Computational Economics
https://doi.org/10.1007/s10614-024-10747-6

A Hybrid Credit Risk Evaluation Model Based on Three-Way Decisions and Stacking Ensemble Approach

Yusheng Li[1] · Ran Zhao[1] · Mengyi Sha[2]

Accepted: 8 October 2024

Abstract

Credit risk evaluation is a binary classification problem, and machine learning algorithms have achieved remarkable results in this field. However, traditional two-way decisions involve a high risk of making a bad decision with insufficient information. This paper applies the three-way decision method to introduce the delayed decision mechanism and proposes a hybrid credit risk evaluation model based on the stacking ensemble approach. First, the decision loss values in the three-way decision are determined based on cash flow data. A multiobjective optimization model is constructed to determine the three-way decision thresholds by minimizing the decision cost and the size of the boundary region. Second, in the hybrid model, the LightGBM algorithm is used to evaluate the default probabilities of samples in the first step, and partial samples are made delayed decisions. Multiple ensemble learning algorithms are integrated to form a stacking model to achieve further decision-making on delayed decision samples. Experiments on the credit dataset show that the proposed model performed better than a variety of popular machine learning algorithms and could classify samples with high decision costs better.

Keywords Credit risk evaluation · Three-way decision · Ensemble learning · Stacking · Multiobjective optimization

✉ Yusheng Li
yusheng.li@hebut.edu.cn

1 School of Economics and Management, Hebei University of Technology, No. 8, Guangrong Road, Dingzigu Street, Hongqiao District, Tianjin 300130, People's Republic of China

2 Department of Industrial Engineering, Tsinghua University, Beijing 100084, People's Republic of China

Published online: 23 October 2024

1 Introduction

Credit risk evaluation is a necessary task in the lending process because it helps financial institutions identify risks and avoid enormous losses (Zhang et al., 2022b). Due to the global financial crisis and regulatory concerns of the Basel III Accord, people are paying increasing amounts of attention to credit risk (Baser et al., 2023). Credit evaluation relying solely on manual judgment is no longer suitable in the era of big data (Lu et al., 2023). The application of machine learning methods is attractive, and they yield high classification performance. Therefore, to strengthen the management and prevention of financial risks, it is highly important for financial institutions to develop intelligent credit risk evaluation models.

Many machine learning algorithms have been applied in this field (Jiang et al., 2019), and they play a role in assisting decision-making and helping financial institutions automatically reject high-risk customers. Credit risk evaluation predicts the performance and default probabilities of customers and can be regarded as a two-class problem in pattern recognition (Akila & Srinivasulu Reddy, 2018). A large number of studies have used machine learning methods in credit risk assessment so that the accumulated big data can provide support. A large number of machine learning methods, such as logistic regression (He et al., 2023; Woo & Sohn, 2022), support vector machines (Lee et al., 2022), least absolute shrinkage and selection operator (Lasso) (Li et al., 2024), K-nearest neighbors (Wang et al., 2022b), decision trees (Dumitrescu et al., 2022; Golbayani et al., 2020), neural networks (Xia et al., 2021), deep learning (Pławiak et al., 2020), gradient boosting decision trees (Liu et al., 2022), light gradient boosting machines (LGBM) (Ma et al., 2018; Wang et al., 2022a), hierarchical clustering (Zeitsch, 2019), and K-means (Pang et al., 2021; Song et al., 2020), have been applied in credit risk evaluation.

Because single methods have limited effectiveness, some studies have combined multiple approaches and built hybrid models to achieve better results. Ensemble learning methods can be regarded as hybrid models and have been applied to evaluate credit risk. Yao et al. (2022a) proposed an ensemble model for enterprise credit risk based on the AdaBoost and oversampling technique, and its prediction effect outperformed that of multiple classifiers. Uddin et al. (2020) employed the typical ensemble learning approach random forest to predict credit defaults and proved that it had a better and more stable effect than single classifiers. In addition, researchers have combined different types of models according to their characteristics or the actual needs of different models. Machado and Karray (2022) compared multiple hybrid models with single supervised models and found that hybrid models achieved better performance in predicting clients' credit scores. Sun et al. (2022) proposed two new combined credit evaluation models by integrating asymmetric bagging and the LGBM. Yao et al. (2022b) built a stacking model for credit risk evaluation, and the base learners used five algorithms, including GBDT, XGBoost, LGBM, random forest, and extreme random tree, and the logistic regression (LR) was used to fine-tune the prediction results. Zhang et al. (2023) designed a hybrid credit risk prediction approach according to a convolutional neural network (CNN) and bidirectional long short-term memory (BiLSTM), where the CNN was applied to extract local features and BiLSTM addressed the long term dependence of the data. Compared with single models, the information learned by multiple models can complement each other, and hybrid models can help improve the evaluation accuracy and classification effect. Therefore, constructing a hybrid credit risk evaluation model will improve the ability to identify defaulters.

Machine learning algorithms can learn the data characteristics of default records from a large amount of historical data and can distinguish default records from non-default records (Jiang et al., 2023; Wang et al., 2020b). All records are predicted under the same information conditions and model, and each record is assigned one of two results, yes or no; this is a two-way decision. However, there are always some samples that cannot be predicted well with the available information (Dai et al., 2023; Wang et al. 2023a). In this case, two-way decisions will bring a high risk of making a bad decision with insufficient information.

To decrease the decision risk caused by two-way decisions, three-way decisions can be introduced into credit risk evaluation. The three-way decision method adds a mechanism of delayed decision to the traditional two-way decision, which means that a part of the sample does not need to be given a definitive decision (yes or no) immediately but makes further decisions after providing additional information (Li & Yang, 2023; Maldonado et al., 2020). The supplementary information includes new data or other learners (Li et al., 2016). Three-way decisions do not require all samples to complete the decision at one time, which reduces the risk of misjudging difficult decision-making samples (Li & Sha, 2024; Zhang et al., 2021). However, it is difficult to judge whether a delayed decision should be made on a sample.

In three-way decisions, delayed decision samples are determined by two thresholds α and β, which are used to divide samples into the positive region (POS), negative region (NEG), and boundary region (BND) (Yao, 2010, 2011). A sample given different decisions may result in different losses, and these losses are the basis for the determination of thresholds. However, the loss values are not easy to quantify; for example, it is difficult to determine the loss value when a sample is divided into the BND region. Several studies have discussed the calculation of three-way decision thresholds. The related literature can be roughly classified into three categories. The first type is to randomly set thresholds. Xu et al. (2023) randomly set the loss values in the range of (0, 30] to calculate the thresholds. These methods can only show the effectiveness of three-way decisions in reducing decision errors, and the threshold setting for randomization is independent of the actual problem. The second type is to replace loss values with other parameters. Zhu et al. (2022) employed regret theory and used regret values, rejoicing values, and total psychological perception values to determine thresholds in three-way decisions. Other theories are also employed, including prospect theory (Wang et al., 2020a), grey decision theory (Du et al., 2021), shadowed set theory (Yang & Yao, 2021; Zhang & Yao, 2020), utility theory (Wang et al. 2023c; Zhan et al., 2022), and interval-valued intuitionistic fuzzy decision (Liu et al., 2023). These methods broaden the way of threshold setting, but some parameters still need to be preset. The third type is to automatically learn thresholds. Researchers simplify the number of loss values in three-way decisions and regard threshold learning as an optimization problem. Jia and Shang (Jia et al., 2014) discussed the relationship between thresholds and loss values and learned the three-way decision thresholds with the optimization goal of minimizing the decision cost. Pan et al. (2016) proposed a multiobjective optimization method and added another objective regarding the threshold difference range to limit the size of the boundary region. Shen et al. (2022) employed information gain to construct an optimization objective and restricted the quantity of samples in the boundary region. Li and Shao (2022) discussed the range of thresholds under different loss values and established the threshold learning optimization goal considering the overall decision cost and the uncertainty in the boundary region. These studies investigated the relationship between the thresholds and the loss values and designed optimization goals based on different factors. However, these methods are not specific to credit risk evaluation, and credit information is underutilized.

The motivations for this study are summarized as follows. (1) Credit risk evaluation models based on two-way decisions may lead to a high risk of decision errors. (2) The threshold setting of existing three-way decision methods is not suitable for credit risk evaluation because they do not consider cash flow data. (3) Previous studies have shown that hybrid models outperform single models in evaluating credit risk. These inspire us to employ three-way decisions in credit risk evaluation to reduce decision errors, design a targeted threshold setting method for three-way decisions by considering cash flow data, and propose a hybrid model to improve the classification performance. In brief, this paper makes the following contributions:

(1) Different from traditional three-way decisions in which all samples use a unified decision loss matrix, the cash flow data in credit risk are considered to infer the decision loss values and make the threshold setting more in line with realistic management.
(2) A multiobjective optimization model is designed to automatically learn the thresholds of the three-way decision based on minimizing the decision cost and the size of the boundary region.
(3) A novel hybrid model is proposed in this paper to improve the performance of credit risk evaluation. Based on three-way decisions and stacking ensemble learning, the decision-making process is divided into two steps. In the first step, a portion of the samples are made for the delayed decision by three-way decisions. A stacking model is built by combining multiple classifiers in the second step to provide additional decision information and participate in further decision-making on samples in the boundary region.

The remaining parts of the paper are arranged as follows. The basic knowledge of three-way decisions and stacking ensemble learning is introduced in Sect. 2. Section 3 discusses the setting of the decision loss and threshold, defines the optimization objective, and describes the proposed model framework. The experiments are conducted and discussed in Sect. 4. The conclusion is presented in Sect. 5.

2 Preliminaries

2.1 Three-Way Decisions

The idea of three-way decisions originates from rough sets (Pawlak, 1982; Wang et al. 2023b). The universal U is divided into three regions, namely, the positive region, the negative region, and the boundary region. The three regions are defined as follows:

$$\begin{aligned} POS(X) &= \{x \in U | \Pr(X|[x]) \geq \alpha\} \\ BND(X) &= \{x \in U | \beta < \Pr(X|[x]) < \alpha\} \\ NEG(X) &= \{x \in U | \Pr(X|[x]) \leq \beta\} \end{aligned} \tag{1}$$

where $\Pr(X|[x])$ means the probability that the object x belongs to X. The thresholds α and β are given by experts, but this does not consider the practical needs and styles of decision makers.

To reflect the rationality of decisions, Yao proposed decision-theoretic rough sets (DTRSs) (Yao, 2010). There are states $\Omega = \{X, \neg X\}$ that represent whether an object x belongs to X or not. There are three actions $\{a_P, a_B, a_N\}$ which denotes that an object x is classified into three regions: $POS(X)$, $BND(X)$, and $NEG(X)$. In different states, the losses of taking different actions on an object vary, and the loss matrix is shown in Table 1.

In general, the loss matrix satisfies the conditions $\lambda_{PP} \leq \lambda_{BP} < \lambda_{NP}$ and $\lambda_{NN} \leq \lambda_{BN} < \lambda_{PN}$. The decision rules can be expressed as follows:

(P) If $\Pr(X|[x]) \geq \alpha$, then $x \in POS$,
(B) If $\beta < \Pr(X|[x]) < \alpha$, then $x \in BND$,
(N) If $\Pr(X|[x]) \leq \beta$, then $x \in NEG$.
where the values of α and β are calculated as Eqs. (2) and (3):

$$\alpha = \frac{(\lambda_{PN} - \lambda_{BN})}{(\lambda_{PN} - \lambda_{BN}) + (\lambda_{BP} - \lambda_{PP})}, \tag{2}$$

$$\beta = \frac{(\lambda_{BN} - \lambda_{NN})}{(\lambda_{BN} - \lambda_{NN}) + (\lambda_{NP} - \lambda_{BP})}. \tag{3}$$

Suppose that the boundary region is nonempty, and then $\alpha > \beta$, so the inequality $(\lambda_{BP} - \lambda_{PP}) \cdot (\lambda_{BN} - \lambda_{NN}) < (\lambda_{NP} - \lambda_{BP}) \cdot (\lambda_{PN} - \lambda_{BN})$ holds. By comparing

Table 1 The loss matrix

Action	State	
	X	$\neg X$
a_P	λ_{PP}	λ_{PN}
a_B	λ_{BP}	λ_{BN}
a_N	λ_{NP}	λ_{NN}

the probability and thresholds, an object can be classified into different regions. The calculation of thresholds is determined by six loss values in the loss matrix (Table 1), and they are related to the setting of three-way decision rules.

2.2 Stacking Ensemble Learning

The basic idea of ensemble learning is to combine multiple weak learners and build a strong learner with better performance. The stacking scheme is usually applied to integrate heterogeneous learners (different classification algorithms). It is a parallel multilayer learning structure. In the two-layer learning framework, the base classifiers are in the first layer, and the classifiers in the second layer are called meta-classifiers. An example of a two-layer stacking ensemble learning structure is shown in Fig. 1. Suppose that the original training dataset $D = \{(x_1, y_1), (x_2, y_3), \ldots, (x_n, y_n)\}$ and the number of base classifiers is S. First, the original training dataset is divided into K subsets of the same size $\{D_1, D_2, \ldots, D_K\}$. Second, $K-1$ datasets are used as the training set, and the remaining dataset is used as the testing set. The prediction results of all the base algorithms on the training samples constitute a new dataset D'. Finally, the new dataset is taken as the input of the algorithm in the second layer to train a classifier, which means that the meta-learner is obtained on the basis of the prediction results of the base classifiers, and the number of features is equal to the number of base classifiers.

3 Methodology

In this section, we describe the proposed model in detail. First, we use the cash flow data in credit management to represent the decision loss in the three-way decision. Second, according to the decision cost and size of the boundary region, a multiobjective optimization model is designed to learn three-way decision thresholds. Third, the framework of the proposed model and decision process are elaborated.

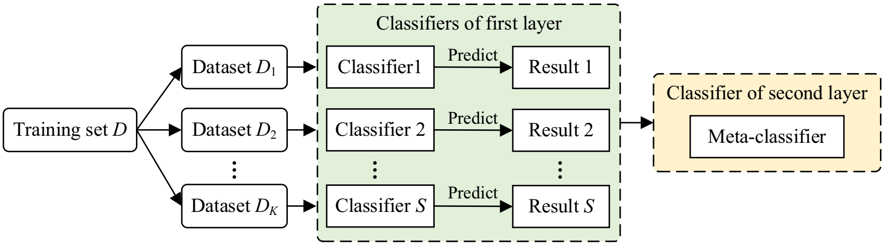

Fig. 1 A two-layer stacking ensemble learning structure

3.1 Thresholds of Three-Way Decisions in Credit Risk Evaluation

There are six loss values in Table 1, which have the following meanings in credit risk evaluation: λ_{PP} and λ_{NN} are the losses of making correct decisions for default and nondefault samples, respectively; λ_{BP} and λ_{BN} are the losses of making delayed decisions for both default and nondefault samples; λ_{NP} represents the loss of making a nondefault decision for a default sample; and λ_{PN} denotes the loss of making a default decision for a nondefault sample. Assuming that the principal and profit of the record r_i are A_i and I_i, respectively, the loss values can be set as follows:

If the right decision is made, then there will be no decision loss, i.e., $\lambda_{NN} = \lambda_{PP} = 0$.
If the default decision is made but it does not actually default, then the decision loss $\lambda_{PN}^{(i)}$ is the potential profit I_i, i.e., $\lambda_{PN}^{(i)} = I_i$. The loss of one sample is $\lambda_{PN}^{(i)}$, and the loss of all samples is expressed as $\boldsymbol{\lambda}_{\mathbf{PN}} = \left\{ \lambda_{PN}^{(1)}, \lambda_{PN}^{(2)}, \ldots, \lambda_{PN}^{(n)} \right\}$, where n represents the number of samples.
If the nondefault decision is made but it actually defaults, then the decision loss is the sum of the principal A_i and the potential profit I_i, i.e., $\lambda_{NP}^{(i)} = A_i + I_i$. The loss of one sample is $\lambda_{NP}^{(i)}$, and the loss of all samples is expressed as $\boldsymbol{\lambda}_{\mathbf{NP}} = \left\{ \lambda_{NP}^{(1)}, \lambda_{NP}^{(2)}, \ldots, \lambda_{NP}^{(i)}, \ldots, \lambda_{NP}^{(n)} \right\}$.
If a delayed decision is made, the decision loss cannot be directly calculated from the cash flow data. Because $\lambda_{PP} \le \lambda_{BP} < \lambda_{NP}$ and $\lambda_{NN} \le \lambda_{BN} < \lambda_{PN}$, we set $\lambda_{BP}^{(i)}$ and $\lambda_{BN}^{(i)}$ as Eq. (4). Therefore, the loss matrix for the record r_i can be simplified to Table 2.

$$\lambda_{BP}^{(i)} = u \cdot \lambda_{NP}^{(i)},\ \lambda_{BN}^{(i)} = v \cdot \lambda_{PN}^{(i)}. \tag{4}$$

The misclassification loss is different for each sample r_i, which is represented as $\lambda_{PN}^{(i)}$ and $\lambda_{NP}^{(i)}$, and the decision loss for samples divided into the boundary region is the same; both are $\lambda_{BP}^{(i)}$ and $\lambda_{BN}^{(i)}$. Therefore, the corresponding thresholds α_i and β_i are not the same. The thresholds of all the samples constitute the sets $\boldsymbol{\alpha}$

Table 2 The simplified loss matrix for the record r_i

Action	State	
	X	$\neg X$
Acceptance a_P	0	$\lambda_{PN}^{(i)}$
Delayed decision a_B	$u \cdot \lambda_{NP}^{(i)}$	$v \cdot \lambda_{PN}^{(i)}$
Rejection a_N	$\lambda_{NP}^{(i)}$	0

and $\boldsymbol{\beta}$, where $\boldsymbol{\alpha} = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$ and $\boldsymbol{\beta} = \{\beta_1, \beta_2, \dots, \beta_n\}$. α_i and β_i are calculated as shown in Eq. (5).

$$\alpha_i = \frac{\left(\lambda_{PN}^{(i)} - \lambda_{BN}^{(i)}\right)}{\left(\lambda_{PN}^{(i)} - \lambda_{BN}^{(i)}\right) + \left(\lambda_{BP}^{(i)} - 0\right)} = \frac{\lambda_{PN}^{(i)} - \lambda_{BN}^{(i)}}{\lambda_{BP}^{(i)} - \lambda_{BN}^{(i)} + \lambda_{PN}^{(i)}} = \frac{\lambda_{PN}^{(i)} - v \cdot \lambda_{PN}^{(i)}}{u \cdot \lambda_{NP}^{(i)} - v \cdot \lambda_{PN}^{(i)} + \lambda_{PN}^{(i)}},$$
$$\beta_i = \frac{\left(\lambda_{BN}^{(i)} - 0\right)}{\left(\lambda_{BN}^{(i)} - 0\right) + \left(\lambda_{NP}^{(i)} - \lambda_{BP}^{(i)}\right)} = \frac{\lambda_{BN}^{(i)}}{\lambda_{BN}^{(i)} + \lambda_{NP}^{(i)} - \lambda_{BP}^{(i)}} = \frac{v \cdot \lambda_{PN}^{(i)}}{v \cdot \lambda_{PN}^{(i)} + \lambda_{NP}^{(i)} - u \cdot \lambda_{NP}^{(i)}}. \tag{5}$$

$\lambda_{PN}^{(i)}$ and $\lambda_{NP}^{(i)}$ of each sample are known exactly, but u and v are inconclusive. In this paper, the thresholds α_i and β_i are calculated by setting the appropriate u and v. Because the thresholds need to meet the condition $\alpha_i \geq \beta_i$, u and v need to meet the following conditions (Eq. 6). If $u = v$, and then $\alpha_i = \beta_i$, the decision degenerates to a two-way decision.

$$\begin{aligned} &\frac{\lambda_{PN}^{(i)} - v \cdot \lambda_{PN}^{(i)}}{u \cdot \lambda_{NP}^{(i)} - v \cdot \lambda_{PN}^{(i)} + \lambda_{PN}^{(i)}} \geq \frac{v \cdot \lambda_{PN}^{(i)}}{v \cdot \lambda_{PN}^{(i)} + \lambda_{NP}^{(i)} - u \cdot \lambda_{NP}^{(i)}} \\ &\Leftrightarrow (1 - v - u) \cdot \lambda_{PN}^{(i)} \cdot \lambda_{NP}^{(i)} \geq 0 \\ &\Leftrightarrow v + u \leq 1 \end{aligned} \tag{6}$$

3.2 Optimization Objective of Threshold Learning

3.2.1 Decision Cost in Credit Risk Evaluation

According to the loss matrix in Table 1, the expected loss $R(a_i|[x])\,(i = P, B, N)$ of a record x for taking different actions is calculated as follows.

$$\begin{aligned} R(a_P|[x]) &= \lambda_{PP} \cdot \Pr(X|[x]) + \lambda_{PN} \cdot \Pr(\neg X|[x]), \\ R(a_B|[x]) &= \lambda_{BP} \cdot \Pr(X|[x]) + \lambda_{BN} \cdot \Pr(\neg X|[x]), \\ R(a_N|[x]) &= \lambda_{NP} \cdot \Pr(X|[x]) + \lambda_{NN} \cdot \Pr(\neg X|[x]). \end{aligned} \tag{7}$$

After the loss matrix is simplified (Table 2), the decision cost for sample r_i to take different actions is calculated as Eq. (8).

$$\begin{aligned} R(a_P|r_i) &= \lambda_{PN}^{(i)} \cdot \left(1 - \Pr\left(X|r_i\right)\right), \\ R(a_B|r_i) &= u \cdot \lambda_{NP}^{(i)} \cdot \Pr\left(X|r_i\right) + v \cdot \lambda_{PN}^{(i)} \cdot \left(1 - \Pr\left(X|r_i\right)\right), \\ R(a_N|r_i) &= \lambda_{NP}^{(i)} \cdot \Pr\left(X|r_i\right). \end{aligned} \tag{8}$$

where $\Pr\left(X|r_i\right)$ denotes the probability of sample r_i being a default record. The decision costs of the POS, BND, and NEG regions are determined by the sum of the decision costs of the samples from the corresponding regions. The region into which a sample r_i is divided is determined by the thresholds (α_i, β_i), and the calculations of α_i and β_i are related to $\lambda_{BP}^{(i)}$ and $\lambda_{BN}^{(i)}$ (see Eqs. 2 and 3). Therefore, the decision cost for different regions is shown in Eq. (9):

$$\begin{aligned}
R_{POS}(u,v) &= \sum_{\Pr(X|r_i)\geq\alpha_i} R(a_P|r_i) = \sum_{r_i\in POS} \lambda_{PN}^{(i)} \cdot (1-\Pr(X|r_i)),\\
R_{BND}(u,v) &= \sum_{\beta_i<\Pr(X|r_i)<\alpha_i} R(a_B|r_i) = \sum_{r_i\in BND} (\lambda_{BP}\cdot\Pr(X|r_i)+\lambda_{BN}\cdot(1-\Pr(X|r_i))),\\
R_{NEG}(u,v) &= \sum_{\Pr(X|r_i)\leq\beta_i} R(a_N|r_i) = \sum_{r_i\in NEG} \lambda_{NP}^{(i)}\cdot\Pr(X|r_i).
\end{aligned} \tag{9}$$

The overall three-way decision cost is composed of the decision costs of the POS, BND, and NEG regions, as shown in Eq. (10).

$$R_{3WD}(u,v) = R_{POS}(u,v) + R_{BND}(u,v) + R_{NEG}(u,v). \tag{10}$$

3.2.2 Multiobjective Optimization Function

When making decisions, it is hoped that the decision cost is as low as possible. On the other hand, the size of the boundary region should not be too large. Therefore, this paper constructs a multiobjective optimization model by minimizing both the decision cost and the size of the boundary region, as expressed in Eq. (11):

$$\min\ F(u,v) = (f1(u,v), f2(u,v)), \tag{11}$$

$$f1(u,v) = R_{3WD}(u,v), \tag{12}$$

$$f2(u,v) = \sum_{i=1}^{N} (\alpha_i - \beta_i). \tag{13}$$

where the difference between $\alpha^{(i)}$ and $\beta^{(i)}$ is used to characterize the size of the boundary region. The objective in Eq. (12) is to minimize the overall decision cost of the three-way decision. The objective in Eq. (13) is to minimize the boundary region size of the samples. These objectives are minimized subject to the following constraints:

$$u < 1, v < 1. \tag{14}$$

$$u + v \leq 1 \tag{15}$$

where $\min(\boldsymbol{\lambda}_{\mathbf{PN}})$ and $\min(\boldsymbol{\lambda}_{\mathbf{NP}})$ are the minimum values of $\boldsymbol{\lambda}_{\mathbf{PN}}$ and $\boldsymbol{\lambda}_{\mathbf{NP}}$ among all the samples, respectively. The constraint in Eq. (14) ensures that the loss values $\lambda_{BN}^{(i)}$

and $\lambda_{BP}^{(i)}$ satisfy the conditions of $\lambda_{PP}^{(i)} \leq \lambda_{BP}^{(i)} < \lambda_{NP}^{(i)}$ and $\lambda_{NN}^{(i)} \leq \lambda_{BN}^{(i)} < \lambda_{PN}^{(i)}$. The constraint in Eq. (15) is from Eq. (6) and ensures that $\alpha^{(i)} \geq \beta^{(i)}$.

Theorem 1 *$\alpha^{(i)} - \beta^{(i)}$ monotonically decreases with u and v.*

Proof In Eq. (5), we know that $\alpha^{(i)} = \left(\lambda_{PN}^{(i)} - v \cdot \lambda_{PN}^{(i)}\right) \Big/ \left(u \cdot \lambda_{NP}^{(i)} - v \cdot \lambda_{PN}^{(i)} + \lambda_{PN}^{(i)}\right)$; if u is larger, the denominator of $\alpha^{(i)}$ is larger, and $\alpha^{(i)}$ is smaller. Because $\beta^{(i)} = v \cdot \lambda_{PN}^{(i)} \Big/ \left(v \cdot \lambda_{PN}^{(i)} + \lambda_{NP}^{(i)} - u \cdot \lambda_{NP}^{(i)}\right)$, a larger u decreases the denominator of $\beta^{(i)}$, and $\beta^{(i)}$ increases. Similarly, a larger v decreases $\alpha^{(i)}$ and increases $\beta^{(i)}$. Because large u and v produce smaller $\alpha^{(i)}$ and larger $\beta^{(i)}$, $\alpha^{(i)} - \beta^{(i)}$ decreases. Therefore, Theorem 1 holds true.

Theorem 1 illustrates that when u and v are large, $\alpha^{(i)} - \beta^{(i)}$ will be small. If the loss of putting a record into the boundary region increases, fewer records would be delayed decisions, which is reasonable in reality. Based on Theorem 1, the objective in Eq. (13) monotonically decreases with u and v.

The loss values u and v are consecutive, so it is not possible to exhaustively obtain the optimal solution. For multiobjective optimization problems, the nondominated sorting genetic algorithm II (NSGA-II) is a popular optimization algorithm. The NSGA-II was proposed by Deb based on the NSGA (Deb et al., 2002). Its fast convergence advantage makes it a benchmark for the performance of other multiobjective optimization algorithms (Hua et al., 2021). This paper applies the NSGA-II to determine the thresholds of three-way decisions. Among the generated multiple solutions, the solution with the least number of samples in the boundary region is selected.

3.3 Framework of the Proposed Model

Researchers have proven that the LightGBM (LGBM) algorithm outperforms many traditional machine learning methods in terms of running speed and classification effect (Ke et al., 2017). It also has excellent performance in the fields of economy and finance (Wang et al., 2022a). Therefore, in the first step of the hybrid model, the LGBM is applied to assess the default probabilities of samples, and samples are divided into different regions by three-way decisions. In the second step, the LGBM and other ensemble models are combined to form a new model, which realizes further decision-making on samples in the boundary region.

The framework of the proposed hybrid model is presented in Fig. 2. In Step 1, the samples in the test set are classified into the POS, NEG, and BND regions. Initially, the LGBM algorithm learns a classification model on the historical dataset and evaluates the default probabilities of samples in the test set. According to the optimization objective function constructed in Eq. (11), the three-way decision thresholds of each sample are obtained. Different decisions are made on samples in the test set, and they are divided into POS, NEG, and BND regions based on their threshold pairs. In Step 2, in addition to LGBM, five popular ensemble learning

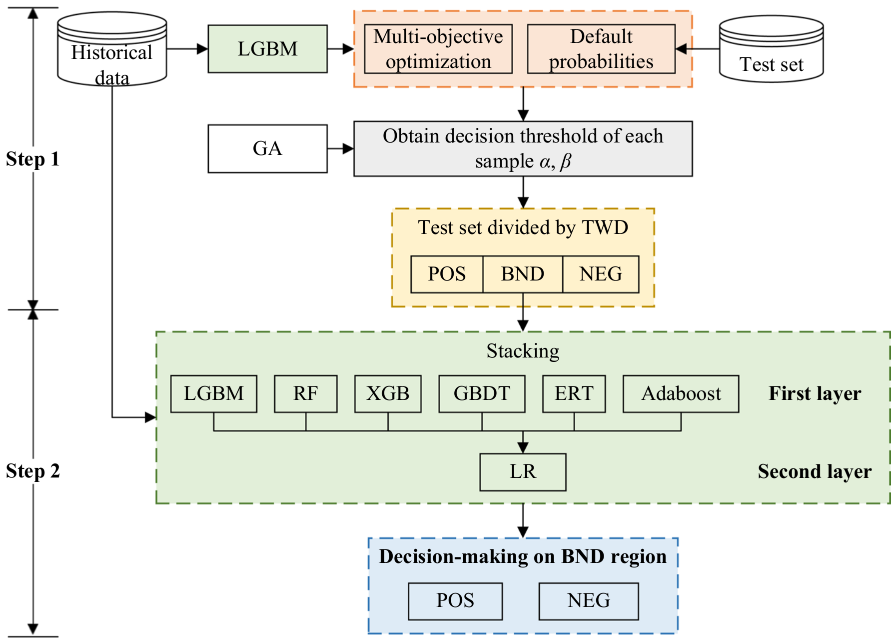

Fig. 2 Framework of the proposed credit risk prediction model

methods, namely, random forest (RF), XGBoost (XGB), gradient boosted decision tree (GBDT), extreme random tree (ERT), and AdaBoost, are utilized as the first-level classifiers of the stacking models. Logistic regression is often applied as a meta-classifier for stacking models, so we used it in the hybrid model. The proposed stacking model combines multiple types of classifiers. This approach provides additional information for decision-making and classifies samples in the BND region of Step 1.

4 Experiments

4.1 Data Description and Experimental Setting

To verify the effectiveness of the proposed model, this paper conducts experiments on a personal loan dataset from a provincial local bank in China. The dataset consists of 6202 records, including 1081 default records and 5121 nondefault records, and the features and related descriptions are shown in Table 3. The values of binary attributes in the dataset are denoted by 0 or 1, and the values of nominal attributes are represented by different integer values. The missing values are filled with the average of existing data. We use recursive feature elimination with cross-validation to select features, and LGBM provides the feature importance in the process of feature selection. The features that achieved the best classification performance were preserved, the features “A2”, “A14”, and “A16” were removed and the number of

Table 3 Features and descriptions of the dataset

Feature	Description
A1	Loan amount
A2	Year of loan
A3	Current loan rates
A4	Installment amount
A5	Company types: higher education institutions, general enterprises, listed companies, fortune 500 companies, early childhood education, primary and secondary schools, government organizations, and other 6 categories
A6	Industry types: traditional industry, commerce, internet, finance, and other 14 categories
A7	Years of work
A8	Availability of housing: yes or no
A9	Type of loan purpose: 14 categories
A10	Postal code of the borrower at the time of application
A11	Region code
A12	Debt-to-income ratio
A13	Accumulated prepayment amount
A14	Number of default events where the borrower was more than 30 days past due in the past 18 months
A15	Number of outstanding credit limits in the borrower's file
A16	Whether individual application: yes or no

selected features is thirteen. The oversampling method synthetic minority over-sampling technique (SMOTE) (Chawla et al., 2002) is applied to balance the dataset.

In this paper, classical algorithms such as the LR, k-nearest neighbors (KNN), Bayesian classification (Bayes), neural network (NN), support vector machine (SVM), LGBM, RF, XGB, GBDT, ERT, and AdaBoost algorithms are also used to conduct comparative experiments. The fivefold cross-validation is applied in the experimental process, where the dataset is divided into five equal parts, four of which are used as the training set and the remaining one as the test set. In the experiments, 10 rounds of fivefold cross-validation are performed, and the average values are taken as the final results. The program is written in Python 3.8, the algorithms are implemented using the scikit-learn, lightgbm, and xgboost libraries, and the parameters used in the algorithms are determined by grid searching, as shown in Table 4. The multiobjective genetic algorithm of the NSGA-II is implemented using geatpy library, the population size is 100, and the maximum number of iterations is 200.

4.2 Evaluation Measures

The evaluation metrics are calculated based on the confusion matrix in Table 5. True Positive (TP) means the number of correctly classified default records. True Negative (TN) means the number of correctly classified nondefault records. False

Table 4 Parameters of the machine learning algorithms

Method	Parameters	Description	Search space	Selected
KNN	n_neighbors	Number of neighbors	3, 5, 7, 9, 11, 13, 15,17, 19	15
	Metric	Metric for distance computation	euclidean, minkowski, mahalanobis	euclidean
LR	Penalty	Choice of penalty	l1,l2,elasticnet	l1
	C	Regularization coefficient	1, 2, 3, 4, 5, 6, 7, 8, 9, 10	1
	Solver	Choice of solver	lbfgs, liblinear, newton-cg, sag, saga, newton-cholesky	liblinear
NN	hidden_layer_sizes	Number of neurons	100, 150, 200, 250, 300	100
	Activation	Activation function	identity, logistic, tanh, relu	logistic
	Solver	Choice of solver	lbfgs, sgd, adam	lbfgs
SVM	C	Regularization coefficient	1, 2, 3, 4, 5, 6, 7, 8, 9, 10	5
	Kernel	Kernel function	linear, poly, rbf, sigmoid	rbf
Adaboost	n_estimators	Number of estimators	100, 150, 200, 250, 300	200
	learning rate	Learning rate	0.05, 0.1, 0.15, 0.2	0.05
ERT	n_estimators	Number of estimators	100, 150, 200, 250, 300	250
	max_depth	Maximum depth of tree	2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, None	12
	max_features	Maximum number of features	sqrt, log2, None	None
LGBM	n_estimators	Number of estimators	100, 150, 200, 250, 300	100
	Learning rate	Learning rate	0.05, 0.1, 0.15, 0.2	0.05
XGB	n_estimators	Number of estimators	100, 150, 200, 250, 300	300
	Learning rate	Learning rate	0.05, 0.1, 0.15, 0.2	0.2
RF	n_estimators	Number of estimators	100, 150, 200, 250, 300	100
	max_features	Maximum number of features	sqrt, log2, None	sqrt
	max_depth	Maximum depth of tree	2, 3, 4, …,15	17
GBDT	n_estimators	Number of estimators	100, 150, 200, 250, 300	100
	Learning rate	Learning rate	0.05, 0.1, 0.15, 0.2	0.05

Table 5 The confusion matrix

Actual label	Predict label	
	Default	Nondefault
Default	True Positive (TP)	False Negative (FN)
Nondefault	False Positive (FP)	True Negative (TN)

Negative (FN) means the number of incorrectly classified default records. False Positive (FP) means the number of incorrectly classified nondefault records.

The following indicators are used in this paper. The quantity of default records is not equal to that of nondefault records, so the dataset is imbalanced. The balanced accuracy is the average of the accuracy over each class (Mahbobi et al., 2021). This paper uses balanced accuracy (B_Acc) to assess the effect of identifying loan default records, and it is calculated as Eq. (16).

$$B\_Acc = \left(\frac{TP}{TP + FP} + \frac{TN}{TN + FN} \right) \Big/ 2. \tag{16}$$

where $TP/(TP+FP)$ denotes the proportion of default records that are correctly classified, $TN/(TN+FN)$ denotes the proportion of nondefault records that are correctly classified, and their average is the balanced accuracy. In addition, the F-measure (FM) and G-mean (GM) are also applied, and they more comprehensively reflect the performance of the algorithms. The computations of FM and GM can be summarized as Eqs. (17) and (18), respectively, and the parameter $b = 1$ is usually used for FM.

$$FM = \frac{\left(1 + b^2\right) \times \frac{TP}{TP+FP} \times \frac{TP}{TP+FN}}{b^2 \times \left(\frac{TP}{TP+FP} + \frac{TP}{TP+FN} \right)}, \tag{17}$$

$$GM = \sqrt{\frac{TP}{TP + FN} \times \frac{TN}{TN + FP}}. \tag{18}$$

The area under the receiver operating characteristic curve (AUC) measures the prediction performance of classifiers (Zhang et al. 2022a). The classical calculation of the AUC is shown in Eq. (19). The samples are ranked in descending order of probability, $rank_i$ is the rank position value of sample r_i, and P and N are the numbers of positive samples and negative samples, respectively.

$$AUC = \frac{\sum_{i \in positiveClass} rank_i - \frac{P \times (1+P)}{2}}{N \times P}, \tag{19}$$

The decision cost (Cost) is also an evaluation indicator. The label y_i of the record r_i in test set D' is known, and the predicted label is denoted as $\hat{y}_i$. For a record of default ($y_i = 1$), the decision error cost is $\lambda_{NP}^{(i)}$; for a record of nondefault ($y_i = 0$), the decision error cost is $\lambda_{PN}^{(i)}$. The overall decision cost of classification can be calculated as Eq. (20).

$$Cost = \sum_{r_i \in D'} \left(\lambda_{NP}^{(i)} \cdot y_i \cdot \left| y_i - \hat{y}_i \right| + \lambda_{PN}^{(i)} \cdot \left(1 - y_i\right) \cdot \left| y_i - \hat{y}_i \right| \right) \tag{20}$$

4.3 Performance Comparison Among Different Credit Risk Evaluation Methods

This paper analyses the thresholds learned by three-way decisions. Taking LGBM as the classification algorithm, changes in the decision cost and number of samples in different regions are investigated in Fig. 3. Because $u + v \leq 1$, the results can only be obtained in the area below the diagonal. In Fig. 3a, c, as u and v increase, the

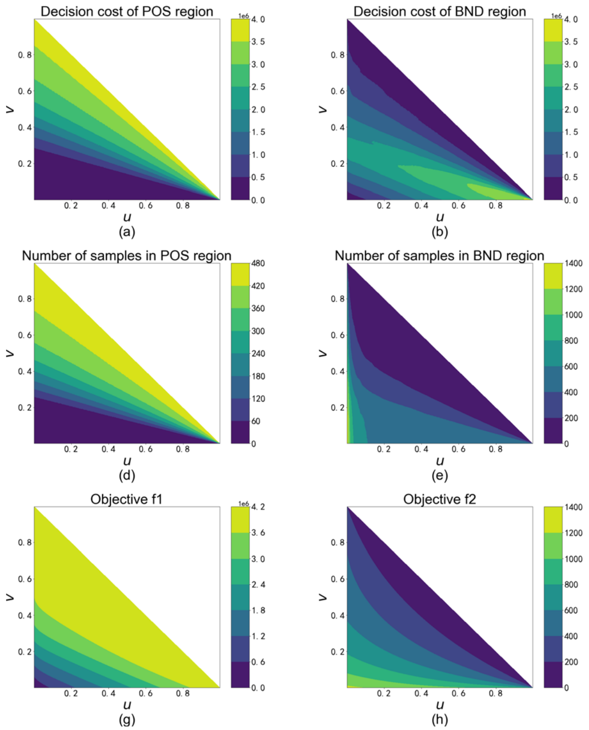

Fig. 3 Changes in results for different λ_{BP} and λ_{BN}

decision costs of the positive and negative regions increase. This is because the numbers of samples in the positive and negative regions monotonically increase with u and v, as demonstrated in Fig. 3d, f. In addition, based on Theorem 1, the number of samples from the boundary region decreases with the increase of u and v (Fig. 3e). Figure 3b shows that there is no monotonic relationship between the decision cost of samples in the boundary region and u, v. This is because the increase in u and v will lead to a decrease in the number of samples in the boundary region. Figure 3g, h show that objective $f1$ increases and objective $f2$ decreases, respectively, with increasing u and v. Overall, the experimental results are consistent with Theorem 1, and objective $f1$ and objective $f2$ have opposite changes with respect to u and v.

The learned thresholds of the samples and their default probabilities are shown in Fig. 4. The decision threshold is not fixed at 0.5, as in traditional two-way decisions, but α and β vary depending on the sample. The threshold α is not very large overall, and the threshold β of the sample is relatively small. This indicates that a nondefault decision will be made when the default probability is small, which is beneficial for finding default records. Moreover, the default probabilities of samples from the POS and NEG regions are concentrated at 0.0 or 1.0, respectively (Fig. 4c). Only when the default probability of a sample is very high or very low will a certain decision (default or nondefault) be made on the sample. In contrast, the distributions of the default probabilities of samples from the BND region (Fig. 4f) are different,

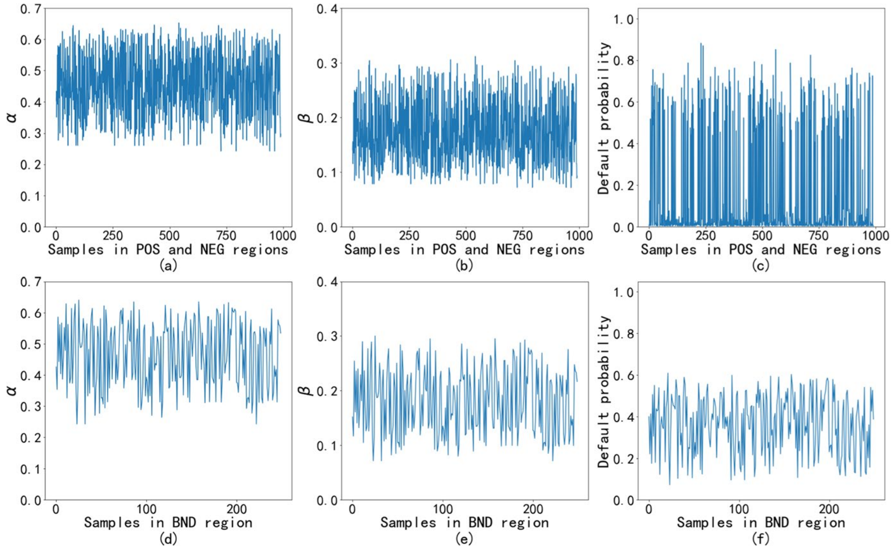

Fig. 4 Three-way decision thresholds and default probability of the sample

and the probability values are more concentrated in the middle rather than at 0.0 or 1.0. Therefore, a three-way decision can make certain decisions (default or nondefault) on samples that are easier to classify; for samples that are difficult to make decisions under available information, delayed decisions will be made on them, and these samples are placed in the boundary region.

Many popular machine learning algorithms are applied and compared with our proposed model, and the results are shown in Table 6. Among the single learners,

Table 6 Classification performance of different algorithms

Method	B_Acc	AUC	FM	GM	Cost	TP	TN	FP	FN
Bayes	0.6599	0.8298	0.4316	0.6079	6.51E+06	92.76	912.96	111.24	123.44
KNN	0.6483	0.8187	0.4142	0.6141	6.60E+06	95.44	875.5	148.7	120.76
LR	0.6615	0.8610	0.4565	0.6023	6.06E+06	84.12	**956.5**	**67.7**	132.08
NN	0.6570	0.8386	0.3367	0.4827	6.44E+06	94.24	894.5	129.7	121.96
SVM	0.6550	0.8159	0.4018	0.6105	6.30E+06	101.82	854.36	169.84	114.38
AdaBoost	0.7217	0.8810	0.5351	0.6997	5.43E+06	118.44	916.78	107.42	97.76
ERT	0.7032	0.8829	0.5184	0.6700	5.55E+06	106.08	937.56	86.64	110.12
GBDT	0.7193	0.8793	0.5274	0.6992	5.48E+06	119.42	907.48	116.72	96.78
LGBM	0.7269	0.8800	0.5373	0.7092	5.36E+06	123.06	906	118.2	93.14
RF	0.6949	0.8776	0.5011	0.6616	5.75E+06	104.56	928.24	95.96	111.64
XGB	0.7131	0.8769	0.5192	0.6912	5.60E+06	116.58	908.46	115.74	99.62
Stacking	0.7115	0.8810	0.5246	0.6846	5.50E+06	112.16	925.28	98.92	104.04
Proposed	**0.7530**	**0.8844**	**0.5625**	**0.7432**	**5.08E+06**	**137.92**	888.82	135.38	**78.28**

The values of TP, TN, FP, and FN are the average results of 10 times fivefold cross-validation

Bold values indicate the best results

LR has the best effect in terms of B_Acc, AUC, FM, and Cost. The KNN algorithm achieves the best performance in the GM indicator. Although NN is weaker than Bayes and KNN in terms of FM and GM, its decision cost is lower than theirs at 6.44E+06. In addition, single learners are weaker than ensemble methods in most cases. The AUC of any ensemble learner is greater than that of any single learner. Specifically, LR has the highest AUC (0.8610) among the single learners, and the AUC (0.8769) of XGB is the smallest compared with that of the other ensemble methods. For the indicators of Cost and TP, all the ensemble methods outperform the single methods. Therefore, the classification effect of the ensemble methods is better than that of single methods.

Compared with the single and ensemble algorithms, the proposed model can further improve the classification performance (Table 6). For the indicators of B_Acc, AUC, FM, GM, and Cost, the proposed model achieves the best performance. Significant improvements are achieved in the B_Acc, FM, and GM indicators, and these indicators increase by at least 3%. In the proposed model, the LGBM is used in the first step, and in the second step, a stacking model composed of six ensemble models is applied. A comparison of the proposed model and LGBM reveals that the proposed model yields better results for most indicators. The stacking model performs mediocrely on various indicators, neither the best nor the worst. For TP, TN, FP, and FN, the proposed model has the best values in terms of TP and FN but does not perform well in TN and FP, and is only better than KNN and SVM for these two indicators. This means that the proposed method has the best performance on most evaluation indicators, especially in identifying default samples; however, more non-default samples are mistakenly classified as default samples.

4.4 Performance Comparison on High Decision Cost Samples

We also analyse the distribution of decision costs of different regions after using three-way decisions, as shown in Fig. 5. The average decision cost value of samples from the positive and negative regions is 15.776 k, and the average value of samples in the boundary region is 24.098 k. The boxes of samples from the positive and negative regions are more concentrated at the bottom, while the box of samples in the boundary region is higher. This is because the threshold setting of the three-way decision is related to the decision cost, and samples with high decision costs tend to be delayed decisions.

From Fig. 5, we know that samples with high decision costs are more likely to be divided into the boundary region. In addition, the “high decision cost” samples are more important because they would bring greater losses to financial institutions if incorrect decisions are made about them. This paper analyses the classification results of “high decision cost” samples. Table 7 shows the classification accuracy rates of samples whose decision cost is in the top 1%, 5%, 10%, and 20%. Whether for all samples or for the default samples, the proposed model has a higher classification accuracy for the “higher decision cost” samples. For the classification accuracy of the “high decision cost” sample in all samples, as the sample ratio increases, the superiority of the proposed model gradually decreases. Specifically, the largest

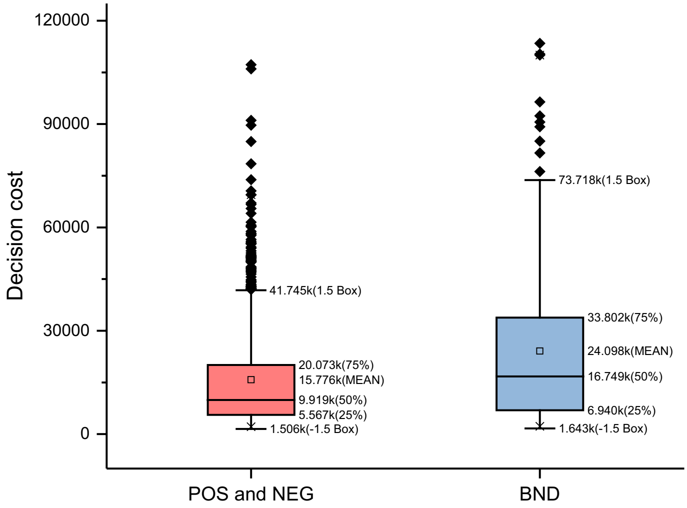

Fig. 5 Distribution of decision costs for samples in different regions

improvements in the top 1%, 5%, 10%, and 20% of the “high decision cost” samples are approximately 7%, 4%, 3%, and 2%, respectively. Except for the accuracy with decision costs in the top 5% among default samples, which is weaker than that of LGBM and ERT, the proposed model outperforms other ensemble methods in the classification of “higher decision cost” default samples. Our model increases the classification accuracy of samples with “high decision cost”, so the final decision cost is reduced, which can be used to explain the phenomenon in which the proposed model performs better on the Cost indicator.

4.5 Discussion and Implications

Based on the experimental results, some implications are discussed in this section.

First, after using three-way decision methods, samples with middle default probabilities are more easily divided into the boundary region, and delayed decisions are made. The cash flow data on credit risk should be taken into consideration in the threshold learning of three-way decisions, which can be more in line with the actual management of credit risk.

Second, because minimizing the overall decision cost is one of the optimization objectives, the average decision cost of samples in the boundary region is greater than that of samples in the positive and negative regions. The proposed hybrid model outperforms existing popular single machine learning classifiers and ensemble learning classifiers. It can also better classify samples with high decision costs. These advantages make the model more reliable and applicable to practical problems.

Third, accurately assessing customer risks and reducing financial losses are concerns of financial institutions, and our proposed model can meet these requirements. Decision makers can adopt the three-way decision method and the hybrid

Table 7 Classification results for samples with high decision costs

Method	Accuracy with decision costs in the top 1% among all samples	Accuracy with decision costs in the top 5% among all samples	Accuracy with decision costs in the top 10% among all samples	Accuracy with decision costs in the top 20% among all samples	Accuracy with decision costs in the top 1% among default samples	Accuracy with decision costs in the top 5% among default samples	Accuracy with decision costs in the top 10% among default samples	Accuracy with decision costs in the top 20% among default samples
AdaBoost	0.8250	0.7255	0.7008	0.6865	0.7400	0.8029	0.7827	0.7264
ERT	0.8100	0.7210	0.6852	0.6677	0.8300	0.8164	0.7774	0.7236
GBDT	0.8033	0.7310	0.7013	0.6934	0.8200	0.8032	0.7723	0.7289
LGBM	0.8383	0.7461	0.7121	0.7008	0.8400	**0.8370**	0.8097	0.7548
RF	0.6817	0.6542	0.6631	0.6735	0.6800	0.6591	0.6565	0.6034
XGB	0.7633	0.7019	0.6885	0.6871	0.7500	0.7557	0.7206	0.6779
Stacking	0.8033	0.7210	0.6908	0.6807	0.8600	0.8053	0.7754	0.7229
Proposed	**0.9083**	**0.7868**	**0.7429**	**0.7202**	**0.8800**	0.8077	**0.8129**	**0.8059**

Bold values indicate the best results

model in the credit evaluation system, which helps financial institutions prevent enormous losses. This study can be used by banks and professional credit institutions as a powerful tool for assessing risk when they offer clients financial services including personal loans, corporate loans, home mortgage loans, and credit cards.

5 Conclusion

Many machine learning classification algorithms have been successfully applied in credit risk evaluation, but they have a high risk of misjudgement under two-way decisions. To reduce decision errors, this paper introduces the three-way decision approach and proposes a hybrid credit risk evaluation model. Unlike traditional three-way decisions in which all the samples share the same threshold, this paper uses cash flow information to simplify the loss value under different decision actions, and each sample has unique three-way decision thresholds. This paper transforms threshold setting into an optimization problem and constructs a multiobjective optimization model by minimizing decision costs and the size of the boundary region. The NSGA-II algorithm is used to determine the thresholds for each sample. Three-way decisions play a role in the first step of the proposed model; the LGBM algorithm evaluates the default probability of the samples, and some samples are delayed decisions. A stacking model composed of multiple algorithms is formed to make further decisions on samples in the boundary region.

This paper uses a variety of machine learning algorithms for comparison experiments, and the following conclusions can be drawn: (1) The learned three-way decision threshold is prone to making certain decisions (yes or no) on easily distinguishable samples and making delayed decisions on difficult-to-distinguish samples, which can achieve the purpose of reducing decision errors. (2) The classification effect of ensemble learning methods is generally better than that of a single method, and the proposed model has the best performance compared with various popular machine learning methods and outperforms the simple stacking model. Moreover, the cost due to decision errors is also reduced after using the proposed model. (3) In the division of the three-way decision, samples with a high decision cost are more likely to fall into the boundary region, and the proposed model can better classify these samples correctly, improving the classification effect and reducing the decision cost.

Our study has several limitations. Although the identification ability of default samples is improved, the number of misjudged default samples is also increased. In the future, more attention will be given to improving the default sample prediction ability. The basic idea is to control the size of the boundary region, as well as the upper and lower bounds of thresholds α and β, to reduce the misclassification of nondefault samples into default samples. We will also explore different ways to combine models and transform two-stage decisions into multistage decisions, and different models will be used in each stage. In addition, credit risk evaluation is a data imbalance problem, and the existing oversampling method is applied in this paper. In subsequent research, we will concentrate on the relevant theories and techniques of imbalanced data mining and develop sampling methods suitable for credit risk.

Author contributions Yusheng Li: Methodology, Software, Investigation, Resources, Writing—Original Draft, Visualization, Supervision, Project administration, Funding acquisition. Ran Zhao: Validation, Formal analysis, Investigation, Data Curation, Writing—Original Draft. Mengyi Sha: Conceptualization, Methodology, Data Curation, Writing—Review & Editing.

Funding This work was supported by the Science Research Project of Hebei Education Department (No. QN2023121), the National Natural Science Foundation of China (No. 72301093), and the Natural Science Foundation of Hebei Province (No. G2023202004).

Data Availability The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declarations

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

References

Akila, S., & Srinivasulu Reddy, U. (2018). Cost-sensitive Risk Induced Bayesian Inference Bagging (RIBIB) for credit card fraud detection. *Journal of Computational Science, 27*, 247–254. https://doi.org/10.1016/j.jocs.2018.06.009

Baser, F., Koc, O., & Selcuk-Kestel, A. S. (2023). Credit risk evaluation using clustering based fuzzy classification method. *Expert Systems with Applications*. https://doi.org/10.1016/j.eswa.2023.119882

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research, 16*, 321–357.

Dai, J., Chen, T., & Zhang, K. (2023). The intuitionistic fuzzy concept-oriented three-way decision model. *Information Sciences, 619*, 52–83. https://doi.org/10.1016/j.ins.2022.11.017

Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation, 6*(2), 182–197. https://doi.org/10.1109/4235.996017

Du, J., Liu, S., & Liu, Y. (2021). A novel grey multi-criteria three-way decisions model and its application. *Computers & Industrial Engineering, 158*, 1–15. https://doi.org/10.1016/j.cie.2021.107405

Dumitrescu, E., Hué, S., Hurlin, C., & Tokpavi, S. (2022). Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects. *European Journal of Operational Research, 297*(3), 1178–1192. https://doi.org/10.1016/j.ejor.2021.06.053

Golbayani, P., Florescu, I., & Chatterjee, R. (2020). A comparative study of forecasting corporate credit ratings using neural networks, support vector machines, and decision trees. *The North American Journal of Economics and Finance, 54*, 1–16. https://doi.org/10.1016/j.najef.2020.101251

He, H., Wang, Z., Jain, H., Jiang, C., & Yang, S. (2023). A privacy-preserving decentralized credit scoring method based on multi-party information. *Decision Support Systems*. https://doi.org/10.1016/j.dss.2022.113910

Hua, Y., Zhu, H., Gao, M., & Ji, Z. (2021). Multiobjective optimization design of permanent magnet assisted bearingless synchronous reluctance motor using NSGA-II. *IEEE Transactions on Industrial Electronics, 68*(11), 10477–10487. https://doi.org/10.1109/tie.2020.3037873

Jia, X., Tang, Z., Liao, W., & Shang, L. (2014). On an optimization representation of decision-theoretic rough set model. *International Journal of Approximate Reasoning, 55*(1), 156–166. https://doi.org/10.1016/j.ijar.2013.02.010

Jiang, C., Lu, W., Wang, Z., & Ding, Y. (2023). Benchmarking state-of-the-art imbalanced data learning approaches for credit scoring. *Expert Systems with Applications*. https://doi.org/10.1016/j.eswa.2022.118878

Jiang, C., Wang, Z., & Zhao, H. (2019). A prediction-driven mixture cure model and its application in credit scoring. *European Journal of Operational Research, 277*(1), 20–31. https://doi.org/10.1016/j.ejor.2019.01.072

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., et al. (2017). *LightGBM: a highly efficient gradient boosting decision tree* Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, California, USA.

Lee, I. G., Yoon, S. W., & Won, D. (2022). A mixed integer linear programming support vector machine for cost-effective group feature selection: Branch-cut-and-price approach. *European Journal of Operational Research, 299*(3), 1055–1068. https://doi.org/10.1016/j.ejor.2021.12.030

Li, W., Huang, Z., & Li, Q. (2016). Three-way decisions based software defect prediction. *Knowledge-Based Systems, 91*, 263–274. https://doi.org/10.1016/j.knosys.2015.09.035

Li, W., & Yang, B. (2023). Three-way decisions with fuzzy probabilistic covering-based rough sets and their applications in credit evaluation. *Applied Soft Computing, 136*, 1–16. https://doi.org/10.1016/j.asoc.2023.110144

Li, X., Zhang, Z., Li, L., & Pan, H. (2024). Combining feature selection and classification using LASSO-based MCO classifier for credit risk evaluation. *Computational Economics*. https://doi.org/10.1007/s10614-023-10535-8

Li, Y., & Sha, M. (2024). Two-stage credit risk prediction framework based on three-way decisions with automatic threshold learning. *Journal of Forecasting, 43*(5), 1263–1277. https://doi.org/10.1002/for.3074

Li, Y., & Shao, X. (2022). Thresholds learning of three-way decisions in pairwise crime linkage. *Applied Soft Computing, 120*, 1–16. https://doi.org/10.1016/j.asoc.2022.108638

Liu, J., Huang, B., Li, H., Bu, X., & Zhou, X. (2023). Optimization-based three-way decisions with interval-valued intuitionistic fuzzy information. *IEEE Transactions on Cybernetics, 53*(6), 3829–3843. https://doi.org/10.1109/tcyb.2022.3151899

Liu, W., Fan, H., & Xia, M. (2022). Credit scoring based on tree-enhanced gradient boosting decision trees. *Expert Systems with Applications*. https://doi.org/10.1016/j.eswa.2021.116034

Lu, J., Wu, D., Dong, J., & Dolgui, A. (2023). A decision support method for credit risk based on the dynamic Bayesian network. *Industrial Management & Data Systems, 123*(12), 3053–3079. https://doi.org/10.1108/imds-04-2023-0250

Ma, X., Sha, J., Wang, D., Yu, Y., Yang, Q., & Niu, X. (2018). Study on a prediction of P2P network loan default based on the machine learning LightGBM and XGboost algorithms according to different high dimensional data cleaning. *Electronic Commerce Research and Applications, 31*, 24–39. https://doi.org/10.1016/j.elerap.2018.08.002

Machado, M. R., & Karray, S. (2022). Assessing credit risk of commercial customers using hybrid machine learning algorithms. *Expert Systems with Applications, 200*, 1–12. https://doi.org/10.1016/j.eswa.2022.116889

Mahbobi, M., Kimiagari, S., & Vasudevan, M. (2021). Credit risk classification: An integrated predictive accuracy algorithm using artificial and deep neural networks. *Annals of Operations Research, 330*(1–2), 609–637. https://doi.org/10.1007/s10479-021-04114-z

Maldonado, S., Peters, G., & Weber, R. (2020). Credit scoring using three-way decisions with probabilistic rough sets. *Information Sciences, 507*, 700–714. https://doi.org/10.1016/j.ins.2018.08.001

Pan, R., Zhang, Z., Fan, Y., Cao, J., Lu, K., & Yang, T. (2016). Multi-objective optimization method for learning thresholds in a decision-theoretic rough set model. *International Journal of Approximate Reasoning, 71*, 34–49. https://doi.org/10.1016/j.ijar.2016.01.002

Pang, P. S., Hou, X., & Xia, L. (2021). Borrowers' credit quality scoring model and applications, with default discriminant analysis based on the extreme learning machine. *Technological Forecasting and Social Change*. https://doi.org/10.1016/j.techfore.2020.120462

Pawlak, Z. (1982). Rough sets. *International Journal of Computer & Information Sciences, 11*(5), 341–356.

Pławiak, P., Abdar, M., Pławiak, J., Makarenkov, V., & Acharya, U. R. (2020). DGHNL: A new deep genetic hierarchical network of learners for prediction of credit scoring. *Information Sciences, 516*, 401–418. https://doi.org/10.1016/j.ins.2019.12.045

Shen, F., Zhang, X., Wang, R., Lan, D., & Zhou, W. (2022). Sequential optimization three-way decision model with information gain for credit default risk evaluation. *International Journal of Forecasting, 38*(3), 1116–1128. https://doi.org/10.1016/j.ijforecast.2021.12.011

Song, Y., Wang, Y., Ye, X., Wang, D., Yin, Y., & Wang, Y. (2020). Multi-view ensemble learning based on distance-to-model and adaptive clustering for imbalanced redit risk assessment in P2P lending. *Information Sciences, 525*, 182–204. https://doi.org/10.1016/j.ins.2020.03.027

Sun, J., Li, J., & Fujita, H. (2022). Multi-class imbalanced enterprise credit evaluation based on asymmetric bagging combined with light gradient boosting machine. *Applied Soft Computing*. https://doi.org/10.1016/j.asoc.2022.109637

Uddin, M. S., Chi, G., Al Janabi, M. A. M., & Habib, T. (2020). Leveraging random forest in micro-enterprises credit risk modelling for accuracy and interpretability. *International Journal of Finance & Economics, 27*(3), 3713–3729. https://doi.org/10.1002/ijfe.2346

Wang, D.-N., Li, L., & Zhao, D. (2022a). Corporate finance risk prediction based on LightGBM. *Information Sciences, 602*, 259–268. https://doi.org/10.1016/j.ins.2022.04.058

Wang, M., Liang, D., & Li, D. (2023a). A Two-Stage Method for Improving the Decision Quality of Consensus-Driven Three-Way Group Decision-Making. *IEEE Transactions on Systems, Man, and Cybernetics: Systems, 53*(5), 2770–2780. https://doi.org/10.1109/tsmc.2022.3222026

Wang, T., Li, H., Zhou, X., Huang, B., & Zhu, H. (2020a). A prospect theory-based three-way decision model. *Knowledge-Based Systems, 203*, 1–16. https://doi.org/10.1016/j.knosys.2020.106129

Wang, T., Liu, R., & Qi, G. (2022b). Multi-classification assessment of bank personal credit risk based on multi-source information fusion. *Expert Systems with Applications, 191*, 1–15. https://doi.org/10.1016/j.eswa.2021.116236

Wang, T., Sun, B., & Jiang, C. (2023b). Kernel similarity-based multigranulation three-way decision approach to hypertension risk assessment with multi-source and multi-level structure data. *Applied Soft Computing*. https://doi.org/10.1016/j.asoc.2023.110470

Wang, Y., He, S., Zamora, D. G., Pan, X., & Martínez, L. (2023c). A large scale group three-way decision-based consensus model for site selection of new energy vehicle charging stations. *Expert Systems with Applications, 214*, 1–12. https://doi.org/10.1016/j.eswa.2022.119107

Wang, Z., Jiang, C., Zhao, H., & Ding, Y. (2020b). Mining semantic soft factors for credit risk evaluation in peer-to-peer lending. *Journal of Management Information Systems, 37*(1), 282–308. https://doi.org/10.1080/07421222.2019.1705513

Woo, H., & Sohn, S. Y. (2022). A credit scoring model based on the Myers-Briggs type indicator in online peer-to-peer lending. *Financial Innovation, 8*(1), 1–19. https://doi.org/10.1186/s40854-022-00347-4

Xia, Y., Zhao, J., He, L., Li, Y., & Yang, X. (2021). Forecasting loss given default for peer-to-peer loans via heterogeneous stacking ensemble approach. *International Journal of Forecasting, 37*(4), 1590–1613. https://doi.org/10.1016/j.ijforecast.2021.03.002

Xu, Y., Wang, M., & Luo, F. (2023). The movement strategy of three-way decisions based on clustering. *Information Sciences, 625*, 175–191. https://doi.org/10.1016/j.ins.2023.01.015

Yang, J., & Yao, Y. (2021). A three-way decision based construction of shadowed sets from Atanassov intuitionistic fuzzy sets. *Information Sciences, 577*, 1–21. https://doi.org/10.1016/j.ins.2021.06.065

Yao, G., Hu, X., Zhou, T., & Zhang, Y. (2022a). Enterprise credit risk prediction using supply chain information: A decision tree ensemble model based on the differential sampling rate, Synthetic Minority Oversampling Technique and AdaBoost. *Expert Systems*. https://doi.org/10.1111/exsy.12953

Yao, J., Wang, Z., Wang, L., Liu, M., Jiang, H., & Chen, Y. (2022b). Novel hybrid ensemble credit scoring model with stacking-based noise detection and weight assignment. *Expert Systems with Applications, 198*, 1–12. https://doi.org/10.1016/j.eswa.2022.116913

Yao, Y. (2010). Three-way decisions with probabilistic rough sets. *Information Sciences, 180*(3), 341–353. https://doi.org/10.1016/j.ins.2009.09.021

Yao, Y. (2011). The superiority of three-way decisions in probabilistic rough set models. *Information Sciences, 181*(6), 1080–1096. https://doi.org/10.1016/j.ins.2010.11.019

Zeitsch, P. J. (2019). A jump model for credit default swaps with hierarchical clustering. *Physica a: Statistical Mechanics and Its Applications, 524*, 737–775. https://doi.org/10.1016/j.physa.2019.04.255

Zhan, J., Ye, J., Ding, W., & Liu, P. (2022). A novel three-way decision model based on utility theory in incomplete fuzzy decision systems. *IEEE Transactions on Fuzzy Systems, 30*(7), 2210–2226. https://doi.org/10.1109/tfuzz.2021.3078012

Zhang, Q., Huang, Z., & Wang, G. (2021). A novel sequential three-way decision model with autonomous error correction. *Knowledge-Based Systems, 212*, 1–14. https://doi.org/10.1016/j.knosys.2020.106526

Zhang, W., Yan, S., Li, J., Tian, X., & Yoshida, T. (2022a). Credit risk prediction of SMEs in supply chain finance by fusing demographic and behavioral data. *Transportation Research Part E: Logistics and Transportation Review*. https://doi.org/10.1016/j.tre.2022.102611

Zhang, X., Ma, Y., & Wang, M. (2023). An attention-based Logistic-CNN-BiLSTM hybrid neural network for credit risk prediction of listed real estate enterprises. *Expert Systems*. https://doi.org/10.1111/exsy.13299

Zhang, X., Yu, L., Yin, H., & Lai, K. K. (2022b). Integrating data augmentation and hybrid feature selection for small sample credit risk assessment with high dimensionality. *Computers & Operations Research, 146*, 1–14. https://doi.org/10.1016/j.cor.2022.105937

Zhang, Y., & Yao, J. (2020). Game theoretic approach to shadowed sets: A three-way tradeoff perspective. *Information Sciences, 507*, 540–552. https://doi.org/10.1016/j.ins.2018.07.058

Zhu, J., Ma, X., & Zhan, J. (2022). A regret theory-based three-way decision approach with three strategies. *Information Sciences, 595*, 89–118. https://doi.org/10.1016/j.ins.2022.02.030