

Research Article

A Novel Ensemble Credit Scoring Model Based on Extreme Learning Machine and Generalized Fuzzy Soft Sets

Dayu Xu ¹, Xuyao Zhang ², Junguo Hu,¹ and Jiahao Chen³

¹College of Information Engineering, Zhejiang A&F University, Hangzhou, Zhejiang, China

²College of Economics and Management, Zhejiang A&F University, Hangzhou, Zhejiang, China

³Fuqua School of Business, Duke University, Durham, NC, USA

Correspondence should be addressed to Xuyao Zhang; zxyzafu@163.com

Received 28 February 2020; Accepted 4 May 2020; Published 30 June 2020

Guest Editor: Zaoli Yang

Copyright © 2020 Dayu Xu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper mainly discusses the hybrid application of ensemble learning, classification, and feature selection (FS) algorithms simultaneously based on training data balancing for helping the proposed credit scoring model perform more effectively, which comprises three major stages. Firstly, it conducts preprocessing for collected credit data. Then, an efficient feature selection algorithm based on adaptive elastic net is employed to reduce the weakly related or uncorrelated variables to get high-quality training data. Thirdly, a novel ensemble strategy is proposed to make the imbalanced training data set balanced for each extreme learning machine (ELM) classifier. Finally, a new weighting method for single ELM classifiers in the ensemble model is established with respect to their classification accuracy based on generalized fuzzy soft sets (GFSS) theory. A novel cosine-based distance measurement algorithm of GFSS is also proposed to calculate the weights of each ELM classifier. To confirm the efficiency of the proposed ensemble credit scoring model, we implemented experiments with real-world credit data sets for comparison. The process of analysis, outcomes, and mathematical tests proved that the proposed model is capable of improving the effectiveness of classification in average accuracy, area under the curve (AUC), H-measure, and Brier's score compared to all other single classifiers and ensemble approaches.

1. Introduction

Nowadays, financial institutions tend to adopt different risk assessment and credit scoring models to reduce potential risk to a certain extent [1]. By analyzing customer credit data to figure out the probability that potential borrowers will default on their loans, the evaluation approaches can be utilized to turn customer data into principle, which could support credit decisions [2]. In that way, an effective credit scoring model can be a reliable supporting system to help managers in making their financial decisions.

To handle the potential risk of financial services, in the past few years, increasingly financial institutions are moving from traditional manual methods to advanced approaches that require building various types of evaluation models. For credit evaluation, three main methods, which are statistical approaches, nonparametric approaches, and AI methods,

are being widely utilized [3–7]. These three methods work efficiently in various circumstances. Statistical methods consist of different models, which include discriminant analysis models, linear probability models, and probit and logit models. Yet nonparametric approaches tend to utilize the decision tree, K -nearest neighbor algorithm, fuzzy logic, Naïve Bayes, and so on. AI methods are more advanced and technology-dependent, such as artificial neural networks, support vector machines (SVM), particle swarm optimization (PSO), and genetic algorithm (GA).

Many researches also indicated that ensemble approaches show more effective performance in the evaluation of credit than single classifiers. To avoid the downsides of single classifiers, an increasing number of researchers have switched to using customized and combined various methods instead of using individual classification models separately. The principle of the hybrid approach is to

perform preprocessing on the data input to the classifiers. The focus is to gather information from group-based classifiers based on the same issue and then export these strengths to get valid credit scoring decisions [8, 9]. In recent years, the research of fuzzy soft sets theory has made great progress, especially in the application field of multiattribute decision making [10, 11]. The development of fuzzy soft sets theory can provide a new perspective for us to build more state-of-the-art ensemble data classification and credit evaluation models [12].

The motivation of our study is to construct a more reliable credit scoring model that can generate accurate outcomes within imbalanced data. Three main approaches will be addressed to achieve this goal: (1) improved elastic net-based feature selection, (2) novel ensemble strategy and learning algorithm for imbalanced credit data, and (3) dynamic weighting method for single ELM classifiers based on new proposed similarity measure of GFSS. Because in real applications of credit risk evaluation, especially in peer-to-peer lending, credit data could be gathered from many different channels, including social networking and judicial administration platforms. The data collected from these channels are usually very sparse, redundant, rough, and imbalanced (good customers generally outnumber bad customers) and often consist of various weakly related or even uncorrelated features [13, 14]. These data characteristics will make the commonly used credit scoring models unstable, which leads to the credit evaluation results become unreliable and inaccurate. Through the above three approaches proposed in this article, the problems arising in credit scoring for imbalanced data can be handled and solved effectively.

In Section 2, we will talk about the construction of a new ensemble credit scoring model. The experimental outcomes will be discussed in Section 3. Finally, Section 4 concludes the paper.

2. New Ensemble Credit Scoring Model

This section mainly talks about the construction of the ensemble classification model for credit scoring.

2.1. Adaptive Elastic Net-Based Feature Selection. A large number of researchers have studied the appropriate feature selection approaches for credit scoring, such as cost-sensitive [15], information gain ratio [16], and genetic algorithm [17]. The Lasso estimator can reduce the regression coefficients to zero in L_1 -norm. This method can also reduce features (variables) as well as select the most important one to build simple but effective models while keeping the high efficiency. Denote historical credit scoring data as (x_i, y_i) , $i = 1, 2, \dots, N$, where $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ are variables for customers and y_i are category tags (binary responses, denote 0 as default and 1 as nondefault). The regression model could be constructed as follows:

$$y_i = \beta_0 + \sum_{j=1}^p \beta_j x_{ij}, \quad (1)$$

where β_0 and β_j are the intercept and regression coefficients, respectively. Suppose that every observation is not correlated

and that all the variables are normalized. The Lasso proper estimate of $\hat{\beta}$ could be constructed as

$$\hat{\beta} = \arg \min \left\{ \sum_{i=1}^N \left(y_i - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\}. \quad (2)$$

Based on the information above, a large λ would reduce some coefficients in β_j to zero. That is, Lasso reduces the coefficients to zero while λ gradually increases. In addition, the Lasso model is able to hold any number of variables. Therefore, both the reduction of coefficients and the selection of features (variables) can be carried out at the same time.

Although Lasso has been proved to be easily interpretable and effective under various circumstances, it still has some shortcomings [18]. Zou and Hastie [19] put forward an expansion approach called elastic net to conduct selection. Similarly, the elastic net is also able to conduct automatic selection of variables and shrinkage of coefficient at the same time and select groups of correlated variables. For any constant and nonnegative λ_1 and λ_2 , the estimation of $\hat{\beta}$ by the elastic net $\hat{\beta}_{\text{Enet}}$ could be carried out as follows:

$$\hat{\beta}_{\text{Enet}} = \arg \min \left\{ \sum_{i=1}^N \left(y_i - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda_1 \sum_{j=1}^p |\beta_j| + \lambda_2 \sum_{j=1}^p \beta_j^2 \right\}, \quad (3)$$

where $\sum_{j=1}^p |\beta_j|$ is an element of L_1 -norm and $\lambda_2 \sum_{j=1}^p \beta_j^2$ is the L_2 -norm element.

In addition, Zou and Zhang [20] also pointed out that the elastic net does not possess the oracle property. They then proposed a new adaptive elastic net which combines the L_2 penalty with the weighted L_1 penalty to penalize the squared error loss. Therefore, the adaptive elastic net could be treated as the package of the adaptive Lasso and elastic net. The valuation of $\hat{\beta}$ by the adaptive elastic net $\hat{\beta}_{\text{Enet}}$ could be calculated as

$$\hat{\beta}_{\text{Enet}} = \arg \min \left\{ \sum_{i=1}^N \left(y_i - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda_1^* \sum_{j=1}^p \hat{\omega}_j |\beta_j| + \lambda_2 \sum_{j=1}^p \beta_j^2 \right\}, \quad (4)$$

where $\hat{\omega}_j = (|\hat{\beta}_{\text{Enet}}|)^{-\gamma}$, γ is positive, while λ_1^* is fixed and nonnegative.

Using formula (4), we will be capable of obtaining the most significant attributes ("big fish") from the variable pool. Then, we can plug them into credit scoring models to get a more precise result but with minimum computational and operational cost.

2.2. ELM-Based Classifier. ELM model, as a single-hidden layer feedforward neural network (SLFN), can select the input weight and hidden biases randomly without any adjustment during its process. The Moore–Penrose generalized inverse matrices of the hidden-layer output matrix can be utilized to analyze and determine the output weights. ELM exhibits excellent performance of generalizing and the reduction in the iterative time of training process. Clearly, it is more effective than any other ANN-type machine learning algorithms [21].

For the historical training credit data set (x_i, y_i) that is mentioned above, the input vector $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T \in R^p$ is the i^{th} sample with p -dimensional features, and $Y = [y_1, y_2, \dots, y_N]$. Then, p is the amount of input neurons. p is also equivalent to the input features. Let L be the amount of hidden neurons. Denote C as the amount of output neurons, which is also equivalent to the category number. Denote the input weight matrix as $K = [k_1, k_2, \dots, k_L]$, where $k_j = [k_{j1}, k_{j2}, \dots, k_{jn}]$ is the vector connecting the p input neurons with the j^{th} hidden neuron. $b = [b_1, b_2, \dots, b_j, \dots, b_L]^T$ is the bias value of the hidden neurons, where b_j is the bias value of the j^{th} hidden neurons. The above parameters do not change during the whole process. The output could be computed by as follows:

$$h(x_i) = G(Kx_i + b), \quad (5)$$

where $G(x)$ is the activation function. Let H be the output of all the samples. It can be calculated by using the following equation:

$$H = \begin{bmatrix} h_1(x_1) & h_2(x_1) & \cdots & h_L(x_1) \\ h_1(x_2) & h_2(x_2) & \cdots & h_L(x_2) \\ \vdots & \vdots & \cdots & \vdots \\ h_1(x_N) & h_2(x_N) & \cdots & h_L(x_N) \end{bmatrix}_{N \times L}. \quad (6)$$

The i^{th} column represents the i^{th} hidden nodes output vector relative to the inputs x_1 to x_N . The j^{th} row represents the output vector of the hidden layer relative to the input x_j . The output of ELM can be calculated by

$$f(x) = \sum_{i=1}^L \alpha_i G(Kx_i + b) = o_j, \quad j = 1, 2, \dots, N, \quad (7)$$

where $\alpha_i = [\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{iC}]^T$ is the weight vector that connects the i^{th} hidden nodes with the output nodes. ELM is able to evaluate those N samples without any mistake. In other words, $\sum_{j=1}^L \|o_j - y_j\| = 0$. Then, the following equation can be obtained:

$$\sum_{i=1}^L \alpha_i G(Kx_i + b) = y_j, \quad j = 1, 2, \dots, N. \quad (8)$$

Equation (8) can also be rewritten as follows:

$$H\alpha = Y. \quad (9)$$

Based on (9), the value of output weight could be estimated using a least square solution as follows:

$$\hat{\alpha} = H^\dagger Y, \quad (10)$$

where H^\dagger stands for the Moore–Penrose generalized inverse of H . For credit scoring classification, the outcome of ELM is as follows:

$$f(x) = \text{sign} \left(\sum_{i=1}^L \hat{\alpha}_i G(Kx_i + b) \right) = \text{sign}(\hat{\alpha} \cdot h(x)). \quad (11)$$

2.3. Ensemble Strategy for Imbalanced Data. To better solve the classification of imbalanced data, a considerable number

of approaches have been used. They can be categorized into three types: preprocessing, cost-sensitive learning, and ensemble methodology. Preprocessing is able to decrease the classification bias based on the bias-variance decomposition to enhance the single classifier. Undersampling [22, 23], oversampling [24, 25], and strategic sampling are extensively utilized to offset imbalanced data.

Ensemble methodology can be viewed as a decision-making process that combines both individual learning algorithms and their outcomes in parallel to obtain the ultimate result. The basic idea behind the ensemble methodology is that the algorithm will get a number of single classifiers from the training set, and then it uses some ensemble strategies to integrate them to raise the accuracy and reliability of classification. Bagging [26], boosting [27], and stacking [28] are the most common ensemble approaches in credit scoring.

A novel ensemble strategy is planned for imbalanced data according to its imbalance ratio, which could determine the number of ELM classifiers that apply as single classifiers to predict credit scoring data, as well as the number of samples that feed into each ELM as training data.

For any given historical credit training data set that contains N samples, there are N^+ “good applicants” and N^- “bad applicants,” such that $N^+ + N^- = N$. Then, the imbalance ratio is called IR, which can be calculated as follows:

$$\text{IR} = \frac{N^+}{N^-}. \quad (12)$$

After obtaining the IR, the amount of single ELM classifiers M in the ensemble model also can be calculated as follows:

$$M = \lceil \text{IR} \rceil = \left\lceil \frac{N^+}{N^-} \right\rceil, \quad (13)$$

where the symbol $\lceil \cdot \rceil$ represents “ceiling” operation.

Equation (13) can help us to not only determine the number of ELM classifiers needed in the ensemble credit scoring model but will also guide us to make the imbalanced data become balanced for each classifier. Regarding the ensemble strategy for imbalanced data, we proposed M -based value. In the remainder of this subsection, we will elaborate the proposed strategy in detail.

Firstly, calculate the imbalance ratio IR for any given historical credit training data set on the basis of N^+ and N^- . Secondly, determine the number of ELM classifiers M using (13).

Thirdly, for the first $M-1$ ELM classifiers, we feed N^- “good applicants” samples and N^- “bad applicants” samples into the ELM classifiers to make sure that the training data sets of the first $M-1$ classifiers are balanced, and random sampling without replacement method is employed to extract N^- samples from N^+ “good applicants” samples for each classifier. After finishing the first $M-1$ training data sets extraction, there are $N^+ - (M-1) \cdot N^-$ “good applicants” sample that have not been extracted.

Finally, for the last ELM classifier, we let the remaining $N^+ - (M-1) \cdot N^-$ “good applicants” samples into the training data sets. Considering that $N^+ - (M-1) \cdot N^- < N^-$, we will employ the SMOTE algorithm to create $M \cdot N^- - N^+$ “good

applicants” samples from N^+ “good applicants” samples. Thus, for the last ELM classifier, there are still N^- “good applicants” samples and N^- “bad applicants” samples fed into it as the training data set.

Through the described processes above, the ensemble strategy for the imbalanced data we proposed has been realized, and the training data for each classifier is balanced. In the next subsection, we will further introduce the GFSS theory-based ensemble credit scoring approach, which utilizes the results of each single ELM classifier.

2.4. GFSS Theory-Based Ensemble Credit Scoring Model. Since we have the results of each ELM classifier, we also need to figure out the weights with respect to their performance. The accuracy of classification is expected to be greatly improved. The theory of soft sets, which is firstly put forward by Molodtsov [29], can be regarded as a way for solving the uncertainties in imprecise environments (e.g., credit scoring area). Maji et al. [30] launched a research focusing on both fuzzy and soft sets. We will firstly introduce the principle of generalized fuzzy soft sets and then put forward a similarity measure of generalized fuzzy soft sets using angular cosine. After that, we can get the weights of each credit scoring model using similarity measure and the accuracy of classification. Finally, we are able to build the generalized fuzzy soft sets theory-based ensemble credit scoring model.

2.4.1. GFSS Theory. Based on the theories proposed by Molodtsov [29] and Maji et al. [30], we can make some definitions for fuzzy soft sets.

Definition 1. Denote U as the initial universal set. Denote P as a set of parameters. $P(U)$ is the power set of U . (F, P) is a soft set over U if P is a mapping given by $F: P \rightarrow P(U)$.

Definition 2. Denote U as the initial universal set. Denote P as a set of parameters. The power set of all fuzzy subsets of U is I^U . Let $A \subset P: A$ pair (F, P) be the fuzzy soft set over U if F is a mapping given by $F: A \rightarrow I^U$.

Then, Maji and Samanta’s [31] definition of GFSS is as follows:

Definition 3. Denote U as the initial universal set. Denote P as a set of parameters. Let (U, P) be the soft universe. Denote $F: P \rightarrow I^U$ and μ as the fuzzy subset of P , i.e., $\mu: P \rightarrow I = [0, 1]$, where I^U is all fuzzy subsets of U . Denote F_μ as the mapping given by $F_\mu: P \rightarrow I^U \times I$, which can also be denoted as $F_\mu(e) = (F(e), \mu(e))$, where $F(e) \in I^U$. In this way, F_μ can be viewed as a GFSS over the soft universe (U, P) .

For every e_i , $F_\mu(e_i) = (F(e_i), \mu(e_i))$ illustrates both the level of belonging of the subsets of U to $F(e_i)$ and the possibility of belonging.

In this paper, U denotes the historical data of customer credit $X = \{x_1, x_2, \dots, x_n\}$, where $x_i \in R^p$, and $F(e_i)$ represents the performance of the classification by a single customer from a single classifier and $\mu(e_i)$ denotes the overall degree of classification of a certain single classifier.

TABLE 1: Confusion matrix.

Real class (%)	Projected class (%)	
	Good	Bad
Good	TP _m	FN _m
Bad	FP _m	TN _m

2.4.2. Similarity Measure of GFSS. It is important to address the similarity measurement of GFSS during the setting of GFSS and the establishment of our model. Thus, we establish a new approach for the similarity measure of GFSS for the scoring of credits.

Definition 4. For M single classifiers, denote η_t and μ_m as the elements of F_μ . The definition is as follows:

$$\eta_t^m = \frac{1 - |\hat{y}_t^m - y_t|}{1 + |\hat{y}_t^m - y_t|}, \quad (14)$$

where y_t ($t = 1, 2, \dots, T$) is the category tag of the t th customer (binary responses denote 0 as default and 1 as nondefault); \hat{y}_t^m ($m = 1, 2, \dots, M$) is the forecasting result of the t th customer predicted by the m th classifier and is between 0 and 1. η_t^m is the accuracy degree of classification of a single classifier for a single customer. This accuracy ranges between 0 and 1. This is in line with our initial intuition. In addition, μ_m is the m th classifier’s overall classification performance, which can be calculated as follows:

$$\mu_m = \sqrt{\frac{TP_m}{TP_m + FN_m} \cdot \frac{TN_m}{TN_m + FP_m}}, \quad (15)$$

where TP_m, FN_m, TN_m, and FP_m are elements of the confusion matrix (Table 1). TP_m is the amount of good customers accurately labeled as good, TN_m is the amount of bad customers accurately labeled as bad, FN_m is the amount of good customers falsely labeled as bad, and FP_m is the amount of bad customers falsely labeled as good. The greater the value of μ_m , the more accurate the result is calculated by the m th classifier.

Based on the above discussion, it can be noted that η_{it} and μ_m from Definition 4 are able to evaluate the performance of classification of every single model. Therefore, we can build the GFSS $F_\mu^m(\eta, \mu)$ of the m th classifier as follows:

$$F_\mu^m(\eta, \mu) = \left(\left(\frac{y_1}{\eta_1^m}, \frac{y_2}{\eta_2^m}, \dots, \frac{y_t}{\eta_t^m}, \dots, \frac{y_T}{\eta_T^m} \right), \mu_m \right). \quad (16)$$

Definition 5. For two GFSS F_μ^i and F_μ^j over the universal set U , where $i, j \in [1, M]$, the similarity measurement of GFSS SG(F_μ^i, F_μ^j) can be calculated as follows:

$$SG(F_\mu^i, F_\mu^j) = \frac{\sum_{t=1}^T (\eta_t^i \mu_i) \times (\eta_t^j \mu_j)}{\sqrt{\sum_{t=1}^T (\eta_t^i \mu_i)^2} \sqrt{\sum_{t=1}^T (\eta_t^j \mu_j)^2}} \quad (17)$$

2.4.3. Ensemble Credit Scoring Modeling. The determination of the weight for every single model is the most important

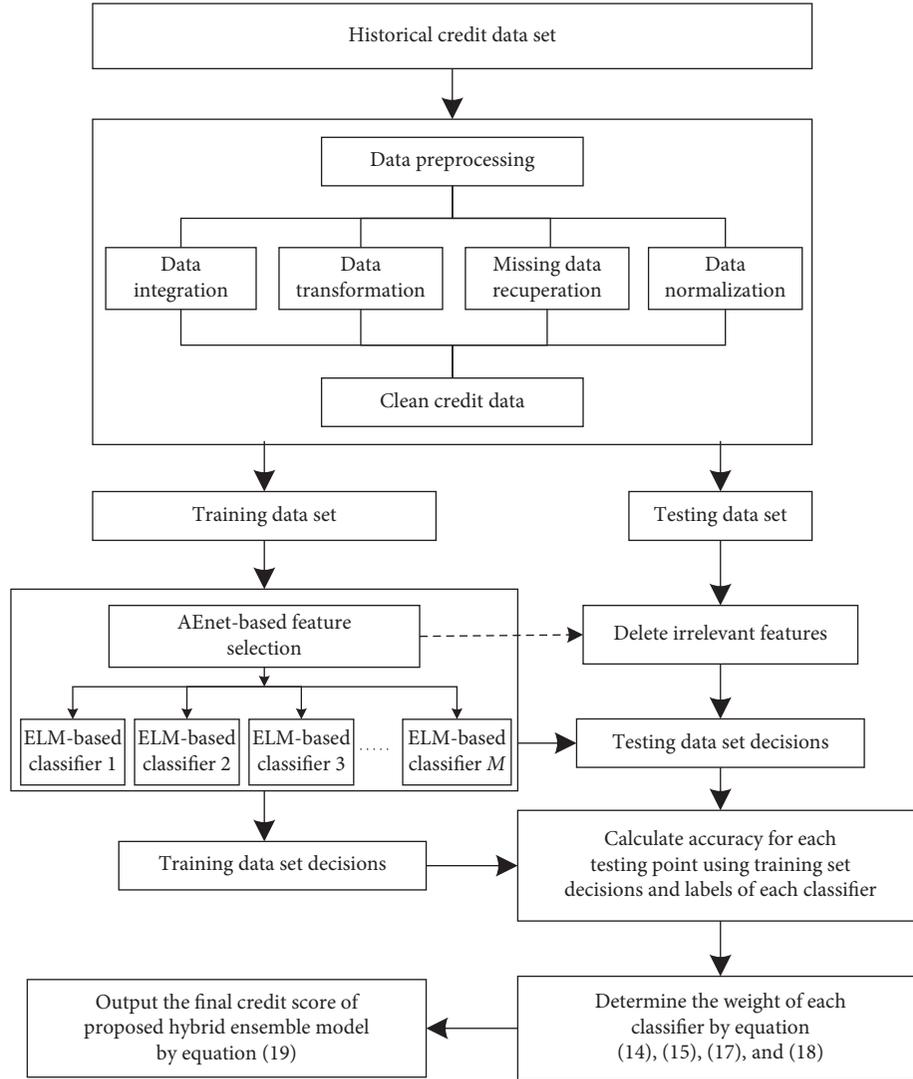


FIGURE 1: Flow diagram of the proposed EGHE model.

step in the establishment of the ensemble credit score modeling. The purpose of this determination is to compare the calculated credit score with the actual records.

The similarity measure of GFSS can be utilized to do the determination, i.e., the calculated $SG(F_\mu^m, F_\mu^*)$ value, where F_μ^* is the GFSS of the actual score of customers. In this way, the weight of m th classifier can be calculated by

$$\bar{\omega}^m = \frac{SG(F_\mu^m, F_\mu^*)}{\sum_{m=1}^M SG(F_\mu^m, F_\mu^*)}, \quad (18)$$

where $0 \leq \bar{\omega}^m \leq 1$ and $\sum_{m=1}^M \bar{\omega}^m = 1$. Thus, the final score could be calculated as

$$\hat{y}_t = \sum_{m=1}^M \bar{\omega}^m \hat{y}_t^m. \quad (19)$$

Figure 1 presents the flow-process diagram of the ensemble model. The algorithm of “ELM and GFSS Theory-Based Hybrid Ensemble” model is described. We will call this EGHE in the following parts (Algorithm 1).

3. Results and Discussion

3.1. Preparation of Dataset. During the evaluation, we collected many different private and public data sets. We have collected a total of six credit data sets and four additional imbalanced data sets with different IR are obtained, i.e., three public and three private. The public sets can be obtained from the UCI Machine Learning Repository. They are real-world credit score data sets and are now widely used by researchers. The German, Australian, and Japanese data sets are used for extra verification. The private data sets consist of the Iranian data set that has also been widely used in many studies and the Bene 1 and 2 data sets, which can be obtained from two key financial institutions in Benelux et al. [32]. This Iranian set has various customer data of many small Iranian private banks [33, 34]. Four additional imbalanced datasets are also from the Machine Learning Repository, UCI. They are Shuttle, Skin_segment, MiniBooNE, and LC2017Q1 which contains loan data of the first quarter in 2017 from Lending Club. The characteristics of all experimental data sets can be found in Table 2.

Input: Historical data of credit scoring (x_i, y_i) .
Output: Score of every single customer \hat{y}_t .
Step 1. Preprocessing of data.
Step 2. Variables selection using AEnet.
Step 3. Imbalanced data rebalancing by using the proposed ensemble strategy.
Step 4. Credit scoring of every single ELM classifier.
Step 5. Compute η_t and μ_m using (14) and (15).
Step 6. Calculate the SG (F_μ^m, F_μ^*) of single classifier using (17).
Step 7. Calculate the weight of m th classifier using (18).
Step 8. Get the final credit score of every customer \hat{y}_t .

ALGORITHM 1: EGHE algorithm.

TABLE 2: Characteristics of data sets.

Data sets	Size	Attributes	Good/bad	Imbalance ratio	Number of classifiers
<i>Credit scoring data sets</i>					
Germany	1000	25	700/300	2.33	3
Australia	690	14	307/383	0.80	1
Japan	690	15	296/357	0.83	1
Iran	1000	27	950/50	19	19
Bene 1	3,123	33	2,082/1,041	2	2
Bene 2	7,190	33	5,033/2,157	2.33	3
<i>Additional imbalanced data sets</i>					
Shuttle	12,380	9	11,428/952	12.004	13
Skin_segment	117,728	3	114,039/3,679	30.99	31
MiniBooNE	201,355	10	196,555/4,800	40.95	41
LC2017Q1	95,633	72	94,414/1,219	77.45	78

In this paper, we compared our proposed model with the other four state-of-the-art models, namely, C5.0 decision tree, SVM with Radial Basis Function, kernel SVM-R, Deep Belief Networks (DBN), and Bayes to validate the performance of our approaches. All the continuous attributes will be discretized into various intervals. Every single data set will be divided into a two-thirds training set and one-third testing set randomly. We use the open source platform *R-statistics* (version R-3.2.2.) to conduct our experiments.

3.2. Experimental Results. Different methods are utilized as comparison models to test the validity of the EGHE credit scoring model.

Firstly, the FS algorithm that is based on AEnet is used to obtain the highly correlated variables after initial data gathering and preprocessing. We could notice that after selection, the variables in these ten data sets are all decreased to various degrees (Table 3). In consideration of the complexity of computation, deleting irrelevant or weakly correlated variables is becoming increasingly important for big-data-oriented credit assessment issues.

After feature selection, we can perform two experiments to further see the effect of feature selection on the classification of each model: (1) all single classifiers with feature selection and (2) all single classifiers without feature selection. Tables 4 and 5 show in detail the AUC, H-measure (HM), Brier's score (BS), and accuracy (ACC) for all single classifiers with and without feature selection.

TABLE 3: Number of features before and after AEnet-based selection.

Data sets	Number of features	
	Before	After
Germany	25	21
Australia	14	10
Japan	15	12
Iran	27	22
Bene 1	33	15
Bene 2	33	20
Shuttle	16	13
Skin_segment	3	3
MiniBooNE	10	9
LC2017Q1	72	48

From the results of every single classifier in Tables 4 and 5, we can see that the feature selection helps enhance the effectiveness of classification of most single classifiers. After feature selection, C5.0 shows accuracy values that are 0.8%, 1.76%, 1.63%, 4.49%, 2.85%, 0.82%, 3.47%, 4.15%, 7.94%, and 4.12% greater than without feature selection for Germany, Australia, Japan, Iran, Bene 1, Bene 2, Shuttle, Skin_segment, MiniBooNE, and LC2017Q1 data sets, respectively. In the same way, SVM-R increased the accuracy values by 0.93%, 1.08%, 0.67%, 4.53%, -0.4%, 1.94%, 0.87%, 1.02%, 2.03%, and 2.54%, respectively, for ten experimental data sets. DBN and Naïve Bayes improve their effectiveness too after feature selection. Only SVM-R on Bene 1 reduced

TABLE 4: Results for all single classifiers without AEnet-based feature selection.

Data set	Performance measures	Single classifiers				
		C5.0	SVM-R	DBN	Bayes	ELM
Germany	AUC	0.680	0.732	0.753	0.761	0.764
	HM	0.336	0.362	0.399	0.338	0.379
	BS	0.251	0.167	0.173	0.199	0.172
	ACC	0.745	0.752	0.749	0.745	0.762
Australia	AUC	0.867	0.812	0.876	0.857	0.852
	HM	0.516	0.603	0.613	0.581	0.610
	BS	0.148	0.113	0.107	0.166	0.127
	ACC	0.736	0.742	0.751	0.734	0.735
Japan	AUC	0.857	0.908	0.911	0.888	0.882
	HM	0.689	0.608	0.608	0.667	0.618
	BS	0.156	0.120	0.138	0.176	0.142
	ACC	0.738	0.749	0.757	0.742	0.751
Iran	AUC	0.616	0.604	0.614	0.715	0.710
	HM	0.111	0.074	0.061	0.192	0.133
	BS	0.071	0.071	0.078	0.077	0.083
	ACC	0.623	0.617	0.640	0.627	0.648
Bene 1	AUC	0.728	0.821	0.767	0.740	0.752
	HM	0.310	0.366	0.250	0.335	0.351
	BS	0.267	0.176	0.168	0.297	0.199
	ACC	0.701	0.749	0.698	0.690	0.753
Bene 2	AUC	0.745	0.789	0.746	0.707	0.762
	HM	0.385	0.309	0.238	0.276	0.270
	BS	0.143	0.178	0.142	0.172	0.153
	ACC	0.728	0.720	0.715	0.711	0.747
Shuttle	AUC	0.604	0.675	0.693	0.714	0.722
	HM	0.299	0.334	0.367	0.317	0.358
	BS	0.224	0.153	0.158	0.186	0.163
	ACC	0.662	0.693	0.740	0.698	0.720
Skin_segment	AUC	0.770	0.749	0.804	0.802	0.805
	HM	0.458	0.556	0.564	0.543	0.576
	BS	0.131	0.105	0.099	0.156	0.119
	ACC	0.653	0.685	0.689	0.687	0.695
MiniBooNE	AUC	0.761	0.837	0.838	0.833	0.833
	HM	0.613	0.560	0.559	0.625	0.584
	BS	0.140	0.112	0.126	0.164	0.134
	ACC	0.655	0.690	0.696	0.695	0.710
LC2017Q1	AUC	0.547	0.557	0.563	0.669	0.671
	HM	0.099	0.067	0.057	0.181	0.126
	BS	0.062	0.066	0.072	0.071	0.078
	ACC	0.655	0.670	0.688	0.687	0.712

by 0.4%, but this does not contradict the improvement on the classification that our AEnet-based feature selection brings. Not only do outliers, redundant, and weakly or even unrelated variables help improve the effectiveness, but also affect the model establishment and cause great computation.

It is noteworthy that, compared with C5.0, SVM-R, DBN, and Bayes, ELM has manifested the superiority in accuracy on the vast majority of data sets. Table 6 reports the average running time (total time for training and testing) for all models. For these experiments, we use an Intel i5-8500 with CPU at 3.0 GHz and 16 GB of RAM.

From Table 6, we can see that ELM costs less time than other single models to carry out credit scoring activities. The efficiency of computing resources also makes ELM a great match for ensemble learning and modeling.

After implementing feature selection, completing the ensemble strategy, and individual model classification, the EGHE model can be achieved. Based on (14), (15), and (17)–(19), the weights of single ELM classifiers are calculated according to their efficiency, respectively. To validate the availability of EGHE, we employed several ensemble models in contrast with EGHE. These models were split into two parts. The first one contains four FS algorithms with GFSS-based combination. They are cost-sensitive, GA, information gain ratio (IGR), and elastic net (Enet). Cost-sensitive, GA, and IGR are popular feature selection approaches in the credit scoring area [16, 35, 36]. The second group applies four other approaches with AEnet-based feature selection, which were weighted average (WAVG) [37], majority voting (MajVot) [38], weighted voting (WVOT) [39], and fuzzy soft set (FSS). Those methods are frequently adopted in the

TABLE 5: Results for all single classifiers with AEnet-based feature selection.

Data set	Performance measurement	Single classifiers				
		C5.0	SVM-R	DBN	Bayes	ELM
Germany	AUC	0.690	0.756	0.762	0.774	0.768
	HM	0.181	0.295	0.248	0.267	0.268
	BS	0.230	0.163	0.172	0.192	0.173
	ACC	0.751	0.759	0.752	0.758	0.765
Australia	AUC	0.882	0.902	0.915	0.912	0.883
	HM	0.614	0.635	0.642	0.623	0.624
	BS	0.121	0.104	0.131	0.121	0.130
	ACC	0.749	0.750	0.755	0.741	0.748
Japan	AUC	0.861	0.910	0.919	0.910	0.878
	HM	0.608	0.621	0.631	0.622	0.622
	BS	0.116	0.112	0.126	0.113	0.119
	ACC	0.749	0.752	0.766	0.753	0.758
Iran	AUC	0.631	0.650	0.612	0.723	0.638
	HM	0.137	0.118	0.108	0.212	0.117
	BS	0.068	0.071	0.078	0.072	0.081
	ACC	0.652	0.645	0.650	0.652	0.658
Bene 1	AUC	0.770	0.810	0.806	0.773	0.798
	HM	0.312	0.345	0.336	0.302	0.320
	BS	0.233	0.177	0.184	0.264	0.241
	ACC	0.721	0.746	0.719	0.721	0.753
Bene 2	AUC	0.763	0.824	0.815	0.780	0.796
	HM	0.354	0.386	0.269	0.285	0.326
	BS	0.132	0.119	0.137	0.125	0.137
	ACC	0.734	0.734	0.715	0.723	0.747
Shuttle	AUC	0.652	0.682	0.686	0.723	0.729
	HM	0.323	0.337	0.363	0.321	0.362
	BS	0.242	0.155	0.156	0.188	0.165
	ACC	0.685	0.699	0.743	0.707	0.727
Skin_segment	AUC	0.832	0.756	0.796	0.812	0.813
	HM	0.495	0.562	0.558	0.550	0.582
	BS	0.141	0.106	0.098	0.158	0.120
	ACC	0.675	0.692	0.712	0.696	0.702
MiniBooNE	AUC	0.822	0.845	0.830	0.844	0.841
	HM	0.662	0.566	0.553	0.633	0.589
	BS	0.151	0.113	0.125	0.166	0.135
	ACC	0.707	0.704	0.689	0.704	0.717
LC2017Q1	AUC	0.591	0.563	0.557	0.678	0.678
	HM	0.107	0.068	0.056	0.183	0.125
	BS	0.067	0.067	0.071	0.072	0.079
	ACC	0.682	0.687	0.681	0.696	0.719

TABLE 6: Running time for all models with feature selection (seconds).

Data sets	Models				
	C5.0	SVM-R	DBN	Bayes	ELM
Germany	1.76	2.89	4.33	1.72	1.38
Australia	1.16	2.02	2.99	1.07	0.97
Japan	1.25	2.05	2.90	1.13	0.94
Iran	1.82	3.01	4.31	1.69	1.44
Bene 1	5.63	10.11	14.29	5.85	4.62
Bene 2	12.71	21.13	28.96	12.32	10.01
Shuttle	21.79	35.78	53.61	21.29	17.11
Skin_segment	207.20	340.24	509.76	202.49	182.46
MiniBooNE	354.38	581.92	871.87	346.33	277.87
LC2017Q1	591.26	970.78	1254.49	560.57	436.04

establishment and utilization of different combination models. They also employ ELM as the classifier but did not take the ensemble strategy that is proposed above, only using

random sampling methods to make all training data sets become balanced. Table 7 displays the results of AUC, H-measure, and Brier's score for all ensemble models.

TABLE 7: Classification outcomes from all data sets for different performance measures with various combinations.

Data set	Performance measurement	Other feature selection methods with GFSS-based combination				Traditional combination methods with AEnet-based feature selection				EGHE
		Cost-sensitive	GA	IGR	Enet	WAVG	MajVot	WVOT	FSS	
Germany	AUC	0.786	0.792	0.777	0.781	0.777	0.786	0.789	0.771	0.823
	HM	0.286	0.177	0.226	0.238	0.222	0.285	0.246	0.305	0.325
	BS	0.166	0.208	0.197	0.182	0.181	0.183	0.192	0.157	0.184
Australia	ACC	0.842	0.862	0.833	0.827	0.845	0.859	0.836	0.844	0.886
	AUC	0.921	0.927	0.933	0.928	0.921	0.922	0.932	0.928	0.945
	HM	0.667	0.519	0.573	0.628	0.637	0.637	0.632	0.653	0.659
	BS	0.092	0.145	0.173	0.103	0.101	0.112	0.107	0.097	0.095
Japan	ACC	0.872	0.878	0.882	0.863	0.872	0.873	0.876	0.874	0.895
	AUC	0.923	0.928	0.918	0.932	0.918	0.927	0.928	0.925	0.937
	HM	0.650	0.492	0.566	0.618	0.602	0.642	0.617	0.648	0.660
	BS	0.099	0.157	0.174	0.109	0.121	0.112	0.116	0.103	0.098
Iran	ACC	0.871	0.862	0.861	0.868	0.853	0.864	0.873	0.868	0.904
	AUC	0.791	0.778	0.781	0.767	0.777	0.779	0.783	0.778	0.802
	HM	0.279	0.150	0.217	0.162	0.283	0.108	0.107	0.294	0.303
	BS	0.042	0.069	0.056	0.049	0.044	0.048	0.049	0.044	0.059
Bene 1	ACC	0.883	0.881	0.874	0.861	0.867	0.885	0.877	0.883	0.915
	AUC	0.843	0.812	0.822	0.821	0.882	0.879	0.886	0.888	0.881
	HM	0.396	0.263	0.338	0.258	0.324	0.441	0.385	0.447	0.351
	BS	0.161	0.240	0.256	0.197	0.188	0.159	0.198	0.154	0.173
Bene 2	ACC	0.881	0.872	0.864	0.872	0.872	0.872	0.865	0.876	0.896
	AUC	0.921	0.838	0.858	0.868	0.878	0.844	0.876	0.883	0.888
	HM	0.537	0.401	0.478	0.489	0.505	0.436	0.432	0.496	0.506
	BS	0.091	0.136	0.168	0.122	0.102	0.115	0.112	0.103	0.114
Shuttle	ACC	0.878	0.860	0.857	0.868	0.881	0.875	0.882	0.884	0.898
	AUC	0.896	0.914	0.914	0.922	0.897	0.919	0.917	0.906	0.943
	HM	0.653	0.510	0.562	0.625	0.620	0.635	0.622	0.636	0.658
	BS	0.091	0.143	0.170	0.103	0.099	0.110	0.104	0.096	0.095
Skin_segment	ACC	0.852	0.866	0.862	0.858	0.849	0.870	0.862	0.851	0.916
	AUC	0.902	0.913	0.899	0.926	0.896	0.922	0.915	0.903	0.936
	HM	0.637	0.486	0.555	0.615	0.586	0.638	0.607	0.631	0.659
	BS	0.096	0.154	0.171	0.107	0.119	0.110	0.113	0.098	0.098
MiniBooNE	ACC	0.853	0.849	0.844	0.863	0.833	0.862	0.859	0.845	0.908
	AUC	0.773	0.767	0.764	0.763	0.757	0.775	0.770	0.758	0.800
	HM	0.272	0.149	0.213	0.159	0.277	0.109	0.104	0.286	0.302
	BS	0.042	0.067	0.054	0.048	0.0439	0.047	0.047	0.041	0.059
LC2017Q1	ACC	0.875	0.877	0.885	0.876	0.884	0.880	0.893	0.890	0.903
	AUC	0.823	0.801	0.804	0.816	0.859	0.874	0.872	0.867	0.879
	HM	0.386	0.260	0.332	0.257	0.315	0.440	0.378	0.435	0.350
	BS	0.158	0.237	0.252	0.195	0.182	0.157	0.196	0.149	0.203
	ACC	0.861	0.888	0.877	0.875	0.871	0.878	0.871	0.873	0.912

From Tables 7 and 5, we can see that, compared with the single classifiers, ensemble methods reveal significant advantages with regard to the accuracy of classification. Compared with other single classifiers and combined approaches in both groups, EGHE has an advantage in all metrics across all datasets. Experiments on several state-of-the-art ensemble models are performed to verify the effectiveness of the EGHE model. They are an EMPNGA-based multistage hybrid model put forward by Zhang and Xia [37]; the heterogeneous ensemble credit model put forward by Xia et al. [40]; EBCA-RF&XGB-PSO model that is put forward by He et al. [41]; heterogeneous ensemble learning-based two-stage credit risk model (TSHE) proposed by Papouskova and Hajek [42]; twin neural networks (TNN) proposed by Jayadeva et al. [43]; and a new rule-based knowledge extraction (RKE)

method proposed by Mahani and Baba [44] recently. Table 8 gives the results of ensemble models in different data sets.

From Table 8, we could tell that the results of these models are very close. The accuracy of EGHE model is better than most of the other models but Iranian. The EBCA-RF&XGB-PSO model achieved a high accuracy of 0.921 in the Iranian data set because it uses the Extended Balance Cascade method that can effectively solve the issue of class imbalance. However, the ensemble strategy and GFSS theory-based model EGHE can deal with the thorny problem of unbalanced data classification better in most experimental data sets; even in some severely skewed data sets, ideal outcomes have been achieved, such as in Shuttle, Skin_segment, MiniBooNE, and LC2017Q1.

TABLE 8: Comparison results of ensemble models in different data sets.

Data set	Performance measurement	Ensemble models						
		EMPNGA-based model	Heterogeneous ensemble	EBCA-RF& XGB-PSO	TSHE	TNN	RKE	EGHE
Germany	AUC	0.802	0.795	0.798	0.775	0.811	0.790	0.823
	HM	0.400	0.386	0.397	0.376	0.320	0.376	0.325
	BS	0.158	0.164	0.163	0.184	0.181	0.158	0.184
	ACC	0.768	0.859	0.869	0.839	0.873	0.874	0.886
Australia	AUC	0.940	0.923	0.931	0.933	0.931	0.929	0.945
	HM	0.672	0.648	0.661	0.628	0.649	0.676	0.659
	BS	0.092	0.101	0.092	0.101	0.095	0.099	0.096
	ACC	0.875	0.842	0.861	0.852	0.882	0.877	0.895
Japan	AUC	0.932	0.925	0.919	0.935	0.923	0.921	0.937
	HM	0.665	0.651	0.636	0.649	0.650	0.649	0.660
	BS	0.095	0.091	0.097	0.090	0.097	0.090	0.098
	ACC	0.872	0.883	0.889	0.887	0.890	0.887	0.904
Iran	AUC	0.876	0.824	0.831	0.824	0.790	0.821	0.802
	HM	0.424	0.384	0.489	0.384	0.298	0.416	0.303
	BS	0.058	0.047	0.061	0.043	0.058	0.053	0.059
	ACC	0.907	0.908	0.921	0.902	0.911	0.910	0.915
Bene 1	AUC	0.824	0.821	0.818	0.821	0.868	0.819	0.881
	HM	0.376	0.383	0.442	0.423	0.346	0.384	0.351
	BS	0.152	0.147	0.147	0.150	0.169	0.139	0.173
	ACC	0.872	0.869	0.869	0.865	0.883	0.872	0.896
Bene 2	AUC	0.866	0.871	0.863	0.881	0.874	0.875	0.887
	HM	0.479	0.488	0.465	0.498	0.498	0.490	0.506
	BS	0.102	0.098	0.101	0.118	0.112	0.110	0.114
	ACC	0.871	0.867	0.875	0.887	0.885	0.882	0.898
Shuttle	AUC	0.913	0.870	0.850	0.865	0.929	0.850	0.943
	HM	0.442	0.405	0.500	0.403	0.648	0.431	0.658
	BS	0.060	0.050	0.062	0.045	0.095	0.055	0.096
	ACC	0.886	0.878	0.892	0.886	0.902	0.892	0.916
Skin_segment	AUC	0.859	0.867	0.837	0.862	0.921	0.848	0.935
	HM	0.392	0.404	0.452	0.444	0.649	0.397	0.659
	BS	0.158	0.155	0.150	0.158	0.096	0.143	0.098
	ACC	0.870	0.877	0.889	0.898	0.894	0.893	0.908
MiniBooNE	AUC	0.903	0.919	0.883	0.925	0.788	0.906	0.800
	HM	0.499	0.515	0.476	0.523	0.297	0.507	0.302
	BS	0.106	0.103	0.103	0.124	0.058	0.114	0.059
	ACC	0.890	0.895	0.901	0.897	0.889	0.901	0.903
LC2017Q1	AUC	0.858	0.846	0.822	0.923	0.866	0.845	0.879
	HM	0.402	0.274	0.340	0.368	0.345	0.266	0.350
	BS	0.165	0.250	0.258	0.213	0.199	0.202	0.203
	ACC	0.888	0.897	0.903	0.895	0.898	0.903	0.912

4. Conclusion

In this paper, we proposed a novel ensemble credit scoring model called EGHE, which integrates efficient feature selection algorithm, novel ensemble strategy, and GFSS-based weighting method for single ELM classifiers. In the proposed model, the adaptive elastic net-based feature selection algorithm was firstly utilized to obtain high-quality training data to improve the evaluation efficiency without reducing the predictive precision. ELM model was employed as basic classifier, and a novel ensemble strategy was generated to make the imbalanced training data sets become balanced for each ELM classifier. Additionally, we proposed a new weighting method to build the GFSS theory-based ensemble credit scoring model. Dual-scale classification accuracy metric that is based on new similarity measurement of GFSS was constructed to compute the final weight of every

single classifier. The biggest contribution of this paper is that the proposed EGHE is able to predict credit risk reliably and accurately, especially for unbalanced credit data. Comparisons between EGHE and other credit scoring models were implemented on ten real-world datasets with four metrics (average accuracy, AUC, H-measure, and Brier's score). A variety of state-of-the-art ensemble models were employed to compare with EGHE to prove its validity. The experiments results demonstrated that the proposed EGHE model was robust and represented a positive development in credit scoring.

Data Availability

(1) The "Germany" data set used to support the findings of this study are included within the following URL: <http://archive.ics.uci.edu/ml/datasets/Statlog+%28German+Credit+Data%29>.

(2) The “Australia” data set used to support the findings of this study are included within the following URL: <http://archive.ics.uci.edu/ml/datasets/Statlog+%28Australian+Credit+Approval%29>. (3) The “Japan” data set used to support the findings of this study are included within the following URL: <https://archive.ics.uci.edu/ml/datasets/Japanese+Credit+Screening>. (4) The “Iran” data set used to support the findings of this study are included in [34, 45] (5) The “Bene 1” and “Bene 2” data set used to support the findings of this study are included within the following article: [32]. (6) The “Shuttle” data set used to support the findings of this study are included within the following URL: [http://archive.ics.uci.edu/ml/datasets/statlog+\(shuttle\)](http://archive.ics.uci.edu/ml/datasets/statlog+(shuttle)). (7) The “Skin_segment” data set used to support the findings of this study are included within the following URL: <http://archive.ics.uci.edu/ml/datasets/Skin+Segmentation>. (8) The “MiniBooNE” data set used to support the findings of this study are included within the following URL: <http://academictorrents.com/details/7fafb101f9c7961f9b840daeb4af43039107ddef>. (9) The “LC2017Q1” data set used to support the findings of this study are included within the following URL and article: [41] <http://www.lendingclub.com>.

Disclosure

Dayu Xu and Xuyao Zhang are co-first authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

Authors' Contributions

Dayu Xu and Xuyao Zhang contributed equally to this work.

Acknowledgments

The authors acknowledge the support for the project (no. 20YJC630173) supported by the Ministry of Education of Humanities and Social Science Project, the project (no. 31971493) by the National Natural Science Foundation of China, the National Key Research and Development Program of China (no. 2018YFD0401403), the Zhejiang Province Key Science and Technology Projects (no. 2018C02050), the Hangzhou Agricultural and Social Development Project (no. 20190101A07), and the Zhejiang Education and Teaching Reform Project (no. jg20180175) supported by the Department of Education of Zhejiang Province.

References

- [1] C. Serrano-Cinca and B. Gutiérrez-Nieto, “The use of profit scoring as an alternative to credit scoring systems in peer-to-peer (p2p) lending,” *Decision Support Systems*, vol. 89, pp. 113–122, 2016.
- [2] C. Liberati and F. Camillo, “Personal values and credit scoring: new insights in the financial prediction,” *Journal of the Operational Research Society*, vol. 69, no. 12, pp. 1–21, 2018.
- [3] J. Sun, H. Li, P.-C. Chang, and Q.-H. Huang, “Dynamic credit scoring using B & B with incremental-SVM-ensemble,” *Kybernetes*, vol. 44, no. 4, pp. 518–535, 2015.
- [4] A. Kammoun, “Credit scoring models for a Tunisian microfinance institution: comparison between artificial neural network and logistic regression,” *Review of Economics & Finance*, vol. 6, pp. 61–78, 2016.
- [5] Z. Zhao, S. Xu, B. H. Kang, M. M. J. Kabir, Y. Liu, and R. Wasinger, “Investigation and improvement of multi-layer perceptron neural networks for credit scoring,” *Expert Systems with Applications*, vol. 42, no. 7, pp. 3508–3516, 2015.
- [6] S. Y. Sohn, D. H. Kim, and J. H. Yoon, “Technology credit scoring model with fuzzy logistic regression,” *Applied Soft Computing*, vol. 43, pp. 150–158, 2016.
- [7] H. Xiao, Z. Xiao, and Y. Wang, “Ensemble classification based on supervised clustering for credit scoring,” *Applied Soft Computing*, vol. 43, pp. 73–86, 2016.
- [8] S. Oreski and G. Oreski, “Genetic algorithm-based heuristic for feature selection in credit risk assessment,” *Expert Systems with Applications*, vol. 41, no. 4, pp. 2052–2064, 2014.
- [9] D. Liang, C.-F. Tsai, and H.-T. Wu, “The effect of feature selection on financial distress prediction,” *Knowledge-Based Systems*, vol. 73, no. 1, pp. 289–297, 2015.
- [10] F. Feng, H. Fujita, M. I. Ali, R. R. Yager, and X. Liu, “Another view on generalized intuitionistic fuzzy soft sets and related multiattribute decision making methods,” *IEEE Transactions on Fuzzy Systems*, vol. 27, no. 3, pp. 474–488, 2019.
- [11] X. Peng and H. Garg, “Algorithms for interval-valued fuzzy soft sets in emergency decision making based on WDBA and CODAS with new information measure,” *Computers & Industrial Engineering*, vol. 119, pp. 439–452, 2018.
- [12] F. Feng, Z. Xu, H. Fujita, and M. Liang, “Enhancing promethee method with intuitionistic fuzzy soft sets,” *International Journal of Intelligent Systems*, vol. 35, no. 7, pp. 1071–1104, 2020.
- [13] R. Emekter, Y. Tu, B. Jirasakuldech, and M. Lu, “Evaluating credit risk and loan performance in online peer-to-peer (p2p) lending,” *Applied Economics*, vol. 47, no. 1, pp. 54–70, 2015.
- [14] Y. Guo, W. Zhou, C. Luo, C. Liu, and H. Xiong, “Instance-based credit risk assessment for investment decisions in p2p lending,” *European Journal of Operational Research*, vol. 249, no. 2, pp. 417–426, 2016.
- [15] S. Benítez-Peña, R. Blanquero, E. Carrizosa, and P. Ramírez-Cobo, “Cost-sensitive feature selection for support vector machines,” *Computers & Operations Research*, vol. 106, pp. 169–178, 2019.
- [16] S. Jadhav, H. He, and K. Jenkins, “Information gain directed genetic algorithm wrapper feature selection for credit rating,” *Applied Soft Computing*, vol. 69, pp. 541–553, 2018.
- [17] N. Kozodoi, S. Lessmann, K. Papakonstantinou, Y. Gatsoulis, and B. Baesens, “A multi-objective approach for profit-driven feature selection in credit scoring,” *Decision Support Systems*, vol. 120, pp. 106–117, 2019.
- [18] L. Cui, L. Bai, Z. Zhang, Y. Wang, and E. R. Hancock, “Identifying the most informative features using a structurally interacting elastic net,” *Neurocomputing*, vol. 336, pp. 13–26, 2018.
- [19] H. Zou and T. Hastie, “Regularization and variable selection via the elastic net,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 2, pp. 301–320, 2005.
- [20] H. Zou and H. H. Zhang, “On the adaptive elastic-net with a diverging number of parameters,” *The Annals of Statistics*, vol. 37, no. 4, pp. 1733–1751, 2009.

- [21] A. Bequé and S. Lessmann, "Extreme learning machines for credit scoring: an empirical evaluation," *Expert Systems with Applications*, vol. 86, pp. 42–53, 2017.
- [22] W.-C. Lin, C.-F. Tsai, Y.-H. Hu, and J.-S. Jhang, "Clustering-based undersampling in class-imbalanced data," *Information Sciences*, vol. 409–410, pp. 17–26, 2017.
- [23] W. W. Y. Jhang, J. Hu, D. S. Yeung, S. Yin, and F. Roli, "Diversified sensitivity-based undersampling for imbalance classification problems," *IEEE Transactions on Cybernetics*, vol. 45, no. 11, pp. 2402–2412, 2017.
- [24] G. Douzas and F. Bacao, "Self-Organizing Map Oversampling (SOMO) for imbalanced data set learning," *Expert Systems with Applications*, vol. 82, pp. 40–52, 2017.
- [25] G. Douzas, F. Bacao, and F. Last, "Improving imbalanced learning through a heuristic oversampling method based on K-means and SMOTE," *Information Sciences*, vol. 465, pp. 1–20, 2018.
- [26] S. Dahiya, S. S. Handa, and N. P. Singh, "A feature selection enabled hybrid-bagging algorithm for credit risk evaluation," *Expert Systems*, vol. 34, no. 6, pp. 1–11, 2017.
- [27] Y. Xia, C. Liu, and N. Liu, "Cost-sensitive boosted tree for loan evaluation in peer-to-peer lending," *Electronic Commerce Research and Applications*, vol. 24, pp. 30–49, 2017.
- [28] M. Doumpos and C. Zopounidis, "Model combination for credit risk assessment: a stacked generalization approach," *Annals of Operations Research*, vol. 151, no. 1, pp. 289–306, 2007.
- [29] D. Molodtsov, "Soft set theory—first results," *Computers & Mathematics with Applications*, vol. 37, no. 4–5, pp. 19–31, 1999.
- [30] P. K. Maji, R. Biswas, and A. R. Roy, "Fuzzy soft sets," *Journal of Fuzzy Mathematics*, vol. 9, no. 3, pp. 589–602, 2001.
- [31] P. K. Maji and S. K. Samanta, "Generalised fuzzy soft sets," *Computers & Mathematics with Applications*, vol. 59, no. 4, pp. 1425–1432, 2010.
- [32] B. Baesens, R. Setiono, C. Mues, and J. Vanthienen, "Using neural network rule extraction and decision tables for credit-risk evaluation," *Management Science*, vol. 49, no. 3, pp. 312–329, 2003.
- [33] K. Setiono, B. Mac Namee, and S. J. Delany, "Using semi-supervised classifiers for credit scoring," *Journal of the Operational Research Society*, vol. 64, no. 4, pp. 513–529, 2012.
- [34] A. I. Marqués, V. García, and J. S. Sánchez, "Exploring the behaviour of base classifiers in credit scoring ensembles," *Expert Systems with Applications*, vol. 39, no. 11, pp. 10244–10250, 2012.
- [35] H. Zhao and S. Yu, "Cost-sensitive feature selection via the $\ell_{2,1}$ -norm," *International Journal of Approximate Reasoning*, vol. 104, pp. 25–37, 2019.
- [36] D. Wang, Z. Zhang, R. Bai, and Y. Mao, "A hybrid system with filter approach and multiple population genetic algorithm for feature selection in credit scoring," *Journal of Computational and Applied Mathematics*, vol. 329, pp. 307–321, 2018.
- [37] D. Zhang and Z. Xia, "Weighted-averaging estimator for possible threshold in segmented linear regression model," *Journal of Statistical Planning and Inference*, vol. 200, pp. 102–118, 2019.
- [38] D. Tripathi, D. R. Edla, V. Kuppili, A. Bablani, and R. Dharavath, "Credit scoring model based on weighted voting and cluster based feature selection," *Procedia Computer Science*, vol. 132, pp. 22–31, 2018.
- [39] A. L. M. Vilela, C. Wang, K. P. Nelson, and H. E. Stanley, "Majority-vote model for financial markets," *Physica A: Statistical Mechanics and Its Applications*, vol. 515, pp. 762–770, 2019.
- [40] Y. Xia, C. Liu, B. Da, and F. Xie, "A novel heterogeneous ensemble credit scoring model based on bstacking approach," *Expert Systems with Applications*, vol. 93, pp. 182–199, 2018.
- [41] H. He, W. Zhang, and S. Zhang, "A novel ensemble method for credit scoring: adaption of different imbalance ratios," *Expert Systems with Applications*, vol. 98, pp. 105–117, 2018.
- [42] M. Papouškova and P. Hajek, "Two-stage consumer credit risk modelling using heterogeneous ensemble learning," *Decision Support Systems*, vol. 118, pp. 33–45, 2019.
- [43] Jayadeva, H. Pant, M. Sharma, and S. Soman, "Twin Neural Networks for the classification of large unbalanced datasets," *Neurocomputing*, vol. 343, pp. 34–49, 2019.
- [44] A. Mahani and A. R. Baba-Ali, "A new rule-based knowledge extraction approach for imbalanced datasets," *Knowledge and Information Systems*, pp. 1–27, 2019.
- [45] A. Marqués, V. García, and J. S. Sánchez, "Two-level classifier ensembles for credit risk assessment," *Expert Systems with Applications*, vol. 39, pp. 10916–10922, 2012.