterns for customer activation.

4.2. Proposed methodology

The proposed knowledge mining model for bank customer analysis is depicted in Fig. 3.

- First, we collected and examined a set of primary data on customer behaviour in the banking industry using different features representing risk factors related to customer patterns and demands. Then, five feature transformation techniques (Dong and Liu, 2018; Abedin et al., 2020), namely logarithm, min–max, Z-score, sine, and cosine methods, were implemented to transform the data into several datasets with different statistical characteristics.
- Then, we applied five feature selection methods (Cai et al., 2018), that is, the correlation-based filter, chi-square, Gini index, ReliefF, and T-score techniques, to the primary and other transformed datasets. After the feature selection procedure, five feature subsets were generated from each dataset.
- Next, in agreement with the existing literature (de Lima Lemos et al., 2022; Kinge et al., 2022; Papouskova and Hajek, 2019; Bhatore et al., 2020), we employed the prominent classification techniques (DT, ELM, GB, k-NN, LR, MLP, NB, RF, SVM, and XGB) to classify the customer behaviour dataset.
- Following the recommendations for statistical comparisons of multiple classifiers (Demšar, 2006), different statistical tests, such as the Friedman test and Wilcoxon signed-rank test, were performed to compare the obtained results.

4.3. Performance analysis

To analyze the performance of the proposed model, we created a confusion matrix that represented the summary of the classification results. The number of correct predictions (i.e., true positive [tp] and true negative [tn]) and incorrect predictions (i.e., false positive [fp] and false negative [fn]) is summarized using counted values and broken down by each class. The confusion matrix shows how the classification model is confounded in making predictions (Keramati et al., 2016). Here, this matrix gave us

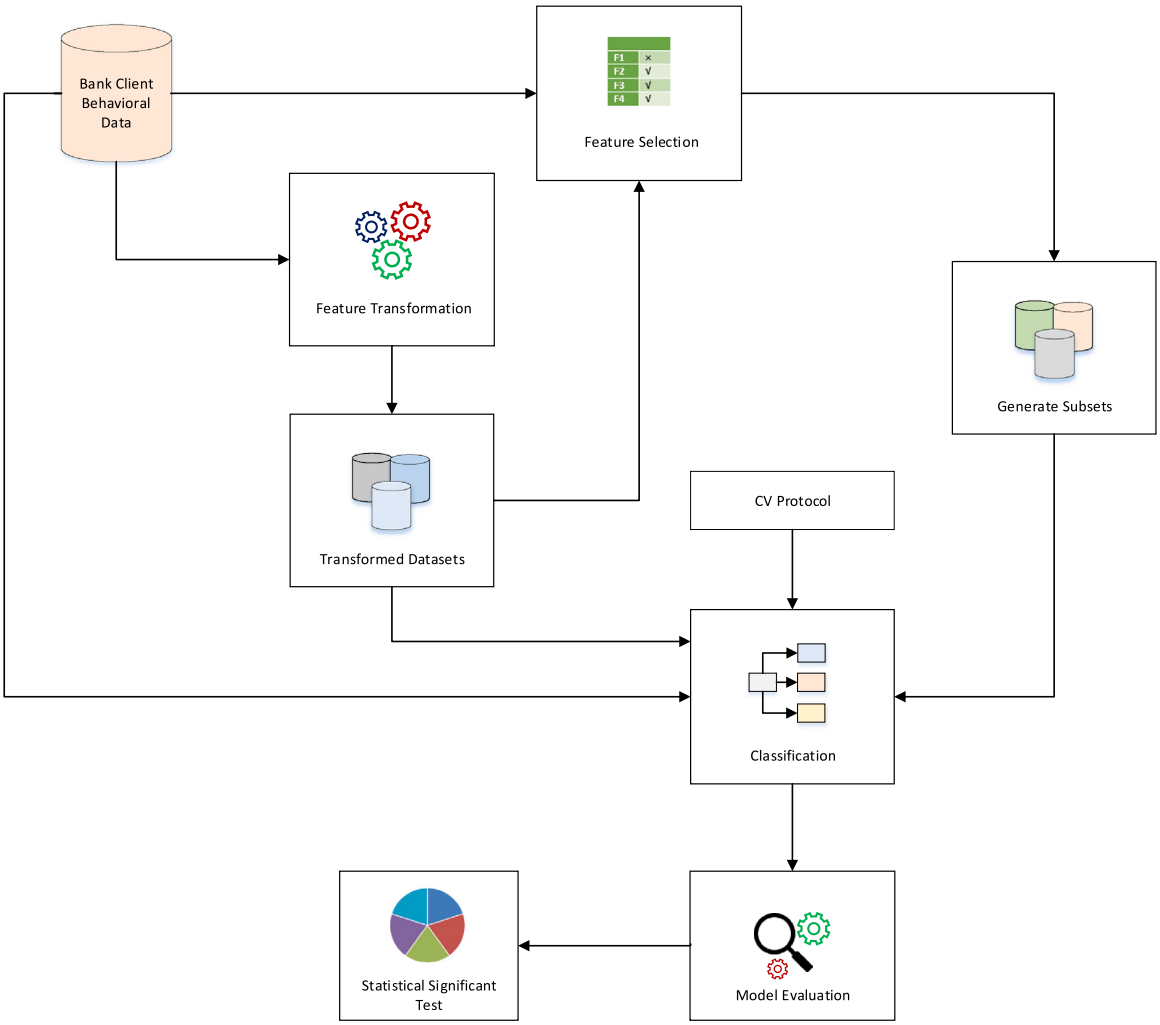

Fig. 3. Proposed knowledge mining methodology. Different experimental scenarios were considered by using primary or transformed data combined with different feature selection methods and classifiers.

Table 4
Confusion matrix for the binary classification problem. Positive and negative observations represent active and inactive bank customers, respectively, and the tp, tn, fp, and fn are the numbers of true positive, true negative, false positive, and false negative observations, respectively.

		Predicted observation	
		Positive	Negative
Actual observation	Positive	tp	fp
	Negative	fn	tn

an overview not only of the errors that the classification algorithm made, but, more importantly, the types of errors it made (see Table 4).

We also measured the accuracy of the classifiers (also known as the classification rate) using the following equation:

$$Accuracy = \frac{tp + tn}{(tp + tn + fp + fn)}. \tag{9}$$

Recall can be defined as the ratio of the total number of correctly classified positive instances divided by the total number of positive instances, where high recall indicates that the given class is correctly recognized:

$$Recall = \frac{tp}{(tp + fn)}. \tag{10}$$

Table 5
Descriptive statistics of attributes. The T-test was performed to compare the mean values of the two classes. This table reports the mean, standard deviation, and T-test results for each attribute.

Attribute	Mean	Std. deviation	T-test	Attribute	Mean	Std. deviation	T-test
A0	0.8233	0.3815	334.3357***	A18	0.4298	0.0108	6174.8926***
A1	6.9630	0.3410	3163.6041***	A19	0.0001	0.0065	1.7644*
A2	53.4017	5.6608	1461.4434***	A20	0.4278	0.0087	7586.4986***
A3	18.5323	1.2006	2391.2567***	A21	0.0002	0.0065	4.0939***
A4	64.7090	2.9309	3420.2957***	A22	0.9240	0.0060	23727.5769***
A5	28.9604	0.7136	6286.7882***	A23	0.9577	0.0062	23900.6556***
A6	0.0183	0.0116	244.8197***	A24	0.2368	0.0063	5790.7611***
A7	0.0133	0.0192	107.4377***	A25	0.1034	0.0058	2750.3757***
A8	0.0017	0.0159	16.0771***	A26	0.2656	0.0079	5206.9334***
A9	0.0158	0.0483	50.6201***	A27	0.1034	0.0058	2750.3741***
A10	0.0006	0.0096	10.1064***	A28	0.9723	0.0063	23960.3039***
A11	0.0000	0.0065	1.0999	A29	0.9822	0.0067	22549.9192***
A12	0.0009	0.0082	16.0480***	A30	0.0950	0.0072	2048.0339***
A13	0.0000	0.0065	1.1001	A31	0.0059	0.0064	141.4787***
A14	0.0000	0.0066	1.1722	A32	0.1743	0.0060	4526.9265***
A15	0.0000	0.0065	1.0000	A33	0.0047	0.0064	114.3843***
A16	0.9999	0.0068	22614.9683***	A34	0.8776	0.0057	23760.2040***
A17	0.0016	0.0173	13.9526***	A35	0.9215	0.0061	23572.5233***

Note: Significant at 1% (***), 5% (**), 10% (*).

To get the value of precision, we divided the total number of correctly classified positive instances by the total number of predicted positive instances. High precision indicates that an instance labelled as positive is indeed positive:

$$Precision = \frac{tp}{(tp + fp)}. \tag{11}$$

Since we have two measures, precision and recall, the F-measure could be computed as their harmonic mean as follows:

$$F - measure = \frac{2 \times precision \times recall}{(precision + recall)}. \tag{12}$$

The area under the receiver operating characteristic curve (ROC) represents the probability that a bank customer classification model ranks a randomly selected inactive customer higher than a randomly chosen active customer (Papouskova and Hajek, 2019). It can be defined as follows:

$$Area\ under\ ROC = \int_0^1 Recall(T) \times \frac{d}{dT} fpr(T) dT, \tag{13}$$

where fpr is the fp rate and T is the cut-off point.

The Matthews correlation coefficient (MCC) is a preferable evaluation measure for binary classification problems (Theodoridis and Tsadiras, 2022). Like the area under the ROC, the MCC is informative for imbalanced classifications:

$$\text{MCC} = \frac{tp \times tn - fp \times fn}{\sqrt{(tp + fp) \times (tp + fn) \times (tn + fp) \times (tn + fn)}}. \tag{14}$$

5. Experimental results and discussion

In this section, we discuss the results of the analysis of bank customers' behaviour. First, we examined the statistical characteristics of the bank customer intelligence data.

5.1. Descriptive statistics

To obtain descriptive statistics, data are summarized in Table 5. The values of the instances can then be evaluated by examining the mean values for each attribute. Through the standard deviation, the average distance between the values of each attribute and the mean value is measured. A low standard deviation indicates that the data points tend to be close to the mean of the data, while a high standard deviation indicates that the data points are spread over a wider range of values.

5.2. Correlation and risk factor analysis

By using the correlation process, we find the correlation coefficient r for the analysis of the result between different attributes. The more closely the r is ranged into +1 or −1, the more closely two attributes are related. If the number is close to 0, it means there is no relationship between the attributes. If the correlation result is positive, it means that if one attribute increases, the other also increases. If the correlation result is negative, it means that as one attribute gets larger, the other gets smaller (i.e., an inverse

Table 6
Results of the Kaiser–Meyer–Olkin (KMO) and Bartlett's tests. The KMO test was conducted to test for sampling adequacy, and the Bartlett's test was performed to demonstrate that the correlation matrix was not an identity matrix. Both tests confirmed that the data were suitable for factor analysis.

Kaiser–Meyer–Olkin measure of sampling adequacy		0.688
Bartlett's test of sphericity	Approx. Chi-Square	3304884.966
	Df	630
	Sig.	0.000

correlation). A table of correlation results can be found in Tables 1 and 2 in the Appendix, and according to those tables, attribute A13 provided the best correlation result for the output attribute.

We further used the Kaiser–Meyer–Olkin (KMO) test, which measures the suitability of the dataset for factor analysis, and the Bartlett's test of sphericity, which compares an observed correlation matrix with the identity matrix and also determines whether there is a redundancy between the attributes that can be summarized with a few factors. The outcomes of the KMO (i.e., sampling adequacy) and Bartlett's test (i.e., sphericity) are shown in Table 6.

Using the correlation analysis, we first calculated the correlation coefficient r to analyze the relationships between input attributes. The closer r is to +1 or −1, the more closely the two attributes are related. If it is close to 0, it means that there is no relationship between the attributes. If the correlation result is positive, it means that if one attribute increases, the other will also increase and vice versa. A table with the correlation results is attached in the Appendix (see Table 1 in the Appendix). We further used the Kaiser–Meyer–Olkin (KMO) test, which is a procedure for measuring the suitability of a dataset for factor analysis. The Bartlett's test of sphericity, on the other hand, compares the observed correlation matrix with the identity matrix. In addition, it determines whether there is some redundancy between attributes that can be summarized by a few factors. To measure the results of KMO and Bartlett's test, we report a correlation in Table 6, which shows the results of KMO for sampling adequacy and Bartlett's test of sphericity.

Again, we analyzed the risk factors for all 36 attributes by calculating the eigenvalues, variance factors, and cumulative outcome (see Table 2 in the Appendix). We then measured the skewness to measure the lack of symmetry and the kurtosis to detect any high or light tails relative to the normal distribution in the dataset. This analysis was performed with six different types of feature transformation techniques. The results of this analysis have been appended for better understanding.

5.3. Classification results

5.3.1. Performance analysis of different classifiers

Different classifiers were used for the primary customer data, and they were later transformed and their feature subsets were selected. When examining the performance for the primary dataset, the RF had the highest accuracy, at 75.39%, of all the classifiers and feature selection subsets. For the feature selection performed on the primary data, good results were reported most frequently for the XGB and GB methods. Regarding the features selection methods, the highest average accuracy, at 65.36%, across classifiers for the primary data was for the ReliefF method.

For the log-transformed dataset, the RF had the highest result, at 75.44%, against all classifiers and feature subsets, while the XGB method ranked second with the average accuracy at 72.67%. For the dataset normalized using the min–max method, the RF outperformed the other methods with an accuracy of 75.52%. The XGB and GB methods also performed well for the other feature subsets. In addition, all classifiers performed best for the chi-squared feature selection method. In the Z-normalized dataset, the RF had the highest accuracy of 75.85%, while the GB and XGB performed best for the other feature subsets. Under these circumstances, all classifiers embodied the best average performance, at 68.72%, for the Gini index.

Furthermore, the RF performed best also for the sine-transformed dataset, with an accuracy of 75.75%, while the GB and XGB produced the best results for this transformation with the feature selection methods. For the cosine-transformed dataset, the XGB exhibited the highest accuracy, at 73.03%, among all the classifiers and feature selection methods. However, the k-NN and GB showed better average results for this feature transformation. All of the classifiers had the highest average accuracy of 61.84% for the subset of the Gini index features. These results are shown in Tables 4 to 9 in the Appendix.

This observation implied that most of the classifiers did not perform well regardless of the feature engineering techniques used. The NB had the accuracy of almost 57% only, and the SVM had an accuracy of about 53%. Moreover, the LR, ELM, and MLP had an accuracy of almost 61%, 63%, and 66%, respectively. Overall, the RF, XGB, and GB performed best for the primary and transformed datasets.

Without feature selection, the highest accuracy of the classifiers was for the primary and all of the transformed datasets. The RF had the best results for all the primary and transformed datasets except for the cosine-transformed dataset. The Z-transformed dataset had an accuracy of 75.85%, which was the best result in this experiment. The RF showed no highest average performance for all the transformed datasets, whereas the XGB and GB showed the best results for the different subsets of features. When examining the frequencies of the best results of the different classifiers, the RF, XGB, GB, and k-NN had the best results in 6, 17, 11, and 2 cases for the primary and transformed datasets and their subsets of features, respectively. The RF did not achieve good results for the feature subsets compared with the GB, XGB, and k-NN. In most cases, the XGB outperformed the other candidate classifiers. Therefore, the XGB was the classifier that usually generated the best performance.

Table 7
Feature transformation performance evaluation and post-hoc analysis. Iman-Davenport modification of Friedman's test was performed to compare the classification performance over different feature transformation methods. Holm's post-hoc test was used to statistically compare the performance of the best ranked method with those of the remaining feature transformation methods.

	Accuracy		*F*-measure	
	Rank (#Position)	Holm's *P*-value	Rank (#Position)	Holm's *P*-value
Z transformation	2.1500 (#1)	———	2.3000 (#1)	———
Sine transformation	2.4500 (#2)	0.050000**	2.3500 (#2)	0.050000**
Normalized transformation	3.0500 (#3)	0.025000**	2.7500 (#3)	0.025000**
Log transformation	3.2500 (#4)	0.016667***	3.0500 (#4)	0.016667***
Raw (non-transformed) data	4.5000 (#5)	0.012500***	4.7500 (#5)	0.012500***
Cosine transformation	5.6000 (#6)	0.010000***	5.8000 (#6)	0.010000***
Iman-Davenport *P*-value	7911266E−5		6774E-7	

Note: ***significant at 1%, **significant at 5%, *significant at 10%.

Table 8
Results of Wilcoxon signed-rank test over the transformed and nontransformed datasets. The classification performance of the RF was compared with those of the other classification models.

Model X	Model Y	Accuracy		*F*-measure	
		Improvement (%)	*P*	Improvement (%)	*P*
RF	LR	14.83	0.005***	14.89	0.005***
	DT	7.42	0.005***	7.43	0.005***
	NB	21.91	0.005***	27.82	0.005***
	MLP	15.77	0.005***	15.78	0.005***
	k-NN	9.64	0.005***	9.64	0.005***
	ELM	16.62	0.005***	16.70	0.005***
	SVM	23.98	0.005***	32.35	0.005***
	GB	0.72	0.103	0.73	0.093*
	XGB	3.83	0.005***	3.82	0.005***

Note: ***significant at 1%, **significant at 5%, *significant at 10%.

5.3.2. Comparison of feature transformation and feature selection techniques

When the results of all the classifiers were averaged for the feature selection methods (including the primary data), the classifiers showed the best average result, at 65.28%, for the dataset normalized with the min–max transformation. The average of the maximum results of all the classifiers for feature selection (including the primary data) had the best results (69.26%) for the Z-normalized dataset. Thus, it can be seen that the normalized data had the best results on average. In this case, the highest maximum average classification results in the primary, transformed, and feature subsets were found for the primary customer behaviour data, at an accuracy of 73.79%. More precisely, the best result was achieved for the primary and transformed datasets using the RF classifier. In transforming the primary data using the logarithmic, min–max, Z-score, sine, and cosine methods, these data were classified using different classifiers without feature selection.

For feature subsets, the maximum average classification results for the Gini-based feature selection were demonstrated. The results based on ranking showed that attribute 13 (A13) was the most relevant feature. It had a result higher than 0.12 for the raw dataset with all five transformation techniques, while all the other attributes were much lower, with a value of approximately 0.7 for rank two. In contrast, attribute number 16 was the least relevant. In addition, if we consider the ROC performance criterion, we find that the repositioning of the Gini index was very stable and had higher accuracy along the Z-transformation. On the other hand, the normalized transformation had the best result if we found the average result. All these ranks and boxplots of ROC values are shown in Figs. 4 and 5.

For comparison purposes, we experimented with the kernel PCA (kPCA) method with different numbers of principal components; the best results were obtained for six principal components. As for the above experiments, the RF was the most accurate classifier, with an accuracy of 66.31%; this was 9.54% lower than the RF results obtained for the original data after the Z-transformation. For a fair comparison, the data were transformed using the Z-score for the kPCA. An even greater difference was observed in terms of the MCC when the RF with the kPCA reached a value of 0.326; for the original dataset it was 0.518, well above the 0 value of the random guess classifier.

5.4. Statistical tests

Based on the performance accuracy and F-measure of the different feature transformation methods shown in Table 7, the results suggest that the Z-transformation outperformed all the other transformation techniques, with the cosine transformation having the lowest accuracy with rank six.

Based on the results of the Wilcoxon signed-rank test for the total transformed and nontransformed datasets, the RF performed best. In contrast, the SVM had the weakest results based on the accuracy improvement rate and F-measure (see Table 8).

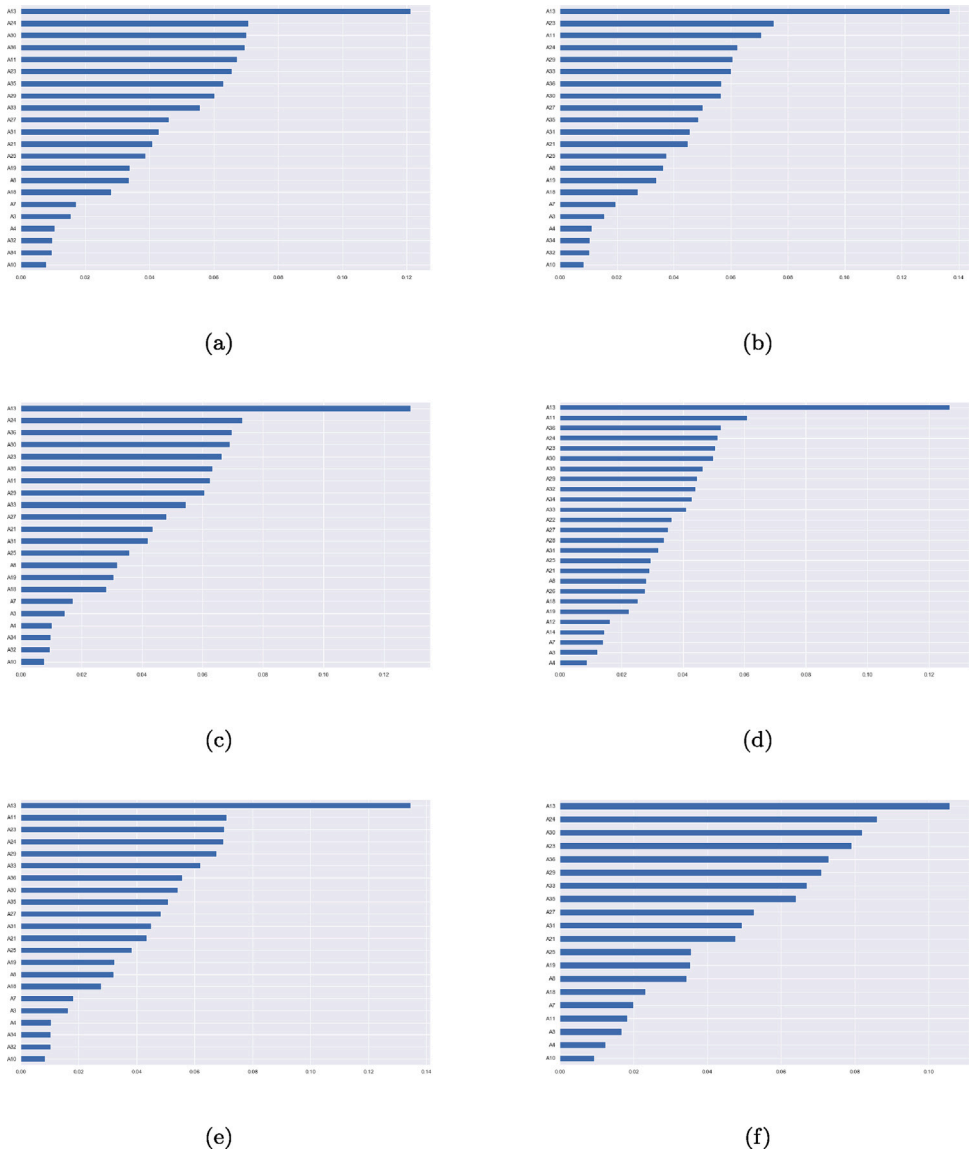

Fig. 4. Feature ranking of (a) primary data, (b) log, (c) min–max, (d) Z-score, (e) sine-transformed, and (f) cosine-transformed datasets. This figure compares the importance of the features for different feature transformation methods in terms of the Gini index.

Finally, we ranked the feature transformation methods according to five different feature selection methods using the Friedman matrix. Table 9 shows that the Gini index had the best result of all the feature selection techniques and that the Z-transformation ranked first with regard to the Gini index, while the log transformation had the second best result and ranked second with regard to the Gini index. Overall, based on the average results, the normalized transformation performed best.

5.5. Comparison with existing studies

In the field of customer behaviour analysis, only a few papers have addressed the topic of the current study. In the EUNITE Competition, the classification accuracy of the test data ranged between 60.0% and 75.5%. In this paper, a detailed and systematic analysis of how different feature engineering methods (i.e., feature transformation and feature selection) affected the results of different classifiers was presented. Factor analysis with principal component analysis and T-tests were also used to identify important features, something that was not done in previous studies (Baumann et al., 2007; Fejza et al., 2017; Rahman and Khan, 2018). This work also presented a post-hoc nonparametric statistical test to demonstrate the efficiency of the proposed data mining methodology.

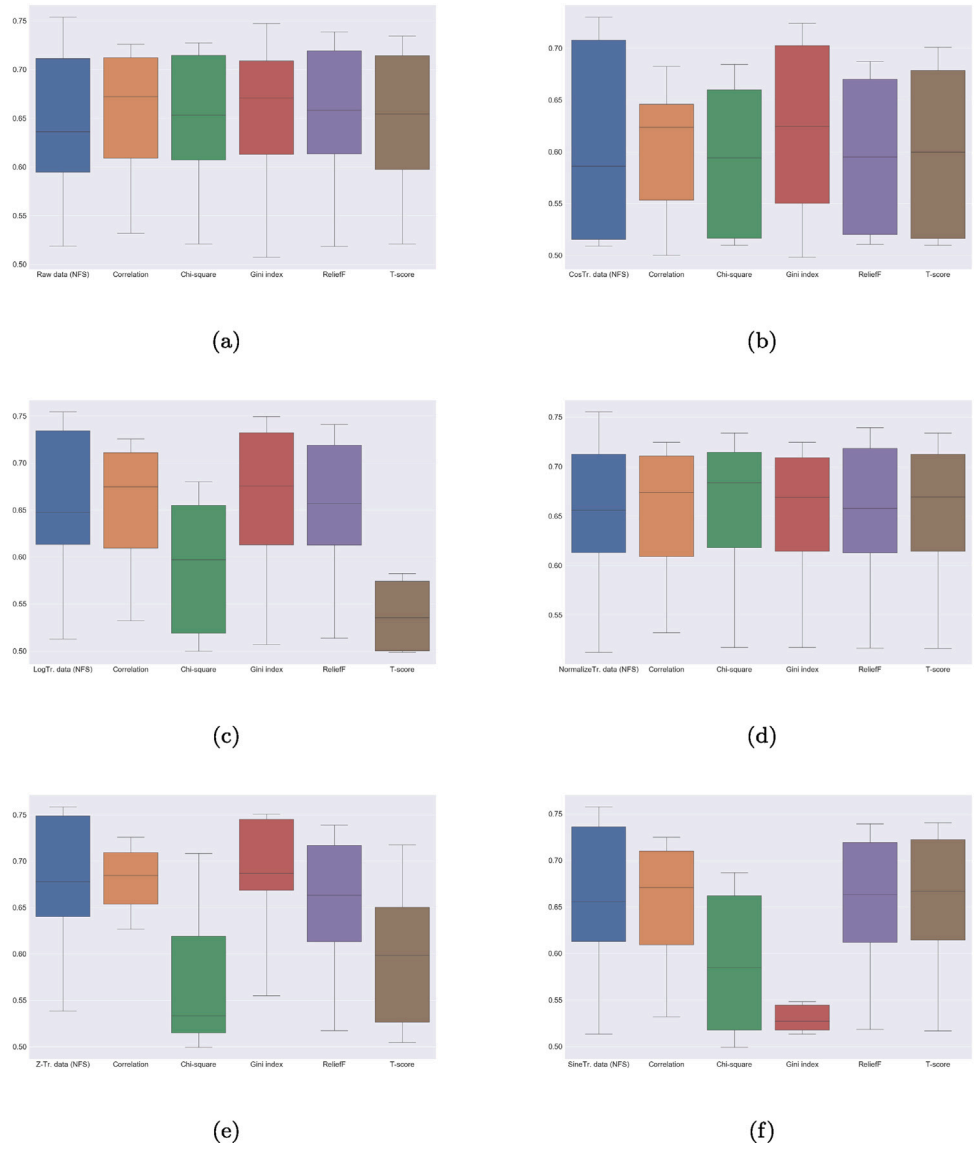

Fig. 5. Boxplots of ROC values for (a) primary data, (b) log, (c) min–max (d), Z-score, (e) sine-transformed, and (f) cosine-transformed datasets. This figure shows the stability of the classification performance on both classes for different feature selection methods.

6. Conclusions

By using the recent developments in information technology, businesses such as banks can collect data and transform them into information to make decisions about their future courses of action. Banks use various data mining techniques to learn information about customer behaviour. Currently, data mining can create a bridge between a bank officer and individual customers.

To achieve this goal, we first used 5 different feature transformation and feature selection methods along with 10 classification techniques, to identify customer activities in the banking industry. We also ranked the feature transformation and feature selection techniques to demonstrate their effects on bank customer classifications. For the benchmark bank customer dataset, the RF had the best classification accuracy, at 75.85%, for the data transformed using the Z-score normalization technique. It was interesting that the performance of most of the classifiers did not improve after applying the feature selection methods; this suggests that all the data gathered by the bank are relevant for predicting bank customer decisions. By performing a detailed systematic analysis of bank customer behaviour with different types of feature transformation, feature selection, and classification techniques, we could provide banks with a decision support tool that might enable them to achieve higher profits through the early detection of a client's intention to stop using the bank's services.

Table 9
Friedman matrix for algorithm ranking and the positions (P#) with different feature selection methods. The average rank values are reported for both the feature transformation and feature selection methods.

	Correlation	Chi-square	Gini index	ReliefF	*T*-score	Average
Z transformation	9.55	26.30	5.9	10.80	22.40	14.99
	(P#4)	(P#27)	(P#1)	(P#9)	(P#21)	(P#3)
Sine transformation	11.55	25.35	26.70	9.95	9.65	16.64
	(P#16)	(P#26)	(P#28)	(P#7)	(P#5)	(P#5)
Normalized transformation	11.40	8.70	11.80	10.80	11.15	10.77
	(P#14)	(P#3)	(P#17)	(P#9)	(P#13)	(P#1)
Log transformation	9.70	25.10	8.65	11.45	27.95	16.57
	(P#6)	(P#25)	(P#2)	(P#15)	(P#29)	(P#4)
Raw (non-transformed) data	10.90	10.85	13.10	10.25	11.00	11.22
	(P#11)	(P#10)	(P#18)	(P#8)	(P#12)	(P#2)
Cosine transformation	23.80	23.60	21.05	22.25	23.35	22.81
	(P#24)	(P#23)	(P#19)	(P#20)	(P#22)	(P#6)
Average	12.82	19.98	14.53	12.58	17.58	
	(P#2)	(P#5)	(P#3)	(P#1)	(P#4)	

There have been many studies in which various analyses have been carried out to improve services in the banking sector. Analyzing customer behaviour is one important area in this regard. However, only a few studies have assessed customer attitudes (Raju and Dhandayudam, 2018; Rahman and Khan, 2018; Zhou et al., 2019; Ho et al., 2019), and a detailed systematic analysis of them was lacking. In the current study, more analytical techniques were applied to customer behaviour data to investigate how they could more accurately predict customer activity in banking services based on customers' behavioural patterns. Therefore, feature transformation, feature selection, and machine learning-based classifiers were used in this paper. This study shows which approaches for data analysis are more suitable for exploring customer behaviour. Different feature transformation and feature selection methods performed differently in this scenario. This suggests which feature transformation and feature selection methods are responsible for contributing to the highest classification performance for identifying customer activity in banking services. In addition, the correlation and risk factor analysis indicate which data transformation, feature selection, and classification methods are the most effective for this analysis.

Another important implication for the banking sector in this study is that a detailed systemic review of customer behaviour analysis has not been conducted so far. Therefore, if any researcher or practitioner wants to analyze customer behaviour in the banking industry, our results suggest how to combine appropriate methods. Many statistical analyses of customer behaviour have been carried out in the past (Abbasimehr and Shabani, 2019; Fejza et al., 2017; Kalaivani and Sumathi, 2019). In this paper, a combination of statistical and machine learning approaches was used, and they could guide researchers to use this methodology for further analyses. We found that the appropriate feature transformation must be applied to achieve an adequate classification performance of the machine learning-based methods, indicating that suitable data representation must first be available before proceeding to the actual customer classification. Consistent with previous studies (Liu et al., 2022), we observed that ensemble-based machine learning methods, such as the RF and XGB, performed best among the benchmarked machine learning methods, including those used in prior research (Raju and Dhandayudam, 2018; Zhou et al., 2019; Karvana et al., 2019). Intriguingly, the XGB method was found to be more robust for the data representation used compared with the RF. These differences can be explained by the invariance of iterative boosting-based methods to such transformations (Hastie et al., 2009). In the case of the RF, then, poorly engineered features increase the likelihood that individual trees are trained on noise rather than on an appropriate representation of the original data (Hastie et al., 2009). Overall, the results of this study show that the revealed combinations of data mining methods allow the ensemble-based methods to achieve a comparable or better performance than the state-of-the-art bank customer classification models based on various neural network architectures (Wojnarski, 2002; Rahman and Khan, 2018).

Our study also provides insights into the effects of feature transformation and feature selection in modelling bank customer data. Consistent with Abedin et al. (2020), we observed a significant effect of the feature transformation methods on the performance of classification methods. Despite the fact that PCA-based approaches were used in previous related studies (Kalaivani and Sumathi, 2019; Rahman and Khan, 2018), the findings of the current study do not support the effectiveness of PCA-based methods for modelling bank customer behaviour. These differences can be explained in part by the use of multiple machine learning-based classification methods in the current study, while previous studies considered only single classifiers, namely, the DT (Kalaivani and Sumathi, 2019) and *k*-NN (Rahman and Khan, 2018). In addition, in contradiction with earlier findings (Keramati et al., 2016), this study was unable to demonstrate the validity of feature selection for modelling bank customer behaviour. This can be attributed mainly to the high relevance of the features monitored by the bank and their low redundancy, which was confirmed by the correlation analysis.

In order to guide bank policies and decision, various steps should be taken by bank authorities to improve their customer service. In this area, the main focus has been on examining the demand and behaviour of customers and also on making the correct decisions about service. Without an awareness of customer demands, banks may lose customers. Our model can help them take the right actions and estimate the risk factors of customer behaviour. It reveals the best machine learning model that could predict customer activity in banking services. Thus, banking authorities can align their services with customer interests and reduce risk factors, and this can provide customer satisfaction with banking services. In addition, customers can have the best facilities and more banking

services than in previous times. With the data mining strategy, this model can make a difference in identifying the significant risk factors for customers and examining behavioural features through machine leaning to make appropriate decisions in the banking sector.

Given the importance of the data mining techniques used in this study for the accuracy of predicting bank customer behaviour across classification methods, these techniques can be used in other cognate subjects with similar data structures, such as loan default prediction, financial fraud detection, customer churn prediction, and bank marketing analyses.

Finally, a number of important limitations need to be considered. First, the current study examined only one benchmark dataset. Moreover, this study was unable to analyze the features in the dataset due to the data confidentiality policy of the bank. Therefore, caution must be applied because the findings might not be transferable to other models addressing customer behaviour and the features used in this study may not be available for other banks. Finally, because the benchmark dataset was balanced, something that may not correspond to reality, the issue of balancing classes in the data was not addressed in this study. It is therefore recommended that further research be undertaken in these areas to validate the proposed data mining methodology.

Funding statement

The authors received no financial support for the research, authorship, and/or publication of this article.

Ethics approval statement

This article does not contain any studies with human participants or animals performed by any of the authors.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ribaf.2023.101913.

References

Abbasimehr, H., Shabani, M., 2019. A new methodology for customer behavior analysis using time series clustering: A case study on a bank's customers. Kybernetes 50 (2), 221–242.

Abedin, M.Z., Chi, G., Uddin, M.M., Satu, M.S., Khan, M.I., Hajek, P., 2020. Tax default prediction using feature transformation-based machine learning. IEEE Access 9, 19864–19881.

Abedin, M.Z., Guotai, C., Moula, F.E., Azad, A.S., Khan, M.S.U., 2019a. Topological applications of multilayer perceptrons and support vector machines in financial decision support systems. Int. J. Finance Econ. 24 (1), 474–507.

Abedin, M.Z., Guotai, C., Zhang, T. Fahmida-E-Moula, Hassan, M.K., 2019b. An optimized support vector machine intelligent technique using optimized feature selection methods: evidence from Chinese credit approval data. J. Risk Model Valid. 13 (2), 1–46.

Akter, T., Satu, M.S., Khan, M.I., Ali, M.H., Uddin, S., Lio, P., Quinn, J.M.W., Moni, M.A., 2019. Machine learning-based models for early stage detection of autism spectrum disorders. IEEE Access 7, 166509–166527.

Alam, N., Gao, J., Jones, S., 2021. Corporate failure prediction: An evaluation of deep learning vs discrete hazard models. J. Int. Final. Inst. Money 75, 101455.

Amin, A., Al-Obeidat, F., Shah, B., Adnan, A., Loo, J., Anwar, S., 2019. Customer churn prediction in telecommunication industry using data certainty. J. Bus. Res. 94, 290–301.

Aslam, F., Hunjra, A.I., Ftiti, Z., Louhichi, W., Shams, T., 2022. Insurance fraud detection: Evidence from artificial intelligence and machine learning. Res. Int. Bus. Finance 62, 101744.

Bahnsen, A.C., Aouada, D., Stojanovic, A., Ottersten, B., 2016. Feature engineering strategies for credit card fraud detection. Expert Syst. Appl. 51, 134–142.

Baumann, C., Burton, S., Elliott, G., 2007. Predicting consumer behavior in retail banking. J. Bus. Manag. 13 (1), 79–96.

Bentéjac, C., Csörgő, A., Marénez-Muñoz, G., 2021. A comparative analysis of gradient boosting algorithms. Artif. Intell. Rev. 54 (3), 1937–1967.

Berggrun, L., Salamanca, J., Díaz, J., Ospina, J.D., 2020. Profitability and money propagation in communities of bank clients: A visual analytics approach. Finance Res. Lett. 37, 101387.

Bhatore, S., Mohan, L., Reddy, Y.R., 2020. Machine learning techniques for credit risk evaluation: a systematic literature review. J. Bank Financ. Technol. 4 (1), 111–138.

Cai, J., Luo, J., Wang, S., Yang, S., 2018. Feature selection in machine learning: A new perspective. Neurocomputing 300, 70–79.

Chandra, B., Gupta, M., 2011. An efficient statistical feature selection approach for classification of gene expression data. J. Biomed. Inform. 44 (4), 529–535.

Charte, D., Charte, F., Herrera, F., 2022. Reducing data complexity using autoencoders with class-informed loss functions. IEEE Trans. Pattern Anal. Mach. Intell. http://dx.doi.org/10.1109/TPAMI.2021.3127698.

Chen, C., Geng, L., Zhou, S., 2021. Design and implementation of bank CRM system based on decision tree algorithm. Neural Comput. Appl. 33, 8237–8247.

Chen, T., Guestrin, C., 2016. Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 785–794.

Clerkin, N., Hanson, A., 2021. Debit card incentives and consumer behavior: evidence using natural experiment methods. J. Financ. Serv. Res. 60 (2), 135–155.

De Caigny, A., Coussement, K., De Bock, K.W., Lessmann, S., 2020. Incorporating textual information in customer churn prediction models based on a convolutional neural network. Int. J. Forecast. 36 (4), 1563–1578.
de Lima Lemos, R.A., Silva, T.C., Tabak, B.M., 2022. Propension to customer churn in a financial institution: A machine learning approach. Neural Comput. Appl. 34, 11751–11768.
Demšar, J., 2006. Statistical comparisons of classifiers over multiple data sets. J. Mach. Learn. Res. 7, 1–30.
Dong, G., Liu, H. (Eds.), 2018. Feature Engineering for Machine Learning and Data Analytics. CRC Press.
Fejza, V., Livoreka, R., Bajrami, H., 2017. Analyzing consumer behavior in banking sector of Kosovo. Eurasian J. Bus. Manag. 5 (4), 33–48.
Hall, M.A., 2000. Correlation-based Feature Selection of Discrete and Numeric Class Machine Learning. University of Waikato, Department of Computer Science.
Han, J., Pei, J., Kamber, M., 2011. Data Mining: Concepts and Techniques. Elsevier.
Hastie, T., Tibshirani, R., Friedman, J.H., Friedman, J.H., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, New York.
Ho, S.C., Wong, K.C., Yau, Y.K., Yip, C.K., 2019. A machine learning approach for predicting bank customer behavior in the banking industry. In: Machine Learning and Cognitive Science Applications in Cyber Security. IGI Global, pp. 57–83.
Jain, H., Yadav, G., Manoov, R., 2021. Churn prediction and retention in banking, telecom and IT sectors using machine learning techniques. In: Advances in Machine Learning and Computational Intelligence. Springer, Singapore, pp. 137–156.
Kalaivani, D., Sumathi, P., 2019. Factor based prediction model for customer behavior analysis. Int. J. Syst. Assur. Eng. Manag. 10 (4), 519–524.
Karvana, K.G.M., Yazid, S., Syalim, A., Mursanto, P., 2019. Customer churn analysis and prediction using data mining models in banking industry. In: 2019 International Workshop on Big Data and Information Security. IWBIS, IEEE, pp. 33–38.
Keramati, A., Ghaneei, H., Mirmohammadi, S.M., 2016. Developing a prediction model for customer churn from electronic banking services using data mining. Financ. Innov. 2 (1), 1–13.
Kinge, A., Oswal, Y., Khangal, T., Kulkarni, N., Jha, P., 2022. Comparative study on different classification models for customer churn problem. In: Machine Intelligence and Smart Systems. Springer, Singapore, pp. 153–164.
Liu, Y., Yang, M., Wang, Y., Li, Y., Xiong, T., Li, A., 2022. Applying machine learning algorithms to predict default probability in the online credit market: Evidence from China. Int. Rev. Financ. Anal. 79, 101971.
Long, W., Lu, Z., Cui, L., 2019. Deep learning-based feature engineering for stock price movement prediction. Knowl.-Based Syst. 164, 163–173.
Moula, F.E., Guotai, C., Abedin, M.Z., 2017. Credit default prediction modeling: an application of support vector machine. Risk Manage. 19 (2), 158–187.
Mujica, L.E., Melendez, J., Colomer, J., 2002. Modeling the bank's client behavior using case based reasoning and self-organizing map. (Accessed 20 December 2016).
Ngai, E.W., Xiu, L., Chau, D.C., 2009. Application of data mining techniques in customer relationship management: A literature review and classification. Expert Syst. Appl. 36 (2), 2592–2602.
Papouskova, M., Hajek, P., 2019. Two-stage consumer credit risk modelling using heterogeneous ensemble learning. Decis. Support Syst. 118, 33–45.
Rahman, A., Khan, M.N.A., 2018. A classification based model to assess customer behavior in banking sector. Eng. Technol. Appl. Sci. Res. 8 (3), 2949–2953.
Raju, S.S., Dhandayudam, P., 2018. Prediction of customer behaviour analysis using classification algorithms. In: AIP Conference Proceedings, Vol. 1952 No. 1. AIP Publishing LLC, 020098.
Sattar, A.M., Ertuğrul, Ö.F., Gharabaghi, B., McBean, E.A., Cao, J., 2019. Extreme learning machine model for water network management. Neural Comput. Appl. 31 (1), 157–169.
Satu, M.S., Ahamed, S., Hossain, F., Akter, T., Farid, D.M., 2017. Mining traffic accident data of N5 national highway in Bangladesh employing decision trees. In: 2017 IEEE Region 10 Humanitarian Technology Conference. R10-HTC, IEEE, pp. 722–725.
Song, Y.Y., Ying, L.U., 2015. Decision tree methods: applications for classification and prediction. Shanghai Arch. Psychiatry 27 (2), 130.
Tan, P.N., Steinbach, M., Kumar, V., 2016. Introduction to Data Mining. Pearson Education India.
Theodoridis, G., Tsadiras, A., 2022. Applying machine learning techniques to predict and explain subscriber churn of an online drug information platform. Neural Comput. Appl. 1–14. http://dx.doi.org/10.1007/s00521-022-07603-9.
Urbanowicz, R.J., Meeker, M., La Cava, W., Olson, R.S., Moore, J.H., 2018. Relief-based feature selection: Introduction and review. J. Biomed. Inform. 85, 189–203.
Vidal, A., Kristjanpoller, W., 2020. Gold volatility prediction using a CNN-LSTM approach. Expert Syst. Appl. 157, 113481.
Wojnarski, M., 2002. Modeling the Bank Client's Behavior with LTF-C Neural Network. Institute of Informatics, Warsaw University.
Yuan, K., Chi, G., Zhou, Y., Yin, H., 2022. A novel two-stage hybrid default prediction model with k-means clustering and support vector domain description. Res. Int. Bus. Finance 59, 101536.
Zhang, X., Han, Y., Xu, W., Wang, Q., 2021a. HOBA: A novel feature engineering methodology for credit card fraud detection with a deep learning architecture. Inform. Sci. 557, 302–316.
Zhang, H., Shi, Y., Yang, X., Zhou, R., 2021b. A firefly algorithm modified support vector machine for the credit risk assessment of supply chain finance. Res. Int. Bus. Finance 58, 101482.
Zhou, X., Bargshady, G., Abdar, M., Tao, X., Gururajan, R., Chan, K.C., 2019. A case study of predicting banking customers behaviour by using data mining. In: 2019 6th International Conference on Behavioral, Economic and Socio-Cultural Computing. BESC, IEEE, pp. 1–6.