Research in International Business and Finance 74 (2025) 102722

Contents lists available at ScienceDirect

Research in International Business and Finance

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

Class imbalance Bayesian model averaging for consumer loan default prediction: The role of soft credit information

Futian Weng [a,b,c], Miao Zhu [d,*], Mike Buckle [e], Petr Hajek [f], Mohammad Zoynul Abedin [e,*]

[a] School of Medicine, Xiamen University, Xiamen 361005, China
[b] National Institute for Data Science in Health and Medicine, Xiamen University, Xiamen 361005, China
[c] Data Mining Research Center, Xiamen University, Xiamen 361005, China
[d] School of Statistics, Huaqiao University, Xiamen, China
[e] School of Management, Swansea University, Bay Campus, Fabian Way, Swansea, Wales SA1 8EN, UK
[f] Faculty of Economics and Administration, University of Pardubice, Studentska 95, Pardubice 53210, Czech Republic

ARTICLE INFO

Keywords:
Consumer loan
Soft credit information
Class imbalance
Bayesian model averaging
Variable contribution

ABSTRACT

This study investigates the predictive value of soft information for consumer loan defaults. We propose a novel framework to address class imbalance by utilizing the concept of Bayesian model averaging. Specifically, we assign unequal weights to machine learning sub-models that incorporate different combinations of variables, thereby creating an accurate and robust model for predicting consumer loan defaults. Additionally, this framework incorporates the Shapley additive explanations (SHAP) method to estimate individual contributions and employs the Bayesian information criterion to assess the variable contributions of the sub-models. We validate the effectiveness and robustness of our proposed method using authentic loan data and publicly available credit default records from a prominent consumer platform in China. Our empirical research suggests that the characteristics of user online behavior are significantly predictive of loan defaults, demonstrating asymmetry at different stages of default.

1. Introduction

The emergence of online consumer credit services in the e-commerce industry has become significant. Shoppers can now use these services on e-commerce platforms in addition to credit cards (Papouskova and Hajek, 2019; Zha et al., 2022). These services enable online consumers to make payments in monthly installments, making them convenient tools that drive purchase conversions while minimizing losses. Gradually, they have become effective marketing tools for practitioners across diverse industries worldwide (Li et al., 2023). However, evaluating consumers' credit risks is crucial in this area because elevated credit risks can significantly affect the profitability of the platforms (Zhou et al., 2021). Therefore, it is essential to utilize effective approaches to predict the likelihood of delinquency for each loan and accurately assess borrowers' credit risks (Chi et al., 2019; Abedin et al., 2022).

The entry threshold for consumer loans is lower than that for traditional bank loans. Some borrowers of consumer loans lack collateral or guarantees (Jiang et al., 2018; Moula et al., 2017). Traditional personal credit evaluation methods are limited by the characteristics of these customer groups. Therefore, e-commerce platforms should identify additional factors in rich user data, such as

* Corresponding authors.
E-mail addresses: zhumiao@hqu.edu.cn (M. Zhu), m.z.abedin@swansea.ac.uk (M.Z. Abedin).

https://doi.org/10.1016/j.ribaf.2024.102722
Received 10 January 2024; Received in revised form 13 August 2024; Accepted 19 December 2024
Available online 25 December 2024
0275-5319/© 2024 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

soft information, that impact loan delinquency (Gao et al., 2024). The term "soft credit information" refers to non-financial data used to supplement traditional hard information in credit risk management processes (Liberti and Petersen, 2019; Kowalewski and Pisany, 2022). Unlike traditional hard credit information, which includes quantitative financial metrics, soft credit information encompasses behavioral and psychological traits, social relationships, and online activities. This information can include the frequency and success of online transactions, social media behavior, email usage, and other digital footprints that reflect the borrower's habits and reliability.

Previous studies have extensively examined the relationship between soft information and borrowers. For instance, Lin et al. (2013) analyzed the significance of social relationships in credit risk assessment and revealed that robust social network connections played a vital role in the success of loans and the mitigation of default risks. Gao et al. (2023) discovered a correlation between loan default probability and the readability of the wording on loan applications, as well as the inclusion of positive emotions in the wording. Additionally, researchers have investigated the predictive relevance of textual features, including the length of the text and the number of spelling errors, on default probability (Wang et al., 2021; Kriebel and Stitz, 2022). However, there is limited research on the role of Internet behavior in predicting defaults on consumer loans. Internet behavior can provide insights into borrowers' consumption habits and other aspects. Analyzing the number of successful transactions by platform users allows for the measurement of borrowers' consumption behavior patterns and stability. Thus, the primary research question of this study is as follows: Does Internet behavior, as an emerging form of soft information, contribute to predicting consumer loan defaults?

Fintech platforms are increasingly incorporating big data and nontraditional data, along with complex algorithms such as machine learning and artificial intelligence, to enhance the precision of identifying borrowers' default risks (Loutfi, 2022; Dai et al., 2023; Rozo et al., 2023; Hasan et al., 2023; Zhao et al., 2022). Since nonperforming customers constitute a small fraction of the overall customer base, consumer loan credit evaluations often involve addressing the class imbalance issue in predicting both overdue and non-overdue loans (Abedin et al., 2023a). This approach allows loans to be assigned based on the confidence level of their respective classifications. Extensive research in the literature focuses on using machine learning techniques in credit scoring, such as artificial neural networks (ANNs), support vector machines (SVMs), decision trees, and Bayesian classifiers (Zhang et al., 2021; Sigrist and Leuenberger, 2023; Li et al., 2023). Moreover, researchers have developed intricate scoring models to comprehensively capture the complex relationships among data variables. Notably, ensemble learning techniques, such as extreme gradient boosting (XGBoost), light gradient-boosting machine (LightGBM), and categorical boosting (CatBoost), have been widely adopted in numerous credit scoring systems (Wang et al., 2020a; Fitzpatrick and Mues, 2021; Hajek et al., 2023; Song et al., 2023). More recently, as the volume of data has expanded, deep learning approaches utilizing deep neural networks have emerged for constructing credit scoring models in the realm of Internet finance (Borchert et al., 2023; Korangi et al., 2023; Che et al., 2024; Shajalal et al., 2023). This highlights the fact that model selection is a fundamental concern when building precise credit scoring systems, as excessively complex or simplistic models can yield significant variances in estimates or predictions.

Researchers have proposed various model selection methods and criteria, including stepwise regression, Akaike information criterion (AIC), cross-validation, Bayesian information criterion (BIC), generalized cross-validation, and Lasso, to address this issue (Ding et al., 2018). However, model selection methods often neglect the uncertainty arising from the selection process, resulting in an underestimation of the actual variance, as well as issues such as instability, loss of valuable information, and target deviation (Zhang and Liu, 2023). Consequently, the model averaging method has emerged. It typically does not consider a chosen model as the actual data-generation process, thereby acknowledging the uncertainty produced by the model selection procedure (Zhang et al., 2016, 2019; Guotai et al., 2017). Moreover, model averaging employs continuous weights to blend diverse models, and during prediction, model selection can be viewed as a characteristic of model averaging, with weights restricted to 0 or 1, resulting in a typically more robust estimation (Zhang et al., 2016, 2019). Nevertheless, the majority of current model averaging techniques depend on conventional statistical approaches, such as linear regression and logistic regression (Li et al., 2024), with limited exploration of artificial intelligence (AI) methods, including machine learning (Zhang et al., 2016, 2019; Lu et al., 2022). To fill this gap, the second research question of this paper is: How can one construct a more robust category imbalance algorithm based on the idea of model averaging for the class imbalance problem in consumer loan credit evaluation?

Within the academic community, a contentious debate has arisen concerning the balance between enhanced accuracy achieved through advanced credit scoring models and the corresponding reduction in interpretability (Chen et al., 2023a, 2023b). Regulatory agencies are devoted to identifying the emerging risks associated with machine learning technology and stressing the imperative nature of transparency and interpretability when modeling the lending sector. Locally interpretable model-independent interpretation methods, such as local interpretable model-agnostic explanations (LIME) and Shapley additive explanations (SHAP), have been employed to elucidate the prediction outcomes produced by "black box" machine learning models, enabling both high predictive accuracy and interpretability (Ribeiro et al., 2016; Lundberg and Lee, 2017; Lundberg et al., 2020; Ding et al., 2023; Wang et al., 2023). These methods specifically explicate the model's predictive results by computing variable contributions (Weng et al., 2022; Yang et al., 2022, 2024). Nevertheless, these methods necessitate a precise and stable model. Research has demonstrated that as category imbalance escalates, the interpretive outputs derived from these model-independent interpretation techniques become less reliable. This suggests that category imbalance negatively affects the interpretability of machine learning models. Consequently, we pose a third research question: How can we develop a resilient machine learning interpretation approach grounded in model-independent interpretation methods to produce consistent variable contributions?

Here, we utilize data obtained from a consumer loan project jointly undertaken by a commercial bank and an e-commerce platform in China. Through empirical analysis, we investigate the predictive significance of Internet behavior in determining consumer loan delinquency. Specifically, we focus on the number of successful transactions conducted by borrowers on the platform, which we refer to as soft credit information. Methodologically, we employ a range of sophisticated class imbalance algorithms to address issues arising from imbalanced classes. Additionally, we utilize machine learning techniques as sub-models of the overarching model, referred to as

candidate models, and select sub-models based on the BIC. Subsequently, we employ the Bayesian model framework to determine the weights of the sub-models, enabling the capture of their nonlinear relationships, and obtain the final credit overdue prediction model through weighted aggregation. Furthermore, we incorporate the SHAP method to elaborate on the predictions made by the sub-models and propose a novel approach for calculating variable contributions by integrating the variables and weights of the sub-models, thus ensuring a reliable assessment of variable contributions.

This study makes several significant contributions to the field of consumer loan risk assessment. Firstly, it investigates the predictive value of soft credit information on borrowers' delayed repayments in different periods using real consumer loan data. This analysis broadens the application of soft information in predicting consumer loan delinquency. Additionally, we developed a Bayesian model averaging framework tailored to address class imbalance by employing a classification algorithm that accounts for imbalanced categories. This framework is further enhanced with the SHAP method. Specifically, we use the SHAP method to assess the contributions of individual variables in sub-models, leading to a comprehensive determination of Shapley values through a weighted approach. The aggregation of interpretations from multiple sub-models enhances the reliability and robustness of determining variable contributions.

Our approach, rooted in the concept of combined forecasting, not only improves the robustness of prediction outcomes compared to individual models but also ensures greater stability in assessing variable contributions. Our findings offer valuable insights for financial institutions and their risk management divisions regarding the prediction of consumer loan delinquency. Moreover, this study highlights the advantages of employing model averaging techniques to achieve more dependable prediction results and precise evaluations of variable contributions. These insights are particularly significant for leaders within financial institutions, especially in navigating complex and uncertain long-term scenarios.

The remainder of this article is organized as follows: In Section 2, we provide a review of the relevant literature on class imbalance and machine learning interpretability. Section 3 presents the methodology proposed in this paper. In Section 4, we describe the analysis of the obtained results. Lastly, Section 5 presents the conclusion and outlines future research prospects.

2. Literature review

Assessing consumer loan risk is crucial for both banks and non-bank financial institutions when making lending decisions (Papouskova and Hajek, 2019; Yang et al., 2023). Among these factors, the information utilized and the model created play significant roles in credit risk prediction (Wang et al., 2021; Abedin et al., 2023b). Hence, this study aims to review the characteristics and methodologies employed in credit risk prediction.

2.1. Predictive features

Previous studies commonly extracted specific features from public information platforms. These features, often referred to as "hard features" (Butler et al., 2017; Cai et al., 2016; Xia et al., 2017; Ghosh et al., 2023), have been confirmed as valuable predictive factors. However, as Internet media have become more advanced, nonstandard or personalized soft features have become factors in credit risk prediction. Early research explored textual features as effective indicators of credit risk, such as analyzing the readability of the borrower's language on applications, the expression of positive emotions in responses to application questions, the length of these answers, and the frequency of spelling errors (Ge et al., 2017; Ma et al., 2018; Lu et al., 2019). Additionally, Jiang et al. (2018) identified a significant correlation between business- and education-related keywords and the probability of credit default, while Wang et al. (2021) explored the predictive value of text information from borrowers' posts on Weibo.

Further investigations have also focused on the role of soft information variables, such as psychological factors and social relationships, in credit assessment. For instance, Lin et al. (2013) examined the impact of social relationships on credit risk evaluation. Burtch et al. (2014) studied the influence of cultural differences on loan outcomes. Li et al. (2021) integrated variables related to loan manager behavior into the development of credit rating systems for small and micro-enterprises, highlighting the importance of soft information in credit models. Djeundje et al. (2021) extracted relevant soft information from user email usage and compared the predictive performance of various statistical and machine-learning methods. Furthermore, Wang et al. (2022b) assessed the significance of social and psychological soft information in predicting defaults within P2P lending markets and its relevance in credit scoring. Goel and Rastogi (2023) conducted a systematic literature review to identify borrower behavior and psychological traits, proposing a conceptual model to elucidate the impact of this soft information on credit defaults. Finally, Wang et al. (2024) investigated the role of social proximity in bank lending using proprietary datasets, while Wu et al. (2024) applied explanatory level theory to explore the impact of geographic and social distance on bid amounts in a controlled experimental setup.

2.2. Methods used in predicting credit default

The methods for predicting credit default can mainly be divided into traditional statistical learning methods and machine learning methods. Traditional methods include the linear discriminant model (LDA), logistic regression, SVM, ANN, and naïve Bayes (NB) classifier. Among machine learning methods, ensemble learning based on tree models is currently one of the popular approaches in the field of credit default. These methods include random forest (RF), extreme gradient boosting (XGBoost), adaptive boosting (AdaBoost), and light gradient boosting machine (LightGBM). Additionally, with the increase in data dimensions and relationship complexity, some scholars have attempted to apply deep learning methods to the field of credit default (Wang et al., 2022a). Gunnarsson et al. (2021) introduced several basic deep learning methods, such as the deep multilayer perceptron neural network and deep belief

network (DBN), for credit default prediction. The experimental results showed that deep learning was not superior to ensemble algorithms. Qian et al. (2023) used one-dimensional convolutional neural networks, which outperformed other benchmark machine learning models on large datasets. Xiao et al. (2024) combined the variational autoencoder (VAE) with deep forest and applied it to the class imbalance dataset of internet finance.

Some interpretable machine learning methods have also been used in the field of credit default prediction, such as the model-independent local interpretations SHAP and LIME. The SHAP method has been especially widely used owing to its theoretical properties in calculating sample-level contributions. For instance, Chen et al. (2023b) proposed a novel selective learning framework that highlights the superiority of machine learning models over logistic regression in credit scoring through both global and local interpretations. Sun et al. (2023) developed an interpretable decision support system for predicting credit defaults utilizing the SHAP method. Zhang et al. (2023) presented a credit default prediction model with profit maximization as the optimization objective and utilized the SHAP method to generate interpretable prediction outcomes. Chen et al. (2024) further examined the influence of class imbalance on the stability of explanations derived from SHAP and LIME and discovered that class imbalance adversely affects the interpretability of machine learning.

2.3. Research gap

Recent research has increasingly acknowledged the significance of soft features. With the widespread use of the Internet, individuals have generated vast amounts of personal behavior data online. This "soft credit information" has the potential to offer valuable supplementary insights, yet there has been limited research exploring its value in predicting consumer loan defaults. While numerous studies have developed advanced credit default prediction methods using machine learning and deep learning, there remains a lack of robust development in class imbalance models specifically tailored to the consumer loan domain. Furthermore, while some studies have employed model-independent interpretable machine learning methods such as SHAP to elucidate the "optimal" classification model in their research, these interpretive methods may exhibit bias and lack robustness in scenarios involving noisy data or imbalanced categories. This limitation becomes more pronounced when the model itself lacks robustness (Alvarez-Melis and Jaakkola, 2018; Chen et al., 2023a). Thus, exploring a more dependable and resilient class imbalance model might be a viable solution.

3. Methodology

3.1. Class imbalance method

Class imbalance algorithms can be placed in three main categories: sampling techniques, cost-sensitive learning methods, and ensemble methods. Sampling techniques involve balancing samples through various methods and building classification models to divide these samples. They include undersampling, oversampling, and combined sampling (Soltanzadeh and Hashemzadeh, 2021; Jiang et al., 2024). Cost-sensitive learning methods focus on enhancing the prediction performance of selected categories by assigning cost weights to different categories during model training (Elkan, 2001). By assigning higher costs to specific categories, the classification importance of the categories can be strengthened. Ensemble methods use a mechanism to balance the data at each iteration of the model (Seiffert et al., 2009; Hasan et al., 2024).

While model-agnostic methods such as SHAP theoretically have the ability to explain any type of model, the computational complexity is greater for sophisticated ensemble models or datasets with high dimensionality due to the necessity of accessing feature values for individual samples when calculating SHAP values (Fatima and Wooldridge, 2008; Luo et al., 2022; Chen et al., 2023a). To facilitate computations when constructing our class imbalance model, we utilize a range of sampling techniques and various classification algorithms. Specifically, we introduce seven classic sampling techniques: random oversampling (RO), synthetic minority oversampling technique (SMOTE), adaptive synthetic sampling (ADASYN), borderline SMOTE analysis (borderline-SMOTE), random undersampling (RU), SMOTE edited nearest neighbors (SMOTE-ENN [SE]), and SMOTE-Tomek (ST). These techniques cover three categories of sampling: upsampling, downsampling, and integrated sampling, and are broadly applicable and representative.

Regarding the selection of classification methods, we opted for eight benchmark classifiers, which encompass traditional statistical methods and machine learning approaches, including logistic regression (logistic), K-nearest neighbor (KNN), NB, multilayer perceptron (MLP), decision tree (DT), RF, LightGBM (LGB), and CatBoost (CAT). These models are commonly used as benchmarks for credit default prediction tasks (Xiao et al., 2024; Yin et al., 2023; Mahbobi et al., 2023).

3.2. Bayesian model averaging

Bayesian model averaging (BMA) is employed to address uncertainty within models, particularly in the context of model selection (Figini and Giudici, 2017). It is customary to choose the optimal model for prediction from a set of candidate models, based on a certain model evaluation criterion. In contrast to the traditional approach of selecting a single best model, BMA uses the Bayesian theorem and BIC to assign weights to each model and compute a weighted average to determine the final prediction value. The BIC is a model information criterion used to select the optimal model for a given dataset by balancing the model's goodness of fit and complexity. Here, the BIC value is employed as the criterion for screening sub-models:

$$BIC = k\ln(n) - 2\ln(\hat{L}) \tag{1}$$

where k represents the number of free parameters in the model, n stands for the sample size of the dataset, and $\widehat{L}$ signifies the maximized value of the likelihood function.

Given the dataset D, y represents the future observed value; the set of predictors consisting of all individual models is $M = \{M_1, M_2, M_3, \ldots, M_k\}$; and the probability density function (PDF) of y under the dataset D is:

$$p(y|D) = \sum_{i=1}^{k} p(M_i|D)p(y|M_i, D) \tag{2}$$

where $p(y|M_i, D)$ is the posterior distribution of y and $p(M_i|D)$ is the posterior probability of the modified model after information from the dataset D has been incorporated (Zou et al., 2021).

Based on the Bayesian theorem, the posterior probability value of the model can be derived:

$$p(M_i|D) = \frac{p(D|M_i)p(M_i)}{\sum_{k=1}^{K} p(D|M_k)p(M_k)} \tag{3}$$

Utilizing methods such as the Laplace's and Taylor series expansions, we can ultimately derive:

$$p(D|M_k) = e^{-BIC/2} \tag{4}$$

By incorporating the BIC into Eq. (3), we obtain:

$$p(M_i|D) = \frac{e^{-BIC_i/2} p(M_i)}{\sum_{k=1}^{K} e^{-BIC_k/2} p(M_k)} \tag{5}$$

where $p(M_i)$ represents the prior probability distribution of the model. In this paper, all prior probability values for the models are set to 1.

Employing the BMA method enables the acquisition of predictive values that are superior to and more robust than the values offered by a single model. In the process of model selection, we cannot guarantee the correctness of the single model we employ. However, the models selected by BMA with higher probability values encompass models that potentially fit the actual data distribution. This enhances the accuracy of the models' predictions. Simultaneously, the weighted combination of multiple models mitigates errors associated with using a single model, thereby increasing robustness.

3.3. The method proposed in this study

3.3.1. Class imbalance model averaging

We have constructed a framework that combines class imbalance techniques with BMA methods. The former is employed to address the issue of imbalanced data, while the latter adaptively selects the best-performing sub-models, enhancing the predictive performance and yielding more stable predictions.

Specifically, in this study, to achieve superior and more stable predictive results while enhancing the discriminative and robust interpretability of features, we employ a BMA method to adaptively select sub-models under different variable combinations. Building upon the model averaging framework, we incorporate strategies to address class imbalance, thereby mitigating the disparity between default and non-default cases and further enhancing the predictive performance of the model (see Fig. 1). The specific calculation steps are as follows:

Introduction: For a dataset D containing feature set W, with the feature variables $\{x_1, x_2, \ldots, x_m\}$. The dataset is divided into a training set W_{train} and a test set W_{test}. The training set data W_{train} undergoes class imbalance handling, while the test set W_{test} is used for final model evaluation.

Step 1: Perform data balancing. Class imbalance is identified using the class imbalance ratio, and an appropriate data balancing technique is selected based on the specific characteristics of the dataset, including RU, RO, SMOTE and its extensions, and ADASYN.

Step 2: Initialize the prior probabilities of model occurrence (in this paper, all model prior probability values are set to 1, indicating an equal initial belief in each model).

Step 3: Select the feature combinations at the i-th level (where the i-th level refers to selecting i features from m features for permutation and combination, denoted as C_m^i). These feature combinations constitute different feature sets, ultimately forming a sample set $D^i\{D_1^i, \ldots, D_j^i; i \le m, j \le C_m^i\}$, where j represents the j-th group of features in the permutation and combination C_m^i, and D_j^i represents the data subset formed by the j-th group of features at the i-th level.

Step 4: Based on Eq. (5), compute the model's Bayesian information values under D_j^i, as well as the posterior probability values $P(M_j^i|D_j^i)$. The Bayesian information values are used to evaluate the model's goodness of fit while penalizing complexity to avoid overfitting.

Step 5: In accordance with the Occam's razor principle $M_k^i : \frac{Max\left\{P\left(M_1^{i-1}|D_1^{i-1}\right), \ldots, P\left(M_j^{i-1}|D_j^{i-1}\right), \ldots, P\left(M_1^1|D_1^1\right), \ldots, P\left(M_1^1|D_1^1\right)\right\}}{P\left(M_k^i||D_k^i\right)} \le C, C = 20$ (Zou et al., 2021), retain the model M_k^i $(i \le m, k \le C_m^i)$ under this feature combination $D^i\{D_1^i, \ldots, D_j^i; i \le m, j \le C_m^i\}$ that meet the selection criteria, forming the model set $M^i = \{M_k^i; i \le m, k \le C_m^i\}$. That is, models are retained that balance complexity and fit well to the data. This involves keeping models with higher posterior probabilities and discarding those that do not meet these criteria.

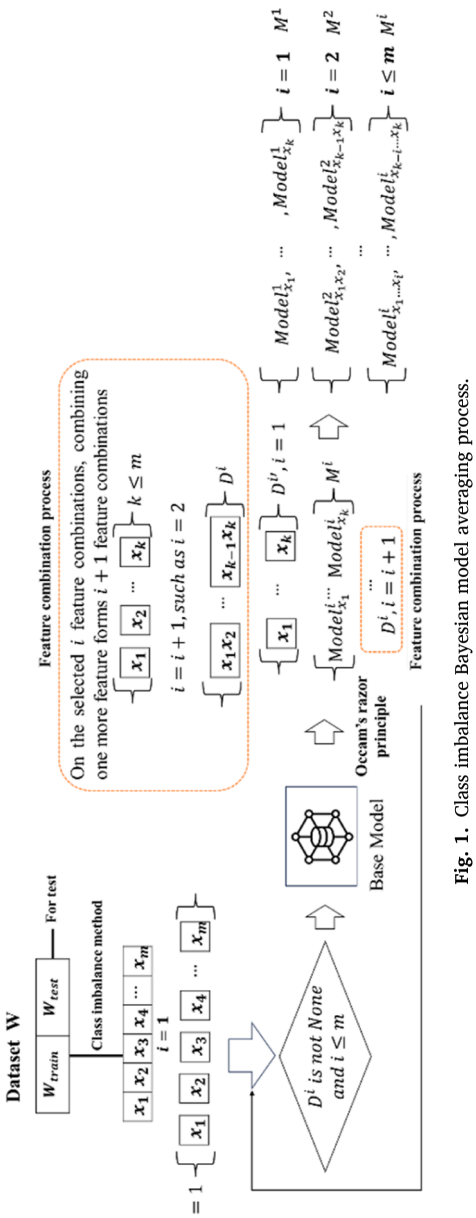

Fig. 1. Class imbalance Bayesian model averaging process.

Step 6: Using the feature sets $D^i\{D_1^i,\ldots,D_j^i; i \le m, j \le C_m^i\}$ from Step 5 that meet the criteria, further to obtain new, higher-level feature combinations at the $i+1$ level (as detailed in the feature combination process illustrated in Fig. 1). Continue this process iteratively to explore all possible combinations and build a comprehensive set of candidate models.

Step 7: Repeat steps 3–6, ultimately selecting K models. Normalize the posterior probability values of K models to obtain the weights α_{M_k} for each model M_k ($k = 1,2,\ldots,K$) under different feature combinations. Utilizing these model weights, make a weighted prediction regarding whether the user will experience default:

$$v_0 = \sum_{k=1}^{K} \alpha_k(M_k(y_i) == 0), \quad v_1 = \sum_{k=1}^{N} \alpha_{M_k}(M_k(y_i) == 1)$$

$$y_i = \begin{cases} 0 & v_0 > v_1 \\ 1 & v_1 <= v_0 \end{cases} \tag{6}$$

This final step ensures that the prediction takes into account the contributions of all selected models, weighted by their reliability.

3.3.2. Measurement of feature contribution

The main concept behind the SHAP values is a method derived from cooperative game theory. Specifically, the calculation method for the contribution value is as follows:

$$\phi_j(x) = \sum\nolimits_{S \subseteq W \setminus j} \frac{(n - |S| - 1)!(|S|)!}{n!} [f(x_S \cup j) - f(x_S)] \tag{7}$$

where j represents a variable, W is the set of all variables, x_S is the vector composed of variable values in the subset S of x, f is the prediction function of a certain model, and $\phi_j(x)$ represents the average contribution value of j to the output result among all possible combinations of variable features.

In this paper, to calculate the contributions of each feature variable to the prediction results more reliably, we have developed a model interpretation approach by combining the SHAP method with model averaging techniques.

Step 1: In Section 3.3.1, we obtained K sets of feature combinations and the corresponding model weight values $(\alpha_1, \alpha_2, \ldots, \alpha_K)$ for these different feature combinations.

Step 2: For the k-th set of features $(k = 1,2,\ldots,K)$, we can obtain the strength of the j-th variable $(j = 1,2,\ldots,m)$ in predicting the result for the i-th sample instance $(i = 1,2,\ldots,N)$:

$$I_{j_k}^i = |\varphi_{j_k}^i| \tag{8}$$

where, for variables j that do not exist in the k-th set of feature combinations, $I_{j_k}^i = 0$.

Table 1
Descriptive statistics of variables.

Variable	Definition	Mean	SD	Min	Max
Dependent Variable					
Overdue of Period 1	Whether users have exceeded the repayment date for 30 days when repaying. Coded as if the user fails to complete in time 1, and 0 otherwise.	0.072	0.259	0	1
Overdue of Period 2	Whether users have exceeded the repayment date for 90 days when repaying. Coded as if the user fails to complete in time 1, and 0 otherwise.	0.012	0.108	0	1
Internet online behavior					
Number of successful transactions	The number of successful transactions on the platform.	2.798	0.506	0	1
Demographics					
Gender	Gender of the user (0 = female, 1 = male)	0.603	0.489	0	1
Age	Age of the user	30.428	5.439	23	53
Marital status	Marriage probability rating of the user.	1.701	0.705	1	3
Working industry	Working industry of the user	4.495	1.856	1	8
Social networking	Number of cities where users have logged in the past 90 days	1.821	0.737	1	3
Historical credit information					
Revolving line	The average value of cc quota	1.636	0.720	1	5
Revolving line utilization	The average level of cc quota utilization rate	1.986	1.096	1	4
Delinquency history	The number of uncleared LN	2.164	0.830	1	4
Past loan number	Number of LN approval cause queries in the last month	1.095	0.310	1	3
Lending time	Maximum mob from card issuance (all credit cards) time to report generation time	3.133	1.164	1	4
Lending organization number	The number of card issuers of CC	1.908	0.734	1	5
Lending number	Number of CC cards	2.670	1.043	1	5
Loan project information					
Repayment periods	Duration of loan	215.291	135.337	29	396
Loan amount	The amount of borrowing	3522.360	5045.965	99	40,000
Debt to income type	Day rate	0.015	0.003	0.008	0.020

Simultaneously, the contributions are weighted:

$$FI_{j_k}^i = \alpha_k I_{j_k}^i \tag{10}$$

In the k-th set of feature combinations, the contribution value of feature j is determined:

$$FI_{j_k} = \sum_{i=1}^{N} FI_{j_k}^i \tag{11}$$

Step 3: The second step is repeated iteratively all the way up to the K-th set of feature combinations to compute the weighted contributions of each feature to the prediction results:

$$FI_j = \sum_{k=1}^{K} FI_{j_k} \tag{12}$$

4. Result analysis

4.1. Data description

The data used in this research were collected from a collaborative online loan program launched by a commercial bank and an e-commerce platform in China. Users browse and purchase products on the online platform and have the option to pay in installments after filling out an online application form when making their purchases. Once reviewed by both the e-commerce platform and the bank, the approved loans are disbursed to the dealers on a pro-rata basis.

As the loan is managed, if a user fails to repay the loan in full by the agreed-upon date, a penalty interest rate is charged daily until the principal and interest are fully paid off. For our study, a total of 16 independent variables and 2 dependent variables were included, as shown in Table 1. The dependent variables were defined by the platform's data showing whether a user had repaid the loan by 30 days past the original repayment date (Period 1) or by 90 days past the original repayment date (Period 2).

There is currently no universally established standard for defining the threshold of overdue days, which may involve trade-offs among interests. On one hand, an excessive number of overdue payments poses risks to stakeholders, such as banks; on the other hand, the potential interest income from late payments can also benefit stakeholders (Jiang et al., 2024). Additionally, this paper aims to investigate whether soft credit information contributes differently to predicting overdue personal consumer loans across different periods. Hence, we consider two distinct periods in our empirical analysis.

Given the class imbalance in the dataset, the imbalance ratio (IR) was employed to describe the degree of class imbalance. Notably, a longer period results in a significantly lower proportion of individuals with overdue payments compared to Period 1. In Period 1, the IR was 14.301 (74450:5206), while in Period 2, the IR was 93.829 (78816:840).

Online behavioral characteristics reflect individuals' personality traits and their ability to acquire information. This paper utilizes the number of successful transactions in which a borrower has engaged to characterize aspects of users' soft credit information. Regarding other objective information, we describe the characteristics of loan users and loan projects from three perspectives.

The first area is borrowers' personal attributes, including gender, age, marital status, employment sector, and the number of cities in which they were active during the past 90 days (on social media). Marital status is represented by estimated probability values on the platform, where values of 1, 2, and 3 signify the probabilities of marriage being within the intervals [0.0, 0.3], [0.3, 0.7], and [0.7, 1.0], respectively. Industries are coded from 1 to 9 for the categories of agriculture, forestry, and livestock (1), manufacturing (2), social services (3), public services (4), real estate (5), wholesale and retail sales (6), finance (7), and unspecified industries (8). Additionally, the number of active cities is encoded as 1, 2, or 3 by the platform, with higher values reflecting a greater number of active cities.

The second area is the borrower's credit history, which comprises variables such as the presence of a revolving line, revolving line utilization, delinquency history, numbers of past loans, lending time, and numbers of lending organizations. The bank considers these indicators as component-type variables, with higher values indicating a greater number or score.

The third area is the characteristics of the loan item, including repayment periods, loan amounts, and debt-to-income ratios. Repayment periods are represented by continuous variables, which denote the number of days between the maturity date and the borrowing date. The loan amount refers to the specific sum borrowed by the user. These two indicators are not graded by the bank; hence, we utilized the rawest available data. The interest rate type is portrayed by the daily interest rate charged by the lender.

4.2. Evaluation metrics and details

To assess the performance of the imbalanced class model averaging methods on the two datasets mentioned above, this study employs the area under the curve (AUC) as the evaluation metric for model accuracy. Its mathematical expression is given by:

$$\mathrm{AUC} = \int_0^1 \mathrm{TPR}\left(\mathrm{FPR}^{-1}(t)\right) dt \tag{13}$$

where TPR represents the true positive rate, FPR is the false positive rate, and $\mathrm{FPR}^{-1}(t)$ denotes the inverse function of the false positive rate. The integration covers all possible decision thresholds.

In addition, we also used metrics such as Specificity, Sensitivity, G-mean, and Accuracy to evaluate the performance of subsequent

models. This study aims to comprehensively identify individuals who may default using the model. Therefore, we place greater emphasis on the Sensitivity metric. In cases where other metrics are similar, this research prioritizes models with superior Sensitivity.

$$\text{Specificity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{14}$$

$$\text{Sensitivity} = \frac{TP}{\text{TP} + \text{FP}} \tag{15}$$

$$\text{G} - \text{mean} = \sqrt{\text{Specificity} * \text{Sensitivity}} \tag{16}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{17}$$

In this study, we assessed the performance of each baseline model by calculating average Specificity, Sensitivity, G-mean, Accuracy, and AUC values obtained from 30 repetitions of 10-fold cross-validation experiments. Specifically, in each 10-fold cross-validation iteration, eight randomly selected subsets were used as the training set, while the remaining two subsets served as the validation and test sets. We employed traditional statistical methods such as logistic regression, machine learning, and a combination of class imbalance algorithms and machine learning as benchmark models. Notably, the class imbalance algorithm was fused with the best machine learning model to establish the benchmark for our report. To ensure fairness in comparison, we calculated the average values across 30 repetitions of the process.

Regarding our proposed Bayesian class imbalance model averaging framework, we included the aforementioned class imbalance combined with the machine learning approach as a candidate sub-model. For equitable evaluation, we also reported the average values of each evaluation indicator based on 30 repetitions of 10-fold cross-validation. Additionally, the selection of hyperparameters plays a vital role in determining the predictive performance of machine learning models. We employed the grid search method to optimize the selection of essential parameters across different models. The specific parameter search ranges are provided in Table 2.

4.3. Performance of classification models

Tables A1(a) and A1(b) present the performance of different baseline models on the test set during periods 1 and 2, respectively. We conducted a comparative analysis of different baseline models, grouping each baseline model with its respective class imbalance handling methods. In total, there were 9 groups, each consisting of 8 schemes (the original model and the original model combined with various class imbalance methods). The model with the highest average AUC, Sensitivity, Specificity, G-mean, and Accuracy in each group is highlighted in bold. The model with the highest proportion of optimal indicators from each group was then combined with the BMA.

The experimental results indicate that, firstly, in Period 1, models combined with the BMA framework performed the best and generally outperformed the original models without the BMA framework. Secondly, the combination of random undersampling (RU) with the baseline model yielded the best performance for most models, and SE was often the best class imbalance method for logistic regression. Thirdly, the BMA-RU-LGB (Bayesian model averaging using LightGBM with undersampling) scheme emerged as the best-performing model among all models,[1] with the following metrics: 0.7635 (AUC), 0.7283 (sensitivity), 0.6608 (specificity), 0.6937 (G-mean), and 0.6652 (accuracy). Consequently, the LGB model, which demonstrates the optimal performance, was selected as the baseline model for subsequent integration with methods addressing class imbalance. While some models exhibit inferior performance compared to the original LGB model in other metrics, considering the overall performance across all metrics, the RU-LGB (LightGBM with undersampling) model emerges as the top performer (Table 3a). Based on this comprehensive comparison, the RU-LGB model surpasses the original LGB model and is selected as the baseline model for the BMA framework.

The results for Period 2 were very similar to those for Period 1, with some differences in the selection of the best class imbalance schemes for certain models. For example, in Period 1, the best class imbalance scheme for the NB model was RU, while in Period 2, it was BSMOTE. During Period 2, we observed that the overall AUC score of the model was lower than that of Period 1, which could be attributed to the highly imbalanced class proportions in Period 2. Nevertheless, the BMA-RU-LGB scheme remained the best-performing model across all models, with the following metrics: 0.7191 (AUC), 0.6586 (sensitivity), 0.6637 (specificity), 0.6607 (G-mean), and 0.6636 (accuracy), see Table 3b.

4.4. Empirical analysis

Finally, Fig. 2 illustrates the globally weighted variable contributions calculated by BMA. In Period 1, the top three variables contributing most to the prediction are the repayment periods, loan amount, and revolving line utilization. The debt-to-income ratio type, lending number, age, and gender of the borrower also make certain contributions. Regarding the soft credit information, we find that this variable, being a soft factor, has a relatively lower impact compared with other variables.

[1] The Bayesian model averaging process involves LightGBM models trained on an undersampled dataset to effectively address class imbalance and enhance prediction robustness.

Table 2
Hyperparameter settings of various machine methods.

Parameter	Logistic	GaussianNB	MLP	KNN	Decision Tree	Random Forest	XGBoost	LGB	CAT
The number of decision trees						[10,20,50,100, 200,500]	[10,20,50,100, 200,500]	[10,20,50,100, 200,500]	[10,20,50,100, 200,500]
Tree depth					[1,2,3,5,10]	[1,2,3,5,10]	[1,2,3,5,10]	[1,2,3,5,10]	[1,2,3,5,10]
Learning rate					[0.001,0.01, 0.1,0.2]	[0.001,0.01, 0.1,0.2]	[0.001,0.01, 0.1,0.2]	[0.001,0.01, 0.1,0.2]	[0.001,0.01, 0.1,0.2]
L2 regularization							1.0	1.0	1.0
No. of neurons in the 1st hidden layer			[1,2,5,10,20,30]						
No. of neurons in the 2nd hidden layer			[1,2,3,5]						
N neighbors				[10,20,30,40,50]					

Table 3a
Evaluation metrics of LGB-based classification models for Period 1.

Model	AUC	Sensitivity	Specificity	G-mean	Accuracy
LGB	0.7355	0.0223	0.9955	0.1468	0.9320
RO-LGB	0.7475	0.5423	0.7940	0.6561	0.7775
SMOTE-LGB	0.7464	0.0158	0.9960	0.1234	0.9320
BSMOTE-LGB	0.7456	0.0144	0.9960	0.1175	0.9319
ADASYN-LGB	0.7465	0.0127	**0.9967**	0.1100	**0.9324**
RU-LGB	0.7489	0.7045	0.6648	0.6842	0.6674
SMOTEENN-LGB	0.7463	0.2200	0.9447	0.4556	0.8973
SMOTETOMK-LGB	0.7427	0.0172	0.9946	0.1283	0.9307
RU-BMA-LGB	**0.7635**	**0.7283**	0.6608	**0.6937**	0.6652

Table 3b
Evaluation metrics of LGB-based classification models for Period 2.

Model	AUC	Sensitivity	Specificity	G-mean	Accuracy
LGB	0.6733	0.0043	0.9992	0.0367	0.9887
RO-LGB	0.6979	0.3071	0.8924	0.5221	0.8862
SMOTE-LGB	0.6927	0.0002	0.9998	0.0022	0.9893
BS-LGB	0.6918	0.0010	0.9997	0.0087	0.9892
ADASYN-LGB	0.6944	0.0000	**0.9999**	0.0000	**0.9894**
RU-LGB	0.6972	0.6364	0.6472	0.6412	0.6471
SMOTEENN-LGB	0.6910	0.0048	0.9980	0.0423	0.9876
SMOTETOMK-LGB	0.6871	0.0026	0.9989	0.0240	0.9884
RU-BMA-LGB	**0.7191**	**0.6586**	0.6637	**0.6607**	0.6636

In Period 2, the top three variables contributing most to the prediction are the loan amount, loan term, and debt-to-income ratio. Compared with Period 1, the importance of the revolving line utilization decreases significantly, while the relative contribution of the debt-to-income ratio becomes third in importance. This suggests that over time, relative to the revolving line utilization, the debt-to-income ratio may make a greater contribution to longer-term credit default predictions. There is considerable variation in the ranking of the other control variables between the two periods, with age and gender making certain contributions. The revolving line, industry of employment, and lending time emerge as new high-ranking variables in terms of importance. This indicates that when predicting default situations in the long term and short term, caution should be exercised, and different features should be chosen depending on the circumstances. The impact of the soft credit information increases, with the contribution value rising from 11.2834 to 13.4771. This increase indicates the importance of soft credit information in long-term credit default predictions. The primary explanation for this phenomenon is that soft credit information can have a persistent impact on personal consumption loans.

4.5. Robustness check

To further validate the effectiveness of our proposed framework, we perform robustness testing using the soft credit information variable. Here, soft credit information is defined as the frequency of successful transactions within 90 days. We replace this variable with the number of successful transactions within 180 days and 360 days to assess the robustness of the framework.

Tables A2 and A3 present the evaluation metrics for different variables in predicting results over different time periods. Firstly, the selection of the optimal class imbalance schemes was largely consistent. Secondly, the best models in each group were still those combined with the BMA. Lastly, the BMA-RU-LGB model remained the best overall model. Figs. 3 and 4 illustrate the overall distributions of different variables and their contributions to the prediction results for the two different time periods. The results are highly consistent compared with the previous results for soft credit information: over time, the importance of online behavior gradually increases.

5. Conclusion

This study proposes a novel class imbalance algorithm using real consumer loan data from a major Chinese consumer platform and bank. The algorithm integrates model-agnostic interpretation methods for model interpretation. Users' soft credit information provides additional insights that could predict consumer loan defaults, contributing significantly to long-term default predictions. These findings enhance existing research on the impact of soft information on credit risk management and deepen our understanding of the influence of soft credit information on consumer credit risk.

This study makes several contributions to existing literature. Firstly, it strengthens the field of consumer loan risk management by emphasizing the impact of soft information on credit delinquency. Additionally, it addresses the lack of in-depth discussion on the relationship between soft credit information and overdue payments. The findings reveal that soft credit information contributes more significantly to long-term overdue forecasting compared to short-term forecasting. These insights provide valuable guidance for optimizing and integrating the role of soft information, enabling more effective risk assessment and decision-making by management.

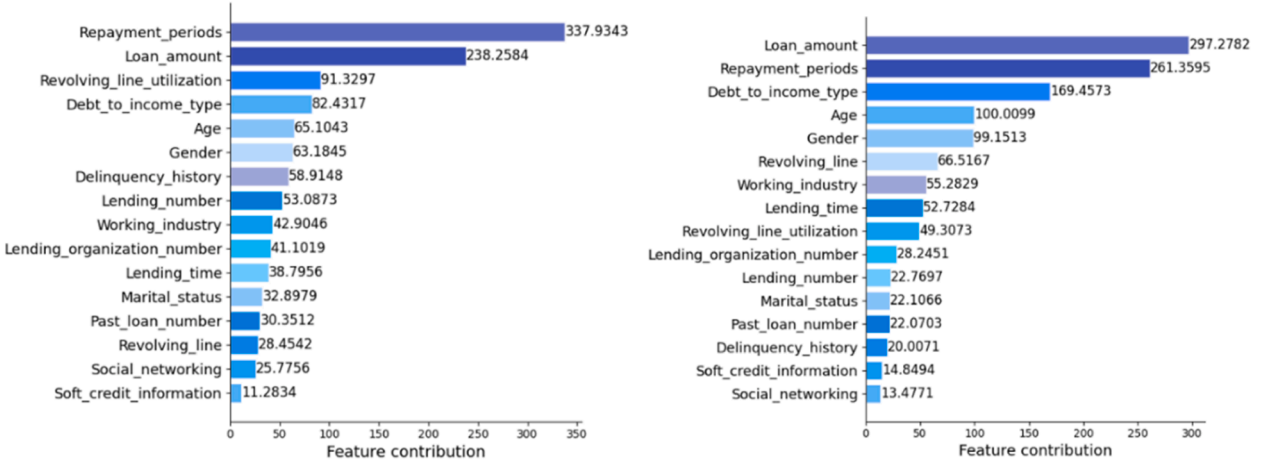

Fig. 2. Globally weighted variable contributions as calculated by the BMA in different periods (Period 1 on the left, Period 2 on the right).

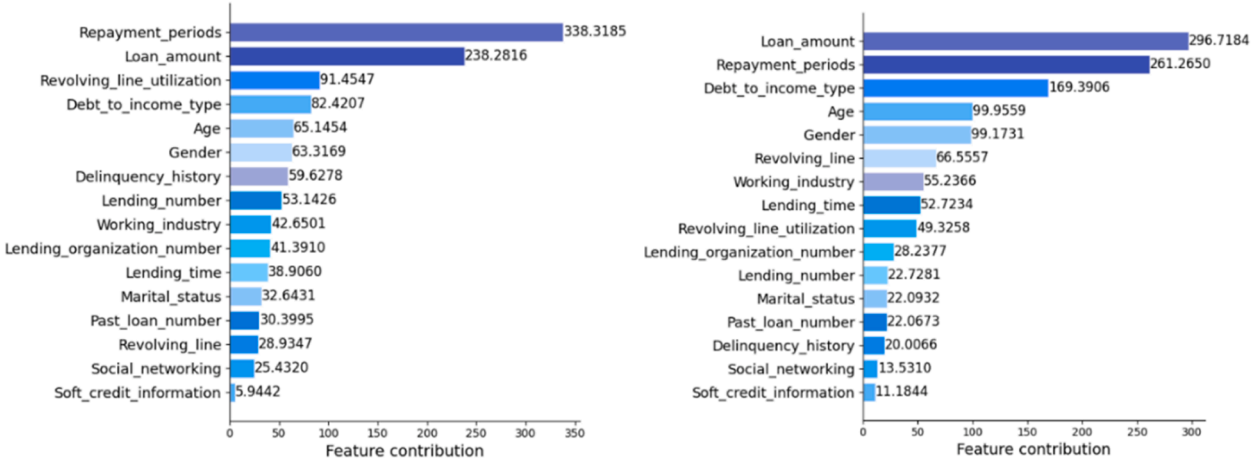

Fig. 3. Globally weighted contributions for different periods (robustness check 1, 180 days) (Period 1 on the left, Period 2 on the right).

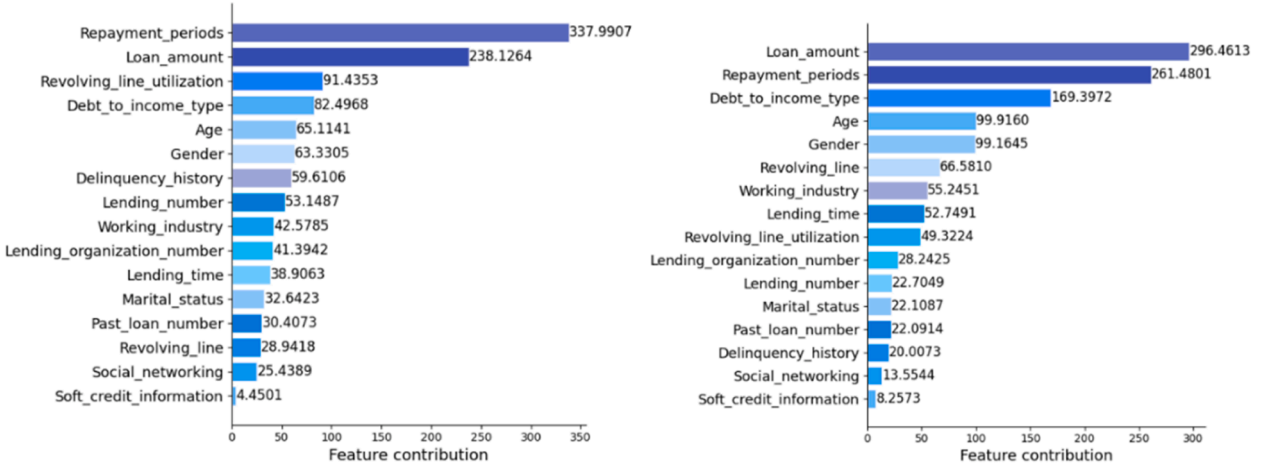

Fig. 4. Globally weighted contributions for different periods (robustness check 2, 360 days) (Period 1 on the left, Period 2 on the right).

In other words, soft information demonstrates a persistent predictive effect on consumer loan delinquency and allows for early detection of financial risks.

Moreover, in terms of methodology, we have developed a framework that combines class imbalance techniques with BMA methods. We have also incorporated the SHAP method to enhance the robustness of model interpretation. This approach achieves improved and stable prediction results while enhancing interpretability.

Variable contribution calculation, also known as feature attribution, is a widely adopted tool for explaining machine learning models. However, recent advancements in adversarial machine learning have underscored the limitations and fragility of state-of-the-art interpretation methods, such as SHAP, Grad CAM, and LIME (Baniecki et al., 2024). This lack of robustness is particularly concerning in the context of adversarial manipulation of credit scoring applications, as it has the potential to compromise safety and undermine credibility (Wang et al., 2020b). To address this issue, aggregating the explanations generated by multiple sub-models can offer more robust variable contribution assessments, considering that attacks typically target a single interpretation (Rieger and Hansen, 2020). The experimental results demonstrate that the proposed method outperforms the benchmark class imbalance model and showcases competitive performance across different datasets. Furthermore, by integrating Bayesian methods with the model-agnostic explanatory approach of SHAP, we can obtain reliable insights into variable contributions. This combination provides a dependable explanatory tool for the application of complex machine learning models within the field of consumer loans.

The findings and methodology presented in this study hold significant importance for Fintech lending institutions. Firstly, users' soft credit information can act as a valuable indicator for decision-makers determining whether to offer loans. Given the relatively low entry threshold for online consumer loans compared with loans from traditional banks, financial institutions and e-commerce platforms should prioritize soft information. Secondly, the proposed interpretable method using class imbalance model averaging quantifies variable contributions and enables the interpretation of complex models. This offers practitioners a novel tool with potential applications in various practical scenarios, including financial fraud detection (Hajek et al., 2023) and customer churn prediction (Jiang et al., 2024).

This study also has several limitations. Firstly, it utilizes consumer loan data from a major Chinese consumer e-commerce platform and bank, which might introduce sample specificity and limitations. Hence, it is essential to validate the generalizability of the research findings on diverse datasets. Moreover, this study employs a combination of the Bayesian method and SHAP method to elucidate the predictive outcomes of the model. While the provided variable contribution explanations are reliable, they are just one aspect of model interpretation. Future research can explore alternative perspectives for model interpretation, including visual analysis, adversarial learning, and more.

Declarations

None.

Ethics approval and consent to participate

Not applicable

Consent for publication

All authors are very positive to publish this manuscript on this journal.

Funding

There is no funding for this research. However, authors will use personal fund for the associated costs of publications in this journal.

CRediT authorship contribution statement

Mike Buckle: Writing – review & editing, Visualization, Validation, Supervision, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Petr Hajek:** Writing – review & editing, Validation, Supervision. **Mohammad Zoynul Abedin:** Writing – review & editing, Validation, Supervision, Software, Resources, Methodology. **Futian Weng:** Writing – original draft, Software, Resources, Methodology, Formal analysis, Data curation, Conceptualization. **Miao Zhu:** Writing – original draft, Formal analysis, Conceptualization.

Acknowledgments

This work is supported by the financial support provided by the major project of National Social Science Foundation (20&ZD137). We are grateful to the anonymous reviewers who commented on this manuscript.

Competing interests

There is no competing interest among the authors.

Appendix A

Table A1 (a)
Evaluation metrics of different baseline classification models (90 days, Period 1)

Model	AUC	Sensitivity	Specificity	G-mean	Accuracy
Logistic	0.7145	0.0003	**1.0000**	0.0061	**0.9346**
RO-Logistic	0.7335	0.6882	0.6655	0.6655	0.6669
SMOTE-Logistic	0.7325	0.6897	0.6630	0.6762	0.6647
BSMOTE-Logistic	0.7322	0.6464	0.7016	0.6734	0.6980
ADASYN-Logistic	0.7326	0.6947	0.6572	0.6756	0.6597
RU-Logistic	0.7329	0.6885	0.6644	**0.6763**	0.6660
SE-Logistic	**0.7341**	**0.7718**	0.5795	0.6687	0.5920
ST-Logistic	0.7328	0.6872	0.6653	0.6761	0.6668
BMA-SE-Logistic	**0.7466**	**0.7754**	0.5851	0.6735	0.5975
KNN	0.7200	0.0000	**1.0000**	0.0000	**0.9346**
RO-KNN	0.7205	0.0000	1.0000	0.0000	0.9346
SMOTE-KNN	0.7172	0.6662	0.6415	0.6536	0.6431
BSMOTE-KNN	0.7140	0.5982	0.7209	0.6566	0.7129
ADASYN-KNN	0.7146	0.6814	0.6206	0.6501	0.6245
RU-KNN	**0.7286**	0.6496	0.6931	**0.6708**	0.6903
SE-KNN	0.7102	**0.7607**	0.5443	0.6434	0.5585
ST-KNN	0.7175	0.6668	0.6416	0.6540	0.6433
BMA-RU-KNN	**0.7388**	0.6613	**0.6973**	**0.6780**	0.6942
NB	0.7195	0.0000	**1.0000**	0.0000	**0.9346**
RO-NB	0.7195	0.7573	0.6003	0.6742	0.6106
SMOTE-NB	0.7188	0.7596	0.5948	0.6721	0.6056
BSMOTE-NB	0.7177	0.7342	0.6209	0.6751	0.6283
ADASYN-NB	0.7188	**0.7614**	0.5917	0.6711	0.6027
RU-NB	**0.7200**	0.7545	0.6037	**0.6747**	0.6135
SE-NB	0.7178	0.7594	0.5951	0.6721	0.6058
ST-NB	0.7186	0.7587	0.5960	0.6724	0.6067
BMA-RU-NB	**0.7411**	0.6923	**0.6642**	**0.6781**	0.6660
MLP	0.6954	0.0121	0.9944	0.0251	0.9302
RO-MLP	0.7065	0.0008	0.9998	0.0065	**0.9345**
SMOTE-MLP	0.7158	0.6727	0.6501	0.6388	0.6516
BSMOTE-MLP	0.7154	0.6349	**0.6883**	0.6273	0.6848
ADASYN-MLP	0.7101	0.6640	0.6531	**0.6531**	0.6538
RU-MLP	0.7052	0.6588	0.5707	0.4896	0.5765
SE-MLP	**0.7180**	**0.7145**	0.5489	0.6233	0.5637
ST-MLP	0.7032	0.5853	0.7206	0.6031	0.7117
BMA-SE-MLP	**0.7321**	**0.7201**	0.5987	**0.6570**	0.5834
DT	0.7228	0.0126	**0.9962**	0.1094	**0.9319**
RO-DT	0.7292	0.0176	0.9935	0.1307	0.9298
SMOTE-DT	0.7232	0.0524	0.9833	0.2195	0.9225
BSMOTE-DT	0.7295	0.1353	0.9559	0.3575	0.9023
ADASYN-DT	0.7250	0.0382	0.9859	0.1863	0.9239
RU-DT	**0.7324**	**0.6716**	0.6366	**0.6536**	0.6389
SE-DT	0.7310	0.3410	0.8786	0.5460	0.8435
ST-DT	0.7280	0.0535	0.9827	0.2217	0.9219
BMA-RU-DT	**0.7483**	**0.7039**	0.6404	**0.6680**	0.6445
RF	0.7273	0.0012	**0.9999**	0.0227	**0.9346**
RO-RF	0.7321	0.0008	**0.9999**	0.0156	**0.9346**
SMOTE-RF	0.7236	0.0024	0.9996	0.0384	0.9344
BSMOTE-RF	0.7282	0.0038	0.9991	0.0574	0.9341
ADASYN-RF	0.7239	0.0014	0.9997	0.0250	0.9344
RU-RF	**0.7407**	**0.6995**	0.6612	**0.6800**	0.6637
SE-RF	0.7398	0.1703	0.9580	0.4035	0.9065
ST-RF	0.7250	0.0024	0.9995	0.0422	0.9344
BMA-RU-RF	**0.7514**	**0.7072**	0.6597	0.6798	0.6615
XGB	0.7325	0.0196	0.9962	0.1386	0.9324
RO-XGB	0.7329	0.0179	0.9962	0.1322	0.9322
SMOTE-XGB	0.7347	0.0158	**0.9967**	0.1232	0.9326
BSMOTE-XGB	0.7352	0.0177	0.9965	0.1317	0.9325
ADASYN-XGB	0.7348	0.0172	**0.9967**	0.1290	**0.9327**
RU-XGB	**0.7462**	**0.7022**	0.6626	**0.6820**	0.6653
SE-XGB	0.7428	0.2273	0.9433	0.4627	0.8965
ST-XGB	0.7360	0.0157	0.9967	0.1230	0.9326
BMA-RU-XGB	**0.7600**	**0.7233**	0.6544	0.6881	0.6589

(*continued on next page*)

Table A1 (a) (*continued*)

Model	AUC	Sensitivity	Specificity	G-mean	Accuracy
CAT	0.7283	0.0000	**1.0000**	0.0000	**0.9346**
RO-CAT	0.7291	0.0000	**1.0000**	0.0000	**0.9346**
SMOTE-CAT	0.7146	0.1033	0.9729	0.3124	0.9161
BSMOTE-CAT	0.7319	0.2311	0.9257	0.4615	0.8803
ADASYN-CAT	0.7143	0.0732	0.9822	0.2619	0.9227
RU-CAT	**0.7438**	**0.7291**	0.6352	**0.6804**	0.6413
SE-CAT	0.7332	0.5421	0.7775	0.6489	0.7621
ST-CAT	0.7139	0.1043	0.9722	0.3149	0.9155
BMA-RU-CAT	**0.7597**	**0.7351**	0.6393	**0.6900**	0.6362
LGB	**0.7355**	**0.0223**	0.9955	**0.1468**	0.9320
RO-LGB	0.7475	0.5423	0.7940	0.6561	0.7775
SMOTE-LGB	0.7464	0.0158	0.9960	0.1234	0.9320
BSMOTE-LGB	0.7456	0.0144	0.9960	0.1175	0.9319
ADASYN-LGB	0.7465	0.0127	**0.9967**	0.1100	**0.9324**
RU-LGB	**0.7489**	**0.7045**	0.6648	**0.6842**	0.6674
SE-LGB	0.7463	0.2200	0.9447	0.4556	0.8973
ST-LGB	0.7427	0.0172	0.9946	0.1283	0.9307
BMA-RU-LGB	**0.7635**	**0.7283**	0.6608	**0.6937**	0.6652

Table A1(b)
Evaluation metrics of different baseline classification models (90 days, Period 2)

Model	AUC	Sensitivity	Specificity	G-mean	Accuracy
Logistic	0.6321	0.0000	**1.0000**	0.0000	0.9895
RO-Logistic	0.6598	0.4917	0.7283	0.5973	0.7258
SMOTE-Logistic	0.6575	0.6437	0.5684	0.6042	0.5692
BSMOTE-Logistic	0.6600	0.5813	0.6389	**0.6082**	0.6383
ADASYN-Logistic	0.6576	0.6452	0.5672	0.6044	0.5680
RU-Logistic	0.6526	0.6548	0.5524	0.6003	0.5535
SE-Logistic	**0.6608**	**0.7340**	0.4756	0.5902	0.4783
ST-Logistic	0.6591	0.6488	0.5678	0.6063	0.5687
BMA-SE-Logistic	**0.6764**	**0.7524**	0.4578	**0.5967**	0.4609
KNN	0.6206	0.0000	**1.0000**	0.0000	**0.9895**
RO-KNN	0.6217	0.3552	0.7986	0.5311	0.7940
SMOTE-KNN	0.6318	0.4361	0.7526	0.5719	0.7493
BSMOTE-KNN	0.6310	0.2175	0.9133	0.4439	0.9059
ADASYN-KNN	0.6326	0.4365	0.7516	0.5718	0.7482
RU-KNN	**0.6658**	**0.5544**	0.6741	**0.6100**	0.6729
SE-KNN	0.6364	0.4848	0.7148	0.5880	0.7124
ST-KNN	0.6318	0.4349	0.7527	0.5711	0.7493
BMA-RU-KNN	**0.6836**	**0.5557**	**0.6831**	**0.6122**	**0.6862**
NB	0.6357	0.0000	**0.9999**	0.0000	**0.9894**
RO-NB	0.6472	0.9274	0.2199	0.4515	0.2274
SMOTE-NB	0.6417	0.9238	0.2287	0.4596	0.2360
BSMOTE-NB	**0.6494**	0.8000	0.3937	**0.5606**	0.3980
ADASYN-NB	0.6418	0.9242	0.2275	0.4584	0.2348
RU-NB	0.6459	**0.9250**	0.2278	0.4586	0.2351
SE-NB	0.6425	0.9167	0.2425	0.4713	0.2496
ST-NB	0.6417	0.9238	0.2302	0.4611	0.2375
BMA-BSMOTE-NB	**0.6655**	0.6386	**0.5986**	**0.6182**	**0.5990**
MLP	0.6229	0.0007	**0.9997**	0.0038	**0.9892**
RO-MLP	0.6275	0.4254	0.7270	0.4958	0.7238
SMOTE-MLP	0.6238	0.5845	0.5802	0.4918	0.5803
BSMOTE-MLP	0.6306	0.5028	0.6698	0.5459	0.6680
ADASYN-MLP	0.6280	0.7079	0.4590	0.5030	0.4616
RU-MLP	0.6296	0.5159	0.6201	0.4684	0.6190
SE-MLP	**0.6322**	**0.7164**	0.4822	**0.5852**	0.4845
ST-MLP	0.6269	0.5921	0.5828	0.5351	0.5828
BMA-SE-MLP	**0.6585**	**0.7237**	0.4819	**0.5863**	**0.5003**
DT	0.6358	0.0012	**0.9994**	0.0109	**0.9888**
RO-DT	0.6371	0.3845	0.7740	0.5446	0.7699
SMOTE-DT	0.6368	0.0187	0.9914	0.1149	0.9811
BSMOTE-DT	0.6382	0.0948	0.9695	0.2970	0.9602
ADASYN-DT	0.6367	0.0202	0.9909	0.1313	0.9807
RU-DT	**0.6467**	**0.5619**	0.6046	0.5804	0.6042
SE-DT	0.6347	0.0190	0.9914	0.1186	0.9812
ST-DT	0.6447	0.0190	0.9914	0.1186	0.9812

(*continued on next page*)

Table A1(b) (*continued*)

Model	AUC	Sensitivity	Specificity	G-mean	Accuracy
BMA-RU-DT	**0.6549**	**0.5845**	0.6463	**0.6136**	**0.6457**
RF	0.6373	0.0000	**1.0000**	0.0000	**0.9895**
RO-RF	0.6508	0.0000	**1.0000**	0.0000	**0.9895**
SMOTE-RF	0.6494	0.0000	**1.0000**	0.0000	0.9894
BSMOTE-RF	0.6438	0.0000	0.9999	0.0000	0.9894
ADASYN-RF	0.6521	0.0000	**1.0000**	0.0000	**0.9895**
RU-RF	**0.6699**	**0.6183**	0.6314	**0.6243**	**0.6313**
SE-RF	0.6551	0.0007	**1.0000**	0.0065	**0.9895**
ST-RF	0.6463	0.0000	**1.0000**	0.0000	**0.9895**
BMA-RU-RF	**0.6954**	**0.6272**	**0.6597**	**0.6430**	**0.6615**
XGB	0.6679	0.0007	**0.9998**	0.0065	**0.9893**
RO-XGB	0.6680	0.0012	**0.9998**	0.0109	**0.9893**
SMOTE-XGB	0.6731	0.0004	**0.9998**	0.0036	**0.9893**
BSMOTE-XGB	0.6701	0.0004	0.9997	0.0036	0.9892
ADASYN-XGB	0.6697	0.0004	**0.9998**	0.0036	**0.9893**
RU-XGB	**0.6791**	**0.6148**	0.6271	**0.6203**	0.6270
SE-XGB	0.6775	0.0033	0.9992	0.0241	0.9887
ST-XGB	0.6728	0.0000	0.9999	0.0000	**0.9893**
BMA-RU-XGB	**0.7005**	**0.6276**	**0.6348**	**0.6208**	**0.6345**
CAT	0.6640	0.0000	**1.0000**	0.0000	**1.0000**
RO-CAT	0.6869	0.5066	0.7548	0.6163	0.7521
SMOTE-CAT	0.6780	0.0310	0.9884	0.1625	0.9783
BSMOTE-CAT	0.6735	0.1349	0.9634	0.3560	0.9547
ADASYN-CAT	0.6880	0.0317	0.9878	0.1649	0.9777
RU-CAT	**0.6878**	**0.6540**	0.6031	**0.6262**	0.6037
SE-CAT	0.6796	0.0325	0.9883	0.1690	0.9782
ST-CAT	0.6867	0.0000	**1.0000**	0.0000	**0.9895**
BMA-RU-CAT	**0.7032**	**0.6581**	0.6001	**0.6283**	**0.6203**
LGB	**0.6733**	**0.0043**	0.9992	**0.0367**	0.9887
RO-LGB	0.6979	0.3071	0.8924	0.5221	0.8862
SMOTE-LGB	0.6927	0.0002	0.9998	0.0022	0.9893
BSMOTE-LGB	0.6918	0.0010	0.9997	0.0087	0.9892
ADASYN-LGB	0.6944	0.0000	**0.9999**	0.0000	**0.9894**
RU-LGB	**0.6972**	**0.6364**	0.6472	**0.6412**	0.6471
SE-LGB	0.6910	0.0048	0.9980	0.0423	0.9876
ST-LGB	0.6871	0.0026	0.9989	0.0240	0.9884
BMA-RU-LGB	**0.7191**	**0.6586**	**0.6637**	**0.6607**	**0.6636**

Table A2(a)
Evaluation metrics of different baseline classification models (180 days, Period 1)

Model	AUC	Sensitivity	Specificity	G-mean	Accuracy
Logistic	0.7167	0.0002	**0.9999**	0.0044	**0.9346**
RO-Logistic	0.7323	0.6340	0.7116	0.6716	0.7066
SMOTE-Logistic	0.7325	0.6881	0.6641	**0.6759**	0.6657
BSMOTE-Logistic	0.7318	0.6467	0.7013	0.6734	0.6978
ADASYN-Logistic	0.7326	0.6934	0.6586	0.6757	0.6609
RU-Logistic	0.7337	0.6865	0.6659	0.6760	0.6672
SE-Logistic	**0.7343**	**0.7718**	0.5809	0.6695	0.5934
ST-Logistic	0.7326	0.6845	0.6656	0.6748	0.6668
BMA-SE-Logistic	**0.7462**	**0.7860**	**0.5831**	**0.6726**	**0.5957**
KNN	0.7199	0.0000	**1.0000**	0.0000	**0.9346**
RO-KNN	0.7279	0.6005	0.7342	0.6639	0.7255
SMOTE-KNN	0.7279	0.7189	0.6251	0.6702	0.6312
BSMOTE-KNN	0.7270	0.6459	0.6972	0.6710	0.6939
ADASYN-KNN	0.7265	0.7385	0.6012	0.6662	0.6101
RU-KNN	**0.7285**	0.6493	0.6951	**0.6716**	0.6921
SE-KNN	0.7202	**0.7594**	0.5450	0.6432	0.5590
ST-KNN	0.7200	0.6645	0.6415	0.6528	0.6430
BMA-RU-KNN	**0.7370**	**0.6633**	**0.6994**	**0.6820**	**0.6966**
NB	0.7194	0.0008	**0.9997**	0.0143	**0.9344**
RO-NB	0.7199	**0.9742**	0.0803	0.2794	0.1388
SMOTE-NB	0.7188	0.7596	0.5949	0.6722	0.6057
BSMOTE-NB	0.7177	0.7330	0.6217	0.6729	0.6289
ADASYN-NB	0.7188	0.7614	0.5916	0.6711	0.6027
RU-NB	**0.7200**	0.7534	0.6042	**0.6746**	0.6140
SE-NB	0.7178	0.7593	0.5951	0.6721	0.6058

(*continued on next page*)

Table A2(a) (*continued*)

Model	AUC	Sensitivity	Specificity	G-mean	Accuracy
ST-NB	0.7185	0.7586	0.5960	0.6723	0.6067
BMA-RU-NB	**0.7384**	**0.7613**	**0.6345**	**0.6871**	**0.6398**
MLP	0.7052	0.0008	**0.9986**	0.0055	**0.9334**
RO-MLP	0.7057	0.5734	0.7340	0.5984	0.7235
SMOTE-MLP	0.7055	0.5783	0.7227	0.5935	0.7132
BSMOTE-MLP	**0.7214**	0.6391	0.6880	**0.6418**	0.6848
ADASYN-MLP	0.7140	0.6399	0.6850	0.6390	0.6820
RU-MLP	0.6866	0.6726	0.6063	0.5983	0.6106
SE-MLP	0.7154	**0.8005**	0.5160	0.6136	0.5345
ST-MLP	0.7131	0.6805	0.6373	0.6363	0.6402
BMA-BSMOTE-MLP	**0.7343**	**0.6521**	0.6820	**0.6620**	0.6832
DT	0.7228	0.0106	**0.9962**	0.1010	**0.9317**
RO-DT	0.7253	0.5910	0.7334	0.6580	0.7241
SMOTE-DT	0.7240	0.0581	0.9816	0.2296	0.9212
BSMOTE-DT	0.7294	0.1339	0.9558	0.3548	0.9020
ADASYN-DT	0.7272	0.0393	0.9859	0.1886	0.9240
RU-DT	**0.7342**	**0.6694**	0.6372	**0.6528**	0.6393
SE-DT	0.7279	0.3467	0.8782	0.5503	0.8435
ST-DT	0.7300	0.0575	0.9813	0.2283	0.9209
BMA-RU-DT	**0.7483**	**0.7039**	**0.6404**	**0.6713**	**0.6445**
RF	0.7286	0.0012	**0.9998**	0.0216	**0.9345**
RO-RF	0.7399	0.1229	0.9663	0.3439	0.9112
SMOTE-RF	0.7322	0.0020	0.9996	0.0377	0.9344
BSMOTE-RF	0.7392	0.0044	0.9991	0.0605	0.9341
ADASYN-RF	0.7336	0.0012	0.9996	0.0242	0.9343
RU-RF	**0.7420**	**0.7007**	0.6607	**0.6803**	0.6633
SE-RF	0.7352	0.0022	0.9995	0.0362	0.9343
ST-RF	0.7334	0.0186	0.9964	0.1345	0.9325
BMA-RU-RF	**0.7510**	**0.7198**	0.6554	**0.6878**	0.6596
XGB	0.7329	0.0156	0.9963	0.1235	0.9322
RO-XGB	0.7357	0.0172	**0.9967**	0.1296	**0.9327**
SMOTE-XGB	0.7348	0.0165	0.9966	0.1256	0.9326
BSMOTE-XGB	0.7356	0.0169	0.9966	0.1282	0.9326
ADASYN-XGB	0.7358	0.6783	0.6534	0.6656	0.6550
RU-XGB	**0.7460**	**0.6784**	0.6527	0.6653	0.6544
SE-XGB	0.7341	0.0179	0.9966	0.1316	0.9327
ST-XGB	0.7358	0.0181	0.9966	0.1235	0.9322
BMA-RU-XGB	**0.7617**	**0.7247**	0.6508	0.6866	0.6557
CAT	0.7294	0.0000	**1.0000**	0.0000	**0.9346**
RO-CAT	0.7382	0.6449	0.7188	**0.6807**	0.7140
SMOTE-CAT	0.7329	0.0955	0.9751	0.3024	0.9176
BSMOTE-CAT	0.7319	0.2346	0.9270	0.4655	0.8817
ADASYN-CAT	0.7343	0.0693	0.9827	0.2567	0.9230
RU-CAT	**0.7425**	**0.7284**	0.6343	0.6795	0.6405
SE-CAT	0.7325	0.5424	0.7768	0.6489	0.7615
ST-CAT	0.7343	0.1035	0.9725	0.3108	0.9157
BMA-RU-CAT	**0.7580**	**0.7393**	0.6279	**0.6810**	0.6352
LGB	**0.7362**	0.0216	0.9935	0.1381	0.9317
RO-LGB	0.7481	0.5468	0.7924	0.6581	0.7764
SMOTE-LGB	0.7448	0.0130	0.9961	0.1124	0.9318
BSMOTE-LGB	0.7454	0.0143	0.9961	0.1176	0.9319
ADASYN-LGB	0.7459	0.0117	**0.9966**	0.1050	**0.9323**
RU-LGB	**0.7482**	**0.7052**	0.6626	**0.6835**	0.6654
SE-LGB	0.7465	0.2191	0.9449	0.4547	0.8975
ST-LGB	0.7407	0.0160	0.9942	0.1237	0.9303
BMA-RU-LGB	**0.7628**	**0.7291**	0.6607	**0.6939**	0.6652

Table A2(b)
Evaluation metrics of different baseline classification models (180 days, Period 2)

Model	AUC	Sensitivity	Specificity	G-mean	Accuracy
Logistic	0.6355	0.0000	**1.0000**	0.0000	**0.9895**
RO-Logistic	0.6608	0.4952	0.7287	0.6001	0.7262
SMOTE-Logistic	0.6583	0.6472	0.5664	0.6049	0.5673
BSMOTE-Logistic	0.6602	0.5738	0.6421	0.6057	0.6414
ADASYN-Logistic	0.6590	0.6536	0.5644	**0.6067**	0.5653
RU-Logistic	0.6574	0.6639	0.5503	0.6031	0.5515

(*continued on next page*)

Table A2(b) (*continued*)

Model	AUC	Sensitivity	Specificity	G-mean	Accuracy
SE Logistic	**0.6622**	**0.7374**	0.4793	0.5938	0.4820
ST- Logistic	0.6595	0.6567	0.5663	0.6090	0.5673
BMA-SE-Logistic	**0.6764**	**0.7548**	0.4572	0.5873	0.4604
KNN	0.6198	0.0000	**1.0000**	0.0000	**0.9895**
RO-KNN	0.6122	0.3528	0.7982	0.5293	0.7935
SMOTE-KNN	0.6339	0.4397	0.7536	0.5747	0.7502
BSMOTE-KNN	0.6331	0.2206	0.9142	0.4470	0.9069
ADASYN-KNN	0.6344	0.4397	0.7513	0.5739	0.7480
RU-KNN	**0.6631**	**0.5635**	0.6689	**0.6076**	0.6677
SE-KNN	0.6352	0.4848	**0.7147**	0.5879	0.7123
ST-KNN	0.6338	0.4393	0.7533	0.5744	0.7500
BMA-RU-KNN	**0.6878**	**0.5717**	0.6643	**0.6153**	**0.6918**
NB	0.6342	0.0000	**0.9999**	0.0000	**0.9894**
RO-NB	0.6472	**0.9282**	0.2200	0.4518	0.2275
SMOTE-NB	0.6416	0.9242	0.2289	0.4599	0.2362
BSMOTE-NB	**0.6495**	0.7952	0.3995	**0.5630**	0.4037
ADASYN-NB	0.6416	0.9246	0.2275	0.4586	0.2349
RU-NB	0.6451	0.9254	0.2251	0.4560	0.2325
SE-NB	0.6426	0.9167	0.2426	0.4714	0.2498
ST-NB	0.6416	0.9238	0.2304	0.4613	0.2377
BMA-BSMOTE-NB	**0.6673**	**0.8039**	**0.4001**	**0.5705**	**0.4230**
MLP	0.6239	0.0014	**0.9996**	0.0053	**0.9891**
RO-MLP	0.6275	0.4254	0.7270	0.4958	0.7238
SMOTE-MLP	0.6248	0.5845	0.5802	0.4918	0.5803
BSMOTE-MLP	0.6306	0.5028	0.6698	**0.5459**	0.6680
ADASYN-MLP	0.6280	0.7079	0.4590	0.5030	0.4616
RU-MLP	0.6316	0.5159	0.6201	0.4684	0.6190
SE-MLP	**0.6325**	**0.7326**	0.4458	0.5169	0.4488
ST-MLP	0.6269	0.5921	0.5828	0.5351	0.5828
BMA-SE-MLP	**0.6620**	**0.7421**	0.4421	**0.5660**	**0.4599**
DT	0.6375	0.0007	**0.9993**	0.0065	**0.9887**
RO-DT	0.6531	0.3845	0.7740	0.5446	0.7699
SMOTE-DT	0.6458	0.0187	0.9914	0.1149	0.9811
BSMOTE-DT	0.6482	0.0948	0.9695	0.2970	0.9602
ADASYN-DT	0.6467	0.0202	0.9909	0.1313	0.9807
RU-DT	**0.6567**	**0.5619**	0.6046	**0.5804**	0.6042
SE-DT	0.6433	0.0464	0.9790	0.1997	0.9692
ST-DT	0.6507	0.0190	0.9914	0.1186	0.9812
BMA-RU-DT	**0.6643**	**0.5833**	**0.6290**	**0.6027**	**0.6285**
RF	0.6581	0.0000	**1.0000**	0.0000	**0.9895**
RO-RF	0.6608	0.0000	**1.0000**	0.0000	**0.9895**
SMOTE-RF	0.6694	0.0000	**1.0000**	0.0000	0.9894
BSMOTE-RF	0.6638	0.0000	0.9999	0.0000	0.9894
ADASYN-RF	0.6721	0.0000	**1.0000**	0.0000	**0.9895**
RU-RF	**0.6799**	**0.6183**	0.6314	**0.6243**	0.6313
SE-RF	0.6733	0.0014	**1.0000**	0.0131	0.9894
ST-RF	0.6663	0.0000	**1.0000**	0.0000	**0.9895**
BMA-RU-RF	**0.6912**	**0.6333**	**0.6431**	**0.6368**	**0.6430**
XGB	0.6681	0.0000	0.9998	0.0000	**0.9893**
RO-XGB	0.6780	0.0348	0.9998	0.1715	**0.9893**
SMOTE-XGB	0.6731	0.0017	**0.9999**	0.0140	**0.9893**
BSMOTE-XGB	0.6701	0.0007	0.9998	0.0037	0.9892
ADASYN-XGB	0.6805	0.0010	**0.9999**	0.0087	**0.9893**
RU-XGB	**0.6879**	**0.6143**	0.6271	**0.6203**	0.6270
SE-XGB	0.6782	0.0043	0.9993	0.0354	0.9888
ST-XGB	0.6728	0.0000	**0.9999**	0.0000	**0.9893**
BMA-RU-XGB	**0.7079**	**0.6286**	**0.6406**	**0.6327**	**0.6405**
CAT	0.6650	0.0000	**1.0000**	0.0000	**0.9895**
RO-CAT	0.6820	0.4900	0.7711	0.6132	0.7681
SMOTE-CAT	0.6747	0.0310	**0.9884**	0.1625	**0.9783**
BSMOTE-CAT	0.6735	0.1349	0.9634	0.3560	0.9547
ADASYN-CAT	0.6813	0.1455	0.9671	0.3701	0.9527
RU-CAT	**0.6855**	**0.6521**	0.6104	**0.6292**	0.6109
SE-CAT	0.6798	0.0431	0.9841	0.1984	0.9782
ST-CAT	0.6757	0.0343	0.9875	0.1754	0.9774
BMA-RU-CAT	**0.7029**	**0.6633**	0.6071	**0.6393**	0.5881
LGB	**0.6779**	0.0048	0.9992	0.0372	0.9887
RO-LGB	0.6968	0.2948	0.8950	0.5122	0.8886
SMOTE-LGB	0.6961	0.0005	**0.9999**	0.0044	**0.9894**
BSMOTE-LGB	0.6906	0.0005	0.9996	0.0044	0.9891

(*continued on next page*)

Table A2(b) (*continued*)

Model	AUC	Sensitivity	Specificity	G-mean	Accuracy
ADASYN-LGB	0.6950	0.0002	**0.9999**	0.0022	0.9893
RU-LGB	**0.6977**	**0.6362**	0.6469	**0.6410**	0.6468
SMOTEENN-LGB	0.6919	0.0019	0.9980	0.0162	0.9875
SMOTETOMK-LGB	0.6851	0.0014	0.9989	0.0131	0.9884
BMA-RU-LGB	**0.7178**	**0.6555**	**0.6599**	**0.6571**	**0.6598**

Table A3(a)
Evaluation metrics of different baseline classification models (360 days, Period 1)

Model	AUC	Sensitivity	Specificity	G-mean	Accuracy
Logistic	0.7172	0.0002	**1.0000**	0.0035	**0.9346**
RO-Logistic	0.7327	0.6329	0.7118	0.6711	0.7066
SMOTE-Logistic	0.7322	0.6888	0.6630	0.6757	0.6647
BSMOTE-Logistic	0.7318	0.6467	0.7014	0.6734	0.6978
ADASYN-Logistic	0.7321	0.6940	0.6585	0.6759	0.6609
RU-Logistic	0.7328	0.6882	0.6656	**0.6768**	0.6671
SE-Logistic	**0.7344**	**0.7715**	0.5802	0.6689	0.5927
ST-Logistic	0.7328	0.6876	0.6650	0.6761	0.6665
BMA-SE-Logistic	**0.7462**	**0.7772**	**0.5824**	**0.6726**	**0.5952**
KNN	0.7196	0.0000	**1.0000**	0.0000	**0.9346**
RO-KNN	0.7295	0.6008	0.7349	0.6643	0.7261
SMOTE-KNN	0.7272	0.7191	0.6235	0.6695	0.6297
BSMOTE-KNN	0.7268	0.6454	0.6970	0.6706	0.6936
ADASYN-KNN	0.7262	0.7376	0.6007	0.6655	0.6097
RU-KNN	0.7285	0.6538	0.6921	**0.6726**	0.6896
SE-KNN	**0.7304**	**0.8035**	0.5328	0.6542	0.5505
ST-KNN	0.7273	0.7190	0.6239	0.6697	0.6301
BMA-SE-KNN	**0.7464**	0.7856	**0.5975**	**0.6860**	**0.6981**
NB	0.7195	0.0008	**0.9997**	0.0144	**0.9344**
RO-NB	0.7199	0.7565	0.6012	0.6743	0.6113
SMOTE-NB	0.7188	0.7595	0.5950	0.6722	0.6058
BSMOTE-NB	0.7177	0.7331	0.6218	0.6751	0.6291
ADASYN-NB	0.7188	**0.7614**	0.5915	0.6710	0.6026
RU-NB	**0.7202**	0.7537	0.6040	**0.6756**	0.6138
SE-NB	0.7178	0.7592	0.5952	0.6721	0.6059
ST-NB	0.7185	0.7586	0.5961	0.6724	0.6067
BMA-RU-NB	**0.7373**	**0.7602**	0.6023	0.6740	**0.6388**
MLP	0.7052	0.0008	**0.9986**	0.0055	**0.9334**
RO-MLP	0.7068	0.6072	0.6862	0.5958	0.6811
SMOTE-MLP	0.7102	0.6531	0.6538	0.6090	0.6537
BSMOTE-MLP	0.7183	0.6218	0.7073	0.6356	0.7018
ADASYN-MLP	0.7142	0.6317	0.6917	0.6337	0.6878
RU-MLP	0.7156	0.6704	0.5705	0.5352	0.5771
SE-MLP	**0.7276**	**0.7659**	0.5763	**0.6524**	0.5887
ST-MLP	0.7019	0.7086	0.5947	0.6073	0.6021
BMA-SE-MLP	**0.7391**	**0.7720**	0.5740	**0.6650**	0.6003
DT	0.7213	0.0104	**0.9964**	0.0992	**0.9319**
RO-DT	0.7234	0.5852	0.7356	0.6559	0.7258
SMOTE-DT	**0.7283**	0.0558	0.9808	0.2264	0.9203
BSMOTE-DT	0.7223	0.1322	0.9574	0.3538	0.9035
ADASYN-DT	0.7281	0.0321	0.9896	0.1710	0.9270
RU-DT	0.7230	**0.6748**	0.6384	**0.6607**	0.6314
SE-DT	0.7278	0.3381	0.8811	0.5444	0.8456
ST-DT	0.7279	0.0555	0.9809	0.2267	0.9204
BMA-RU-DT	**0.7421**	**0.6961**	**0.6526**	**0.6737**	**0.6554**
RF	0.7283	0.0008	**0.9998**	0.0158	**0.9345**
RO-RF	0.7288	0.1214	0.9667	0.3420	0.9115
SMOTE-RF	0.7254	0.0022	0.9995	0.0373	0.9343
BSMOTE-RF	0.7299	0.0054	0.9992	0.0692	0.9343
ADASYN-RF	0.7241	0.0021	0.9996	0.0370	0.9344
RU-RF	**0.7408**	**0.6996**	0.6594	**0.6792**	0.6621
SE-RF	0.7402	0.1694	0.9582	0.4026	0.9066
ST-RF	0.7265	0.0026	0.9995	0.0437	0.9344
BMA-RU-RF	**0.7540**	**0.7179**	**0.6613**	**0.6830**	**0.6650**
XGB	0.7344	0.0187	0.9964	0.1351	0.9325
RO-XGB	0.7167	0.3479	0.8770	0.5521	0.8424
SMOTE-XGB	0.7352	0.0157	**0.9967**	0.1236	0.9325

(*continued on next page*)

Table A3(a) (*continued*)

Model	AUC	Sensitivity	Specificity	G-mean	Accuracy
BSMOTE-XGB	0.7357	0.0189	0.9965	0.1358	0.9326
ADASYN-XGB	0.7354	0.0162	0.9965	0.1259	0.9325
RU-XGB	0.7400	**0.6748**	0.6528	**0.6637**	0.6543
SE-XGB	**0.7429**	0.2291	0.9427	0.4644	0.8961
ST-XGB	0.7360	0.0183	0.9966	0.1336	**0.9327**
BMA-RU-XGB	**0.7600**	**0.7155**	**0.6554**	**0.6847**	**0.6593**
CAT	0.7294	0.0000	**1.0000**	0.0000	**0.9346**
RO-CAT	**0.7432**	0.6457	0.7199	0.6816	0.7150
SMOTE-CAT	0.7334	0.0913	0.9760	0.2955	0.9182
BSMOTE-CAT	0.7323	0.2374	0.9261	0.4681	0.8811
ADASYN-CAT	0.7330	0.0723	0.9826	0.2616	0.9231
RU-CAT	0.7425	**0.7285**	0.6363	**0.6827**	0.6423
SE-CAT	0.7321	0.5454	0.7735	0.6493	0.7586
ST-CAT	0.7345	0.1020	0.9727	0.3083	0.9158
BMA-RU-CAT	**0.7598**	**0.7351**	**0.6393**	**0.6900**	0.6362
LGB	**0.7360**	0.0204	0.9936	0.1410	0.9300
RO-LGB	0.7489	0.5459	0.7906	0.6569	0.7746
SMOTE-LGB	0.7457	0.0146	0.9959	0.1180	0.9318
BSOMTE-LGB	0.7454	0.0143	0.9961	0.1176	0.9319
ADASYN-LGB	0.7470	0.0137	**0.9965**	0.1148	**0.9323**
RU-LGB	**0.7485**	**0.7052**	0.6626	**0.6835**	0.6654
SE-LGB	0.7457	0.2198	0.9451	0.4554	0.8977
ST-LGB	0.7418	0.0181	0.9942	0.1322	0.9304
BMA-RU-LGB	**0.7627**	**0.7272**	0.6596	**0.6925**	0.6640

Table A3(b)
Evaluation metrics of different baseline classification models (360 days, Period 2)

Model	AUC	Sensitivity	Specificity	G-mean	Accuracy
Logistic	0.6338	0.0000	**1.0000**	0.0000	**0.9895**
RO-Logistic	0.6605	0.4921	0.7286	0.5977	0.7261
SMOTE-Logistic	0.6585	0.6524	0.5689	**0.6084**	0.5698
BSMOTE-Logistic	0.6597	0.5718	0.6423	0.6050	0.6416
ADASYN-Logistic	0.6574	0.6520	0.5650	0.6061	0.5659
RU-Logistic	0.6563	0.6560	0.5571	0.6037	0.5582
SE-Logistic	**0.6613**	**0.7386**	0.4751	0.5918	0.4779
ST-Logistic	0.6604	0.6528	0.5668	0.6074	0.5677
BMA-SE-Logistic	**0.6766**	**0.7583**	0.4554	0.5875	0.4586
KNN	0.6181	0.0000	**1.0000**	0.0000	**0.9895**
RO-KNN	0.6192	0.3528	0.7982	0.5293	0.7935
SMOTE-KNN	0.6339	0.4397	0.7536	0.5747	0.7502
BSMOTE-KNN	0.6331	0.2206	0.9142	0.4470	0.9069
ADASYN-KNN	0.6344	0.4397	0.7513	0.5739	0.7480
RU-KNN	**0.6700**	**0.5635**	0.6708	**0.6137**	0.6696
SE-KNN	0.6362	0.4852	0.7150	0.5883	0.7126
ST-KNN	0.6338	0.4393	0.7533	0.5744	0.7500
BMA-RU-KNN	**0.6829**	0.5362	**0.7009**	**0.6163**	**0.7064**
NB	0.6327	0.0019	**0.9997**	0.0175	**0.9892**
RO-NB	0.6472	**0.9282**	0.2200	0.4518	0.2275
SMOTE-NB	0.6416	0.9242	0.2289	0.4599	0.2362
BSMOTE-NB	**0.6495**	0.7952	0.3995	**0.5630**	0.4037
ADASYN-NB	0.6416	0.9246	0.2275	0.4586	0.2349
RU-NB	0.6451	0.9254	0.2251	0.4560	0.2325
SE-NB	0.6424	0.9160	0.2435	0.4721	0.2506
ST-NB	0.6416	0.9238	0.2304	0.4613	0.2377
BMA-BSMOTE-NB	**0.6657**	**0.8064**	**0.4026**	**0.5680**	**0.5664**
MLP	0.6250	0.0000	**1.0000**	0.0000	**0.9895**
RO-MLP	0.6344	0.0000	1.0000	0.0000	0.9895
SMOTE-MLP	0.6341	0.6139	0.5576	0.5104	0.5582
BSMOTE-MLP	0.6350	0.5230	0.6360	0.5357	0.6349
ADASYN-MLP	0.6430	0.5821	0.5680	0.4761	0.5681
RU-MLP	0.6389	0.5909	0.5194	0.4400	0.5202
SE-MLP	**0.6445**	**0.7260**	0.4558	0.5341	0.4587
ST-MLP	0.6333	0.6349	0.5434	**0.5428**	0.5443
BMA-SE-MLP	**0.6679**	**0.7364**	0.4490	**0.5740**	**0.4677**
DT	0.6339	0.0014	**0.9993**	0.0131	**0.9887**
RO-DT	0.6515	0.3754	0.7740	0.5378	0.7698

(*continued on next page*)

Table A3(b) (*continued*)

Model	AUC	Sensitivity	Specificity	G-mean	Accuracy
SMOTE-DT	0.6453	0.0190	0.9909	0.1188	0.9807
BSMOTE-DT	0.6462	0.0770	0.9730	0.2685	0.9635
ADASYN-DT	0.6418	0.0214	0.9904	0.1297	0.9802
RU-DT	**0.6521**	**0.5647**	0.5924	**0.5773**	0.5921
SE-DT	0.6483	0.0538	0.9787	0.2171	0.9689
ST-DT	0.6460	0.0198	0.9909	0.1218	0.9807
BMA-RU-DT	**0.6721**	**0.5893**	**0.6459**	**0.6161**	**0.6453**
RF	0.6587	0.0000	**1.0000**	0.0000	**0.9895**
RO-RF	0.6659	0.0040	0.9982	0.0342	0.9877
SMOTE-RF	0.6675	0.0000	**1.0000**	0.0000	**0.9895**
BSMOTE-RF	0.6647	0.0000	0.9999	0.0000	0.9894
ADASYN-RF	0.6753	0.0004	**1.0000**	0.0036	**0.9895**
RU-RF	**0.6813**	**0.6139**	0.6348	**0.6235**	0.6346
SE-RF	0.6763	0.0010	**1.0000**	0.0087	0.9894
ST-RF	0.6714	0.0000	**1.0000**	0.0000	**0.9895**
BMA-RU-RF	**0.6924**	**0.6204**	0.6597	**0.6310**	0.6208
XGB	0.6684	0.0005	0.9998	0.0044	**0.9893**
RO-XGB	0.6773	0.0341	0.9878	0.1730	0.9778
SMOTE-XGB	0.6795	0.0000	**0.9999**	0.0000	**0.9893**
BSMOTE-XGB	0.6770	0.0008	0.9998	0.0073	**0.9893**
ADASYN-XGB	0.6814	0.0004	0.9998	0.0036	**0.9893**
RU-XGB	**0.6834**	**0.6222**	0.6273	**0.6241**	0.6273
SE-XGB	0.6743	0.0031	0.9992	0.0271	0.9887
ST-XGB	0.6726	0.0012	**0.9999**	0.0109	**0.9893**
BMA-RU-XGB	**0.7023**	**0.6381**	**0.6431**	**0.6401**	**0.6430**
CAT	0.6640	0.0000	**1.0000**	0.0000	**0.9895**
RO-CAT	0.6830	0.4905	0.7716	0.6133	0.7686
SMOTE-CAT	0.6732	0.0313	0.9884	0.1631	0.9783
BSMOTE-CAT	0.6747	0.1516	0.9610	0.3769	0.9525
ADASYN-CAT	0.6810	0.0329	0.9874	0.1675	0.9773
RU-CAT	**0.6852**	**0.6591**	0.6041	**0.6291**	0.6047
SE-CAT	0.6745	0.0438	0.9838	0.1984	0.9739
ST-CAT	0.6768	0.0341	0.9876	0.1742	0.9775
BMA-RU-CAT	**0.7062**	**0.6624**	**0.6105**	**0.6360**	**0.6109**
LGB	**0.6780**	0.0026	0.9993	0.0240	0.9888
RO-LGB	0.6974	0.3086	0.8917	0.5227	0.8856
SMOTE-LGB	0.6944	0.0001	**0.9999**	0.0011	**0.9894**
BS-LGB	0.6923	0.0011	0.9997	0.0098	0.9892
ADASYN-LGB	0.6950	0.0001	**0.9999**	0.0011	**0.9894**
RU-LGB	**0.6990**	**0.6417**	0.6488	**0.6447**	0.6487
SE-LGB	0.6909	0.0064	0.9979	0.0512	0.9875
ST-LGB	0.6851	0.0019	0.9990	0.0174	0.9884
BMA-RU-LGB	**0.7153**	**0.6543**	**0.6602**	**0.6567**	**0.6602**

Data availability

Data will be made available on request.

References

Abedin, M.Z., Hassan, M.K., Khan, I., Julio, I.F., 2022. Feature transformation for corporate tax default prediction: application of machine learning approaches. Asia-Pac. J. Oper. Res. 39 (04), 2140017.

Abedin, M.Z., Hajek, P., Sharif, T., Satu, M.S., Khan, M.I., 2023b. Modelling bank customer behaviour using feature engineering and classification techniques. Res. Int. Bus. Financ. 65, 101913.

Abedin, M.Z., Guotai, C., Hajek, P., Zhang, T., 2023a. Combining weighted SMOTE with ensemble learning for the class-imbalanced prediction of small business credit risk. Complex Intell. Syst. 9, 3559–3579.

Alvarez-Melis, D., Jaakkola, T.S., 2018. On the robustness of interpretability methods. arXiv preprint arXiv:1806.08049.

Baniecki, H., Biecek, P., 2024. Adversarial attacks and defenses in explainable artificial intelligence: a survey. Inf. Fusion 107, 102303.

Borchert, P., Coussement, K., De Caigny, A., De Weerdt, J., 2023. Extending business failure prediction models with textual website content using deep learning. Eur. J. Oper. Res. 306 (1), 348–357.

Burtch, G., Ghose, A., Wattal, S., 2014. Cultural differences and geography as determinants of online prosocial lending. MIS Q. 38 (3), 773–794.

Butler, A.W., Cornaggia, J., Gurun, U.G., 2017. Do local capital market conditions affect consumers' borrowing decisions? Manag. Sci. 63 (12), 4175–4187.

Cai, S., Lin, X., Xu, D., Fu, X., 2016. Judging online peer-to-peer lending behavior: a comparison of first-time and repeated borrowing requests. Inf. Manag. 53 (7), 857–867.

Che, W., Wang, Z., Jiang, C., Abedin, M.Z., 2024. Predicting financial distress using multimodal data: an attentive and regularized deep learning method. Inf. Process. Manag. 61 (4), 103703.

Chen, D., Ye, J., Ye, W., 2023b. Interpretable selective learning in credit risk. Res. Int. Bus. Financ. 65, 101940.

Chen, H., Covert, I.C., Lundberg, S.M., Lee, S.I., 2023a. Algorithms to estimate Shapley value feature attributions. Nat. Mach. Intell. 5, 590–601.

Chen, Y., Calabrese, R., Martin-Barragan, B., 2024. Interpretable machine learning for imbalanced credit scoring datasets. Eur. J. Oper. Res. 312 (1), 357–372.
Chi, G., Uddin, M.S., Abedin, M.Z., Yuan, K., 2019. Hybrid model for credit risk prediction: an application of neural network approaches. Int. J. Artif. Intell. Tools 28 (05), 1950017.
Dai, Q., Liu, J.W., Shi, Y.H., 2023. Class-overlap undersampling based on Schur decomposition for Class-imbalance problems. Expert Syst. Appl. 221, 119735.
Ding, J., Tarokh, V., Yang, Y., 2018. Model selection techniques: an overview. IEEE Signal Process. Mag. 35 (6), 16–34.
Ding, S., Cui, T., Bellotti, A.G., Abedin, M.Z., Lucey, B., 2023. The role of feature importance in predicting corporate financial distress in pre and post COVID periods: evidence from China. Int. Rev. Financ. Anal. 90, 102851.
Djeundje, V.B., Crook, J., Calabrese, R., Hamid, M., 2021. Enhancing credit scoring with alternative data. Expert Syst. Appl. 163, 113766.
Elkan, C., 2001. The foundations of cost-sensitive learning. International Joint Conference on Artificial Intelligence. Lawrence Erlbaum Associates Ltd.
Fatima, S.S., Wooldridge, M., Jennings, N.R., 2008. A linear approximation method for the Shapley value. Artif. Intell. 172 (14), 1673–1699.
Figini, S., Giudici, P., 2017. Credit risk assessment with Bayesian model averaging. Commun. Stat. -Theory Methods 46 (19), 9507–9517.
Fitzpatrick, T., Mues, C., 2021. How can lenders prosper? Comparing machine learning approaches to identify profitable peer-to-peer loan investments. Eur. J. Oper. Res. 294 (2), 711–722.
Gao, Q., Lin, M., Sias, R., 2023. Words matter: The role of readability, tone, and deception cues in online credit markets. J. Financ. Quant. Anal. 58 (1), 1–28.
Gao, R., Yao, X., Wang, Z., Abedin, M.Z., 2024. Sentiment classification of time-sync comments: a semi-supervised hierarchical deep learning method. Eur. J. Oper. Res. 314 (3), 1159–1173.
Ge, R., Feng, J., Gu, B., Zhang, P., 2017. Predicting and deterring default with social media information in peer-to-peer lending. J. Manag. Inf. Syst. 34 (2), 401–424.
Ghosh, I., Jana, R.K., Abedin, M.Z., 2023. An ensemble machine learning framework for Airbnb rental price modeling without using amenity-driven features. Int. J. Contemp. Hosp. Manag. 35 (10), 3592–3611.
Goel, A., Rastogi, S., 2023. Understanding the impact of borrowers' behavioural and psychological traits on credit default: review and conceptual model. Rev. Behav. Financ. 15 (2), 205–223.
Gunnarsson, B.R., Vanden Broucke, S., Baesens, B., Óskarsdóttir, M., Lemahieu, W., 2021. Deep learning for credit scoring: do or don't? Eur. J. Oper. Res. 295 (1), 292–305.
Guotai, C., Abedin, M.Z., Moula, F.E., 2017. Modeling credit approval data with neural networks: an experimental investigation and optimization. J. Bus. Econ. Manag. 18 (2), 224–240.
Hajek, P., Abedin, M.Z., Sivarajah, U., 2023. Fraud detection in mobile payment systems using an XGBoost-based framework. Inf. Syst. Front. 25, 1985–2003.
Hasan, M., Noor, T., Gao, J., Usman, M., Abedin, M.Z., 2023. Rural consumers' financial literacy and access to FinTech services. J. Knowl. Econ. 14, 780–804.
Hasan, M., Abedin, M.Z., Hajek, P., Coussement, K., Sultan, M.N., Lucey, B., 2024. A blending ensemble learning model for crude oil price forecasting. Ann. Oper. Res. https://doi.org/10.1007/s10479-023-05810-8.
Jiang, C., Wang, Z., Wang, R., Ding, Y., 2018. Loan default prediction by combining soft information extracted from descriptive text in online peer-to-peer lending. Ann. Oper. Res. 266 (1-2), 511–529.
Jiang, P., Liu, Z., Abedin, M.Z., Wang, J., Yang, W., Dong, Q., 2024. Profit-driven weighted classifier with interpretable ability for customer churn prediction. Omega 125, 103034.
Korangi, K., Mues, C., Bravo, C., 2023. A transformer-based model for default prediction in mid-cap corporate markets. Eur. J. Oper. Res. 308 (1), 306–320.
Kowalewski, O., Pisany, P., 2022. Banks' consumer lending reaction to fintech and bigtech credit emergence in the context of soft versus hard credit information processing. International Review of Financial. Analysis 81, 102116.
Kriebel, J., Stitz, L., 2022. Credit default prediction from user-generated text in peer-to-peer lending using deep learning. Eur. J. Oper. Res. 302 (1), 309–323.
Li, B., Xiao, B., Yang, Y., 2021. Strengthen credit scoring system of small and micro businesses with soft information: analysis and comparison based on neural network models. J. Intell. Fuzzy Syst. 40 (3), 4257–4274.
Li, Z., Li, A., Bellotti, A., Yao, X., 2023. The profitability of online loans: a competing risks analysis on default and prepayment. Eur. J. Oper. Res. 306 (2), 968–985.
Li, Z., Liang, S., Pan, X., Pang, M., 2024. Credit risk prediction based on loan profit: evidence from Chinese SMEs. Res. Int. Bus. Financ. 67, 102155.
Liberti, J.M., Petersen, M.A., 2019. Information: hard and soft. Review of Corporate Finance. Studies 8 (1), 1–41.
Lin, M., Prabhala, N.R., Viswanathan, S., 2013. Judging borrowers by the company they keep: friendship networks and information asymmetry in online peer-to-peer lending. Manag. Sci. 59 (1), 17–35.
Loutfi, A.A., 2022. A framework for evaluating the business deployability of digital footprint based models for consumer credit. J. Bus. Res. 152, 473–486.
Lu, T., Zhang, Y., & Li, B. , 2019. The value of alternative data in credit risk prediction: Evidence from a large field experiment. ICIS 2019 Proceedings, 10, pp. 1-16.
Lu, Y., Yang, L., Shi, B., Li, J., Abedin, M.Z., 2022. A novel framework of credit risk feature selection for SMEs during industry 4.0. Ann. Oper. Res. https://doi.org/10.1007/s10479-022-04849-3.
Lundberg, S.M., Lee, S.I., 2017. A unified approach to interpreting model predictions. Adv. Neural Inf. Process. Syst. 30, 1–10.
Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Lee, S.I., 2020. From local explanations to global understanding with explainable AI for trees. Nat. Mach. Intell. 2 (1), 56–67.
Luo, C., Zhou, X., Lev, B., 2022. Core, shapley value, nucleolus and nash bargaining solution: a survey of recent developments and applications in operations management. Omega 110, 102638.
Ma, L., Zhao, X., Zhou, Z., Liu, Y., 2018. A new aspect on P2P online lending default prediction using meta-level phone usage data in China. Decis. Support Syst. 111, 60–71.
Mahbobi, M., Kimiagari, S., Vasudevan, M., 2023. Credit risk classification: an integrated predictive accuracy algorithm using artificial and deep neural networks. Ann. Oper. Res. 330, 609–637.
Moula, F.E., Guotai, C., Abedin, M.Z., 2017. Credit default prediction modeling: an application of support vector machine. Risk Manag. 19, 158–187.
Papouskova, M., Hajek, P., 2019. Two-stage consumer credit risk modelling using heterogeneous ensemble learning. Decis. Support Syst. 118, 33–45.
Qian, H., Ma, P., Gao, S., Song, Y., 2023. Soft reordering one-dimensional convolutional neural network for credit scoring. Knowl. -Based Syst. 266, 110414.
Rieger, L., & Hansen, L.K. , 2020. A simple defense against adversarial attacks on heatmap explanations. arXiv preprint arXiv:2007.06381.
M.T. Ribeiro S. Singh C. Guestrin Why should I trust you?" Explaining the predictions of any classifier Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2016 1135 1144.
Rozo, B.J.G., Crook, J., Andreeva, G., 2023. The role of web browsing in credit risk prediction. Decis. Support Syst. 164, 113879.
Seiffert, C., Khoshgoftaar, T.M., Van Hulse, J., Napolitano, A., 2009. RUSBoost: a hybrid approach to alleviating class imbalance. IEEE Trans. Syst., Man, Cybern. - Part A Syst. Hum. 40 (1), 185–197.
Shajalal, M., Hajek, P., Abedin, M.Z., 2023. Product backorder prediction using deep neural network on imbalanced data. Int. J. Prod. Res. 61 (1), 302–319.
Sigrist, F., Leuenberger, N., 2023. Machine learning for corporate default risk: multi-period prediction, frailty correlation, loan portfolios, and tail probabilities. Eur. J. Oper. Res. 305 (3), 1390–1406.
Soltanzadeh, P., Hashemzadeh, M., 2021. RCSMOTE: range-Controlled synthetic minority over-sampling technique for handling the class imbalance problem. Inf. Sci. 542, 92–111.
Song, Y., Wang, Y., Ye, X., Zaretzki, R., Liu, C., 2023. Loan default prediction using a credit rating-specific and multi-objective ensemble learning scheme. Inf. Sci. 629, 599–617.
Sun, W., Zhang, X., Li, M., Wang, Y., 2023. Interpretable high-stakes decision support system for credit default forecasting. Technol. Forecast. Soc. Change 196, 122825.
Wang, D., Wang, X., Abedin, M.Z., Wang, S., Yin, Y., 2023. Interpretable multi-hop knowledge reasoning for gastrointestinal disease. Ann. Oper. Res. https://doi.org/10.1007/s10479-023-05650-6.
Wang, G., Chen, G., Zhao, H., Zhang, F., Yang, S., Lu, T., 2021. Leveraging multisource heterogeneous data for financial risk prediction: a novel hybrid-strategy-based self-adaptive method. MIS Q. 45 (4), 1949–19998.

Wang, J., Wang, Y., Wu, C., Yang, X., Zhao, L., 2024. Social proximity, information, and incentives in local bank lending. Rev. Corp. Finance Stud. 13 (1), 80–146.
Wang, Y., Jia, Y., Tian, Y., Xiao, J., 2022a. Deep reinforcement learning with the confusion-matrix-based dynamic reward function for customer credit scoring. Expert Syst. Appl. 200, 117013.
Wang, Y., Drabek, Z., Wang, Z., 2022b. The role of social and psychological related soft information in credit analysis: evidence from a Fintech Company. J. Behav. Exp. Econ. 96, 101806.
Wang, Z., Crook, J., Andreeva, G., 2020a. Reducing estimation risk using a Bayesian posterior distribution approach: application to stress testing mortgage loan default. Eur. J. Oper. Res. 287 (2), 725–738.
Wang, Z., Wang, H., Ramkumar, S., Mardziel, P., Fredrikson, M., Datta, A., 2020b. Smoothed geometry for robust attribution. Adv. Neural Inf. Process. Syst. 33, 13623–13634.
Weng, F., Zhu, J., Yang, C., Gao, W., Zhang, H., 2022. Analysis of financial pressure impacts on the health care industry with an explainable machine learning method: China versus the USA. Expert Syst. Appl. 210, 118482.
Wu, Y., Ke, W., Li, Y., Lin, Z., Tan, Y., 2024. Understanding lenders' investment behavior in online peer-to-peer lending: a construal level theory perspective. Inf. Syst. Res. https://doi.org/10.1287/isre.2020.0428.
Xia, Y., Liu, C., Liu, N., 2017. Cost-sensitive boosted tree for loan evaluation in peer-to-peer lending. Electron. Commer. Res. Appl. 24, 30–49.
Xiao, J., Zhong, Y., Jia, Y., Wang, Y., Li, R., Jiang, X., Wang, S., 2024. A novel deep ensemble model for imbalanced credit scoring in internet finance. Int. J. Forecast. 40 (1), 348–372.
Yang, C., Abedin, M.Z., Zhang, H., Weng, F., Hajek, P., 2023. An interpretable system for predicting the impact of COVID-19 government interventions on stock market sectors. Ann. Oper. Res. 1–28. https://doi.org/10.1007/s10479-023-05311-8.
Yang, F., Qiao, Y., Abedin, M.Z., Huang, C., 2022. Privacy-preserved credit data sharing integrating blockchain and federated learning for industrial 4.0. IEEE Trans. Ind. Inform. 18 (12), 8755–8764.
Yang, F., Abedin, M.Z., Hajek, P., 2024. An explainable federated learning and blockchain-based secure credit modeling method. Eur. J. Oper. Res. 317 (2), 449–467.
Yin, W., Kirkulak-Uludag, B., Zhu, D., Zhou, Z., 2023. Stacking ensemble method for personal credit risk assessment in Peer-to-Peer lending. Appl. Soft Comput. 142, 110302.
Zha, Y., Wang, Y., Li, Q., Yao, W., 2022. Credit offering strategy and dynamic pricing in the presence of consumer strategic behavior. Eur. J. Oper. Res. 303 (2), 753–766.
Zhang, H., Shi, Y., Yang, X., Zhou, R., 2021. A firefly algorithm modified support vector machine for the credit risk assessment of supply chain finance. Res. Int. Bus. Financ. 58, 101482.
Zhang, L., Wang, J., Liu, Z., 2023. What should lenders be more concerned about? Developing a profit-driven loan default prediction model. Expert Syst. Appl. 213, 118938.
Zhang, X., Liu, C.A., 2023. Model averaging prediction by K-fold cross-validation. J. Econ. 235 (1), 280–301.
Zhang, X., Yu, D., Zou, G., Liang, H., 2016. Optimal model averaging estimation for generalized linear models and generalized linear mixed-effects models. J. Am. Stat. Assoc. 111 (516), 1775–1790.
Zhang, X., Zou, G., Liang, H., Carroll, R.J., 2019. Parsimonious model averaging with a diverging number of parameters. J. Am. Stat. Assoc. 115 (530), 972–984.
Zhao, Y., Goodell, J.W., Dong, Q., Wang, Y., Abedin, M.Z., 2022. Overcoming spatial stratification of fintech inclusion: inferences from across Chinese provinces to guide policy makers. Int. Rev. Financ. Anal. 84, 102411.
Zhou, J., Wang, C., Ren, F., Chen, G., 2021. Inferring multi-stage risk for online consumer credit services: an integrated scheme using data augmentation and model enhancement. Decis. Support Syst. 149, 113611.
Zou, Y., Lin, B., Yang, X., Wu, L., Muneeb Abid, M., Tang, J., 2021. Application of the bayesian model averaging in analyzing freeway traffic incident clearance time for emergency management. J. Adv. Transp. 2021, 1–9.