

## Research Article

# “Dimension Reduction: Feature Subset” Method for Selecting the Best Index Combination in Reputation Evaluation of Crowdsourcing Participants

Yanrong Huang <sup>1</sup>, Zhan Zheng <sup>2</sup>, and Bo Wei <sup>3</sup>

<sup>1</sup>College of Economics & Management, Zhejiang University of Water Resource and Electric Power, Hangzhou 310018, China

<sup>2</sup>School of Media & Communication, Wuhan Textile University, Wuhan 430073, China

<sup>3</sup>School of Computer Science and Technology, Zhejiang Sci-Tech University, Hangzhou 310018, China

Correspondence should be addressed to Zhan Zheng; [czheng@wtu.edu.cn](mailto:czheng@wtu.edu.cn)

Received 6 July 2022; Revised 3 September 2022; Accepted 14 September 2022; Published 12 October 2022

Academic Editor: Xuefeng Zhang

Copyright © 2022 Yanrong Huang 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.

An effective reputation evaluation mechanism is an essential guarantee for the crowdsourcing mode's healthy, orderly, and rapid development. Aiming at the problems of unsound reputation evaluation mechanism, single reputation evaluation index, and poor discrimination ability of crowdsourcing platforms a “dimension reduction feature subset” method for selecting the best reputation evaluation index combination of crowdsourcing participants is proposed. This method first selects the best dimensionality reduction method by empirical method, then uses the classifier as the evaluation function of feature selection, and uses the sequential backward selection strategy (SBS) to select the feature subset and reputation evaluation algorithm with the best classification performance. The experimental results show that the reputation evaluation method of crowdsourcing participants based on ReliefF-SVM has the best performance in terms of accuracy, F1 measure, and stability and can select a comprehensive, objective, and effective evaluation index combination to distinguish the reputation status of crowdsourcing participants.

## 1. Introduction

Crowdsourcing is an online activity in which the task publisher convenes the crowdsourcing participants to complete the task through the crowdsourcing platform. It is an effective way to solve problems from a long distance [1]. Crowdsourcing participants provide knowledge, wisdom, experience, and skills through the crowdsourcing platform, participate in tasks, and receive remuneration. The task publisher publicly convenes the participants of crowdsourcing through the crowdsourcing platform to participate in the completion of the task and pay compensation. Crowdsourcing platform is the intermediary and bridge between task publishers and task participants [2]. Crowdsourcing is an open innovation form that can pool talents from all fields to participate in technological innovation and value creation, stimulate the innovation vitality of skills,

provide valuable achievements [3, 4], and help enterprises solve the problems they face.

Hundreds of crowdsourcing platforms such as Zhu Bajie, Yipin Weike, Upwork, and Freelancer have emerged at home and abroad. Most domestic and foreign crowdsourcing platforms use transaction amounts and positive feedback to evaluate the reputation of crowdsourcing participants [1, 2]. The reputation evaluation index is single, and the reputation evaluation mechanism is not sound, so it is not easy to give comprehensive feedback on the reputation of crowdsourcing participants.

As the reputation evaluation mechanism of the crowdsourcing platform is not sound, the information on both sides of the transaction is asymmetric, and traders' speculative psychology [5] leads to frequent violations of crowdsourcing participants. Crowdsourcing participants ask for added remuneration, submit plagiarized results, provide the same

result to participate in multiple tasks, are unable to complete tasks, fail to complete tasks as required, blackmail task publishers for evaluation, fail to provide follow-up maintenance services, maliciously lower prices to rob customers, conduct false transactions, and guide offline transactions and other violations. Some crowdsourcing platforms in China, such as Zhu Bajie, have taken ex-postpunishment measures such as restricting trading and closing accounts for illegal traders, but the effect is insignificant.

When the reputation evaluation mechanism of the crowdsourcing platform is challenging to regulate and restrict the behavior of participants effectively, a large number of crowdsourcing participants with a poor reputation will disrupt the trading order at low prices, and the crowdsourcing participants with a good reputation will leave the market one after another, resulting in the “Lemon Effect,” causing the collapse of the crowdsourcing market. Whether we can select the key indicators for the reputation evaluation of crowdsourcing participants and establish an effective reputation evaluation mechanism is directly related to the healthy, orderly, and rapid development of crowdsourcing activities [6, 7].

However, the academic research on the reputation of e-commerce platform participants primarily concentrates on physical commodity trading and financial services, which provides a reference for constructing a reputation evaluation mechanism for a crowdsourcing platform. The research on reputation evaluation of crowdsourcing platforms mainly focuses on reputation evaluation methods, ignoring the discussion of evaluation indicators. The premise of practical evaluation is whether to select evaluation indicators that can significantly distinguish the reputation of crowdsourcing participants. Therefore, the vital issue is removing the indicators that have little impact on the reputation of crowdsourcing participants from the numerous evaluation indicators and selecting the best combination that can significantly distinguish the reputation of crowdsourcing participants.

The selection of the best indicator combination for the reputation evaluation of crowdsourcing participants is mainly a research method problem. How to rely on the Internet’s open and real-time participation environment, make use of the large amount of behavior data generated by the online transactions of crowdsourcing participants, and use machine learning technology to select the best indicator combination for the reputation evaluation of crowdsourcing participants? Based on research paradigm III of “computational, experimental paradigm of selection behavior research” in [7], that is, following the idea of “raising questions, data collection, data analysis, and theoretical conclusion,” two major problems need to be solved in this study. First, find the data dimensionality reduction method and find the indicators related to the reputation evaluation of crowdsourcing participants from a large number of index data. Second, find the method to obtain the best feature subset, establish the correlation between the optimal index combination and the machine learning algorithm, and find the best index combination that can significantly distinguish the reputation of crowdsourcing participants.

## 2. Literature Review

In the context of the Internet, the reputation evaluation of network platform participants has attracted extensive attention from the academic community and has become one of the frontier issues in e-commerce. Scholars have made rich achievements in the research of e-commerce platforms, which provides an essential theoretical basis for the reputation evaluation of crowdsourcing participants.

*2.1. Research on Selection Method of Reputation Evaluation Index.* The selection method of the reputation evaluation index is divided into single index selection and index combination selection.

The single index selection method selects a better group of indexes to build an index system according to the discrimination ability of a single index. Jiang et al. [8] analyzed that the number of frauds has a great impact on the reputation of platform traders. Yan et al. [9] proposed to measure the credibility of participants through active factors and historical factors. Liu et al. [10] concluded through literature analysis that professional level, work speed, work attitude, smooth communication, after-sales service, and innovation are the key factors to evaluating reputation. Zhang et al. [11] proposed to take interest and ability as the reputation evaluation index of crowdsourcing participants. Wang et al. [12] proposed that service quality and user score are important indicators affecting the reputation of crowdsourcing participants.

Although the index combination selected by the single index selection method can improve the discrimination ability of the model, it cannot guarantee that the selected index combination has the most robust discrimination ability. To make up for this shortcoming, scholars put forward the selection method of the index group. Jiang et al. [13] used the SBS method to screen P2P platform users’ reputation evaluation indicators. Zhang et al. [14] used, CHAID, C5.0, and CART, three decision tree models to screen farmers’ reputation evaluation indicators. Li et al. [15] extracted key indicators affecting personal credit using the Sparse Bayesian model. Wei et al. [16] proposed an optimal feature subset classification algorithm for the self-tuning particle swarm optimization algorithm. Zhang et al. [17] used the SVM model optimized based on the firefly algorithm to study the financial evaluation indexes of the supply chain of small- and medium-sized enterprises. Zhao et al. [18] mined key default characteristics of farmers based on the least significant difference method. Zhang et al. [19] used elastic network regression to select a credit characteristic index and determine listed enterprises’ credit evaluation index systems in China’s A-share market.

At present, the research on the selection of index combinations mainly focuses on the financial field, and there is less research on service e-commerce. The research results of the reputation evaluation index are difficult to be transplanted to the evaluation of the reputation of crowdsourcing participants.

*2.2. Research on Reputation Evaluation Method.* In recent years, the reputation evaluation method has attracted the attention of academia and has achieved a series of research results. At present, scholars' research mainly focuses on game theory and mathematical models. The research on reputation evaluation using the game theory method includes the following. Zhan et al. [20] established a multi-objective Stackelberg game model under incomplete information, trying to solve the problem of maintaining the reputation of the cross-border e-commerce platform when the reputation maintainer faces one or more damages. Quan et al. [21] discussed a reputation evaluation mechanism considering tolerance based on game theory. Wang et al. [22] established a reputation update method for free-riding and false data, which improved the credibility of the system. Lu et al. [23] designed a new optimal rating protocol based on game theory. Zhu et al. [24] designed an incentive mechanism based on game theory to make candidate nodes more willing to give honest reputation verification results.

The research of using a mathematical model to evaluate reputation includes the following. Yan et al. [9] improved the mean model, considered the activity factor and historical factor in the model, and proposed the worker reputation model based on activity. Bhattacharjee et al. [25] put forward the QNQ reputation model, using a regression method to take quantity and quality as judgment indicators to distinguish honest, selfish, and fraudulent users. Lu et al. [26] incorporated the comment text into the reputation evaluation and constructed the reputation calculation model. Sun et al. [27] studied the vehicle crowdsourcing service that provided users with real-time traffic feedback information and calculated participants' reputation values through trust propagation and feedback similarity. Huang et al. [1] proposed a multidimensional weighted cumulative reputation calculation model, MWCRM.

Scholars have begun to explore the use of machine learning technology for reputation evaluation in recent years. For example, Al Quadri et al. [28] used machine learning methods to predict the reliability of consumers from their data. Rantanen et al. [29] constructed an index system from the dimensions of quality, reliability, responsibility, success, pleasure, and innovation. They used a convolutional neural network (CNN) to classify and evaluate online corporate reputation. Yang et al. [30] used a semi-SVM model to measure the reputation of online service providers. Wang et al. [12] proposed an HMRep reputation evaluation method that could effectively resist malicious evaluation and improve calculation accuracy. In addition, in recent years, scholars have studied issues related to reputation management. Yu et al. [31] proposed an adaptive fog blockchain reputation storage method to improve the security and effectiveness of reputation management systems. Kealeboga [32] used the life cycle theory combined with the geographical location information of participants to judge the quality, reliability, and credibility of the dataset of reputation contribution data of crowdsourcing participants.

In conclusion, the research on crowdsourcing reputation evaluation mechanisms has made great progress. However, the method of game theory shows a strong explanatory

power of social phenomena through deductive reasoning. However, the assumption of the ideal subject of game theory often deviates from the actual phenomena. Mathematical model methods need to make assumptions or judgments first, and the prediction results will be expressed as one or a group of functional relations, but they fail to combine the participant behavior generated by the crowdsourcing platform with massive amounts of data for reputation evaluation.

In the context of big data, how to use the massive data generated by the crowdsourcing platform and consider the relationship between the selection of reputation evaluation indicators and evaluation methods to evaluate the reputation of crowdsourcing participants remains to be discussed. How to objectively and accurately identify the key index combination of reputation evaluation from many indicators and put forward the best evaluation index combination that is objective and can significantly distinguish crowdsourced reputation is a question

Based on the analysis of literature, to start from the reputation characteristics of crowdsourcing participants, firstly, by systematically combining the relevant literature and learning from the reputation evaluation indicators of a crowdsourcing platform, we preliminarily select the reputation evaluation indicators of outstanding crowdsourcing participants. Secondly, using empirical research methods, select the best data dimensionality reduction method. Thirdly, the combination of the search strategy and machine learning algorithm is used to establish the association between the optimal index combination and the optimal machine learning algorithm, mine the combination effect of different data dimensionality reduction methods and machine learning algorithms, and put forward the best index combination selection method.

### 3. Research Procedures

Based on the computational experimental paradigm of choice behavior research, we put forward theoretical conclusions based on experiments, which mainly include three steps.

First is data acquisition and preprocessing. Considering the characteristics of reputation, following the selection principle of evaluation indicators, analyzing the reputation evaluation indicators of China's mainstream crowdsourcing platforms, systematically combining the relevant literature, and preliminarily screening 28 reputation evaluation indicators of crowdsourcing participants [33], we collect reputation evaluation index data from the Zhu Bajie crowdsourcing platform and preprocess the data.

Secondly, the best data dimensionality reduction method is selected. To find the most suitable method for dimensionality reduction of reputation evaluation indicators of crowdsourcing participants, four methods, ReliefF, mean impact value (MIV), principal component analysis (PCA), and linear discriminant analysis (LDA), are selected for comparative analysis. Select PCA method and ReliefF to calculate the index weight, sort according to the weight of the evaluation index, judge the data noise, and delete the

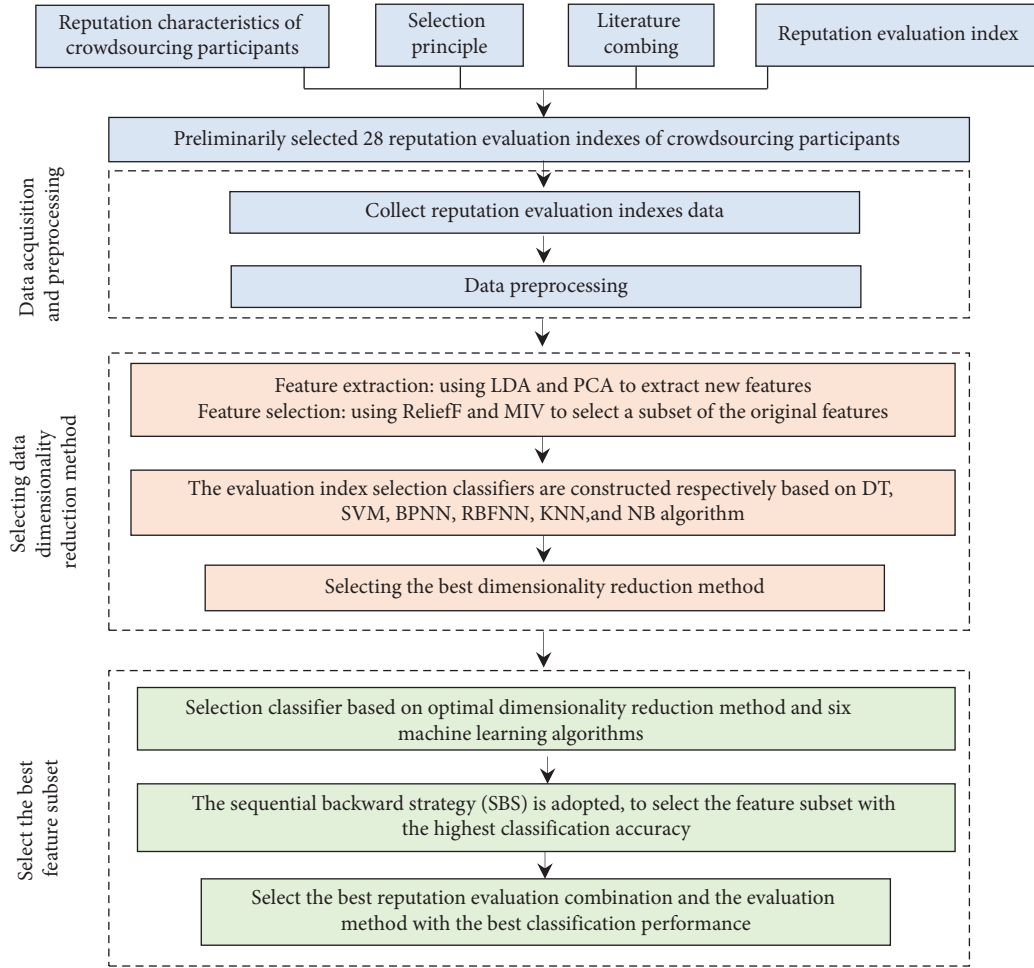


FIGURE 1: Selection steps of the best evaluation index combination of the reputation of crowdsourcing participants.

redundant index with weak reputation discrimination ability and small impact on crowdsourcing participants.

Data mining is the combined effect of a data dimensionality reduction method and a machine learning algorithm. Six commonly used machine learning classification algorithms are selected to cross construct 24 classifiers with four data dimensionality reduction methods. Taking the reputation evaluation dataset of crowdsourcing participants on the Zhu Bajie platform as an example, classification and prediction are carried out. Six machine learning algorithms include Decision Tree (DT), BP Neural Network (BPNN), RBF Neural Network (RBFNN), Support Vector Machine (SVM), K-Nearest Neighbor Classifier (KNN), and Naive Bayes (NB). Taking Zhu Bajie dataset as an example, the data dimensionality reduction method with the best overall classification effect of the six algorithms is selected.

Thirdly, select the best combination of indicators. Based on the selected best data dimensionality reduction method and six machine learning algorithms, a reputation evaluation index selection classifier for crowdsourcing participants is constructed. The sequential backward selection strategy (SBS) is adopted, and the classifier is used as the evaluation function to delete the indicators that have the least impact on the reputation discrimination ability of crowdsourcing

participants one by one until the optimal number of indicators is reached, to determine the feature subset with the highest classification accuracy.

The confusion matrix is used to compare the accuracy, recall, and F1 value of the classifier. Using statistical significance test methods, including Friedman test, Kruskal-Wallis test, and dispersion degree, it compares the classification accuracy and stability of the classifier constructed by different data dimensionality reduction methods and machine learning algorithms, selects the classifier with the best performance, and puts forward the best index combination; the specific steps are shown in Figure 1.

#### 4. Data Acquisition and Preprocessing

We select the Zhu Bajie platform to collect data, mainly based on three reasons: first, in China, the market share of the Zhu Bajie platform is the highest, reaching 50%, and the transaction is the most active, which has practical research significance. Secondly, the Zhu Bajie platform is one of the most widely studied crowdsourcing platforms and has been selected by many scholars as an important platform for data collection and research [4, 9, 26, 27]. Third, the Zhu Bajie platform provides the transaction data of crowdsourcing

participants, the published list of illegal crowdsourcing participants, violation records, and reputation-related data such as work attitude, work speed, and completion quality, which is suitable for this study.

Based on the reputation evaluation index system of crowdsourcing participants established in [2], the evaluation indexes are collected. Collect 28 reputation evaluation index variables of crowdsourcing participants on the Zhu Bajie platform from December 1, 2006, to January 31, 2019, including the city of crowdsourcing participants, store type, margin deposit amount, years of opening, transaction amount in recent three months, transaction times of three months, number of employers served, transaction activity, number of refunds in this month, refund rate in this month, number of refunds in three months, refund rate in three months, comprehensive scoring, task completion quality, work speed, work attitude, employer repurchase rate, positive feedback rate, the number of positive feedback, the number of medium feedback, the number of negative feedback, number of employer recommendation, the number of employers did not choose, the number of employer nonrecommend, growth scoring, number of punishments, number of punishments in three months, and credibility frozen after reporting.

Through data collection, a total of 4357 samples of crowdsourcing participants were obtained. Each sample of data contains 28 index variables. After data preprocessing, 3298 valid samples are finally obtained. Through manual marking, the reputation level of crowdsourcing participants is marked as good, medium, and poor.

## 5. Selection of Data Dimensionality Reduction Methods

The massive data generated by the crowdsourcing platform provides information for the reputation evaluation of crowdsourcing participants, but the high dimension of variables will increase the operation cost. It is necessary to eliminate redundant indicators unrelated to the reputation evaluation and having no significant impact and determine the best data dimensionality reduction method. It selects four data dimensionality reduction methods and six machine learning algorithms for empirical research. Take the participants of the Zhu Bajie platform as the research object, collect the reputation evaluation index data, and propose the evaluation index selection method based on the experiment.

**5.1. Existing Data Dimensionality Reduction Methods.** Data dimensionality reduction methods mainly include feature selection and feature extraction. The goal of both is to reduce the dimension of index variables, but there are differences in the ways.

Feature extraction is to get new features by mixing the original features. For example, the Mean Impact Value (MIV) method reflects the influence of output variables on input variables through the change of the weight matrix in the neural network [33]. For example, the ReliefF algorithm makes a feature extraction method, which gives different

weights to different index variables and removes the index variables with insignificant evaluation results. The ReliefF method has the characteristics of a simple algorithm and high operational efficiency.

The specific calculation process of the ReliefF algorithm is as follows: randomly extract the sample subset  $p$  from the dataset, then extract  $s$  nearest neighbor samples from  $p$  similar sample set and a different sample set, respectively, calculate the feature weight  $w$ , and update the correlation between features and categories in turn. Sort the features according to the weight, set the threshold, and select the feature subset; then, there are the following.

Input:  $p$  samples and corresponding characteristic attributes.

Output: feature weight vector  $w$ .

Initialize  $w = 0$ .

Weight value  $w$  calculation: the formula is

$$w[I] = w[I] - \frac{\sum_{j=1}^s \text{diff}(I, P_i, H_j)}{ms} + \sum_{c \neq \text{class}(P_i)} \left[ \frac{p(C)}{1 - p(\text{class}(P_i))} \sum_{j=1}^s \text{diff} \left( \frac{I, P_i, M_j(C)}{MS} \right) \right], \quad (1)$$

where  $m$  is the number of samples,  $\sum_{j=1}^s \text{diff}(I, p_i, H_j)$  calculate the distance between two sample examples about feature  $I$ ,  $M_j(C)$  is the  $j$ th nearest neighbor sample of different classes, and  $\text{class}(P_i)$  is the category of the sample  $P_i$ . The ReliefF algorithm has the advantages of easy expansion, high computational efficiency, and strong stability. It can process noisy data and quickly process large datasets.

### 5.2. Common Machine Learning Algorithms

**5.2.1. Support Vector Machine.** Support vector machine (SVM) was first proposed by Cortes and Vapnik in 1995. SVM constructs the maximum interval classifier by dividing multiple features and draws lessons from the concepts of maximum interval hyperplane [34], the inner product of kernel function as feature space, use of kernel function, and sparsity. Vapnik Chervonenkis (VC) theory proves the existence of a risk VC boundary.

It is a method to classify linear and nonlinear data. For the linear separable problem, suppose the training set is  $T = \{(x_1, y_1), \dots, (x_n, y_n)\}$ , among  $x_i \in R^p$ ,  $y_i \in \{-1, +1\}$ ,  $i = 1, 2, \dots, n$ . Since the problem is linearly separable, there is

$$\{x: f(x_i) = w^T x + w_0 = 0\}. \quad (2)$$

The positive and negative sample points in the dataset are completely and correctly divided on both sides of the hyperplane. That is, for all sample points of  $y_i = +1$ , we have  $f(x_i) > 0$ , and for all samples of  $y_i = -1$ , we have  $f(x_i) < 0$ . The decision function can be constructed:

$$G(x) = \text{sgn}(f(x)) = \text{sgn}(w^T x + w_0). \quad (3)$$

For any hyperplane  $L = \{x: f(x) = w^T x + w_0\}$ , let  $x_1$  and  $x_2$  be two points on the hyperplane; then,  $w^T(x_1 - x_2) = 0$ . So,  $w^* = w/\|w\|$  is the unit vector of the hyperplane. Let  $x_0$  be a point on the hyperplane; then,  $w^T x_0 + w_0 = 0$ . Therefore, the distance from any point  $x$  to the hyperplane is

$$\begin{aligned} S(L, x) &= |w^{*T}(x, x_0)| = \frac{1}{\|w\|} |w^T(x - x_0)| \\ &= \frac{1}{\|w\|} |w^T x + w_0| = \frac{f(x)}{\|w\|}. \end{aligned} \quad (4)$$

Divide all sample points in the hyperplane; that is, all the sample points of  $y_i = +1$  have  $f(x_i) > 0$ ; for all the samples with  $y_i = -1$ , we have  $f(x_i) < 0$ . Therefore,

$$S(L, x) = \frac{1}{\|w\|} (y_i f(x_i)) = \frac{1}{\|w\|} y_i (w^T x_i + w_0). \quad (5)$$

SVM has advantages in modeling complex nonlinear boundaries. It can control the model by maximizing the edge of the decision boundary. It has good stability, generalization characteristics, and a unique global optimal solution.

**5.2.2. BP Neural Network.** Backpropagation neural network (BPNN) is a multilayer feedforward neural network [35]. The input tuple of BPNN is weighted by the input layer and then given to the middle layer [36]. The output of the middle layer unit can be input to another middle layer. The number of middle layers is arbitrary, and usually, only one layer is used in practice. The weighted output of the last middle layer is used as the input of the unit constituting the output layer, and the output layer publishes the network prediction of a given tuple.

**5.2.3. Radial Basis Function Neural Network.** The radial basis function neural network (RBFNN) is a neural network structure proposed by Broomhead and Lowe in 1988. It is a three-layer feedforward network with a single hidden layer. The RBF neural network reduces the weight update link of error feedback, applies radial basis function as an excitation function in the hidden layer to fit the nonlinearity of the data set, and has the characteristics of simple training and fast learning convergence.

**5.2.4. Decision Tree.** The decision tree (DT) method came into being in the 1960s. It is a learning system built by Quinlan et al. when modeling human concepts. It is a method of decision-making based on the tree structure. The decision tree starts from the root node and tests a certain feature of instance  $x$ ; according to the test results, the instance is assigned to the child node until it reaches the leaf node, and the class to which the leaf node belongs, that is, the label  $y$  of instance  $x$ , is predicted.

**5.2.5. K-Nearest Neighbor Classifier.** The K-nearest neighbor classifier (KNN) is a simple machine learning algorithm.

This method assumes that adjacent points have the same attributes and “proximity” is generally measured by distance, such as Euclidean distance.

**5.2.6. Naive Bayesian.** Naive Bayesian (NB) is a statistical classification method. This method creates the probability distribution of features on each class label, which is characterized by combining a priori probability and a posteriori probability; that is, it avoids the subjective bias of using only a priori probability and the overfitting phenomenon of using sample information alone.

**5.3. Preliminary Selection of Evaluation Indicators.** This study uses empirical methods. Firstly, the filtering method is used to preliminarily select the 28 index variables in the dataset in the ZhuBajie platform and preliminarily eliminate the variables with irrelevant redundancy, weak discrimination ability, and insignificant impact on the reputation evaluation, to reduce the dimension of the variables.

Two feature selection methods, ReliefF and Mean Influence Value (MIV), and two feature extraction methods, principal component analysis (PCA) and linear discriminant analysis (LDA), were selected to compare the dimensionality reduction effect. Among the two-dimensionality reduction methods, ReliefF and PCA are selected to focus on the selection results of indicators in the first stage.

**5.3.1. Selection Results of Phase I Indicators of ReliefF.** The sample subset is selected randomly from the original data sample set, and then, the nearest neighbor samples are selected from a similar sample set of the sample subset. Each feature weight value is calculated and updated in turn. Repeat the above process many times to obtain the weight of features, arrange the features in descending order according to their feature weight value, and select part of the feature set by giving a threshold. That is, when the feature weight value is greater than the given threshold, the feature is used to form a new feature subset. If it is less than the given threshold, the feature is removed. Relevant calculations are carried out on MATLAB2016b, and the sorting results of feature weights are shown in Figure 2 and Table 1.

As shown in Table 1, the most significant indicator affecting the reputation evaluation of crowdsourcing participants is the praise rate, followed by the number of three-month penalties and the three-month refund rate, with weights of 0.054, 0.037, and 0.029, respectively. The evaluation index with negative index feature weight means that the information reflected by this feature increases the data noise and affects the accuracy of classification evaluation, which should be eliminated. Therefore, seven evaluation indexes, including store type, number of transactions in three months, employer served, three-month transaction volume, the employer not selected, total good evaluation, and employer recommendation, should be deleted. Eliminate redundant indicators with weak discrimination ability and little impact on reputation evaluation. The weight of the

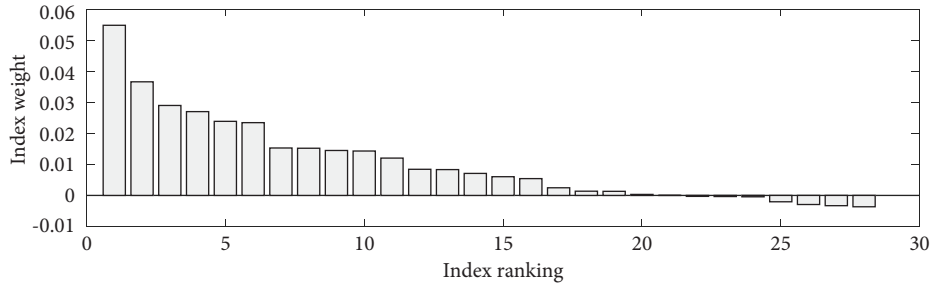


FIGURE 2: Index variable weight table after dimension reduction using ReliefF.

TABLE 1: Weight of index variables after dimension reduction using ReliefF.

Rank	Index variable	Weight
1	$x_1$ positive feedback rate	0.054
2	$x_2$ number of punishments in three months	0.037
3	$x_3$ refund rate in three months	0.029
4	$x_4$ refund rate in this month	0.027
5	$x_5$ credibility frozen after reporting	0.023
6	$x_6$ transaction activity	0.023
7	$x_7$ work attitude	0.015
8	$x_8$ task completion quality	0.015
9	$x_9$ work speed	0.014
10	$x_{10}$ years of opening	0.014
11	$x_{11}$ comprehensive scoring	0.012
12	$x_{12}$ number of employer nonrecommend	0.008
13	$x_{13}$ margin deposit amount	0.008
14	$x_{14}$ number of negative feedback	0.007
15	$x_{15}$ growth scoring	0.006
16	$x_{16}$ number of medium feedback	0.0055
17	$x_{17}$ number of punishments	0.037
18	$x_{18}$ employer repurchase rate	0.001
19	$x_{19}$ number of refunds in this month	0.001
20	$x_{20}$ number of refunds in three months	0.0003
21	$x_{21}$ city	0.00001
22	$x_{22}$ number of employer recommendation	-0.002
23	$x_{23}$ number of positive feedback	-0.0029
24	$x_{24}$ number of employers did not choose	-0.003
25	$x_{25}$ transaction amount in recent three months	-0.0003
26	$x_{26}$ number of employers served	-0.0037
27	$x_{27}$ transaction times of three months	-0.0004
28	$x_{28}$ store type	-0.0004

city is 0.00001, which shall be eliminated. According to the first-stage index selection results of ReliefF, 20 of the 28 indexes, are retained.

5.3.2. *Selection Results of Indicators in the First Stage of the PCA.* The principal component analysis is a commonly used feature extraction method. It extracts new feature variables from the original feature variables, condenses the original many variables into a few variables, and retains the original information as much as possible while reducing the number of feature variables. The principal component analysis first determines whether the characteristic variables are suitable for PCA analysis by the Bartlett sphericity test and KMO test. Using SPSS25.0 software for the Bartlett sphericity test, the observed value of the statistics is 67924.04, and the

probability  $P$  value is close to 0. The KMO value is 0.77, indicating that the original variable is suitable for principal component analysis. The principal component is extracted by the PCA method, and the total variance of the original variables explained by the principal component is shown in Table 1.

In Table 2, the first column is the principal components extracted by the PCA method, which are numbered according to the variance contribution rate. The eigenvalue of the first principal component is 5.484, which explains 19.587% variables. The eigenvalues of the first 10 principal components are greater than 1, which explains 74.738% variables, indicating that the first 10 factors retain most of the original information.

To sum up, the feature selection method ReliefF is used to rank the importance of the evaluation index system, then judge the data noise, and eliminate the redundant indexes with weak reputation discrimination ability and little impact on the crowdsourcing participants. According to the ranking results of the evaluation indicators, 8 indicators are removed from the original 28 indicator variables and 20 indicators are retained. Using the feature extraction method to calculate the total variance of principal components, the eigenvalues of the first 10 of the 28 indicators are greater than 1, indicating that the original feature information is less lost when 10 principal components are extracted.

5.4. *Select the Best Data Dimensionality Reduction Method.*

After the preliminary screening of evaluation indicators through empirical methods, they use the combination method of data dimensionality reduction and machine learning algorithm and then empirically finds the best data dimensionality reduction method. Four data dimensionality reduction methods, including ReliefF, MIV, PCA, and LDA, are selected, together with DT, BPNN, RBFNN, SVM, KNN, and NB six machine learning algorithms cross construct 24 kinds of crowdsourcing participants' reputation evaluation index selection classifier, mine the combined effect of data dimensionality reduction method and machine learning algorithm, and select the best data dimensionality reduction method.

Taking the dataset of the Zhu Bajie platform as an example, the performance of 24 crowdsourcing participant evaluation indexes' selection classifiers constructed cross is verified. Through ten-fold cross-validation and the



TABLE 2: Total variance of original variables explained by principal components.

Principal component	Initial eigenvalue			Extract the sum of squares of loads		
	Total	Variance contribution rate (%)	Cumulative percentage (%)	Feature value	Variance contribution rate (%)	Cumulative percentage (%)
1	5.484	19.587	19.587	5.484	19.587	19.587
2	4.699	16.781	36.369	4.699	16.781	36.369
3	1.983	7.084	43.452	1.983	7.084	43.452
4	1.636	5.842	49.295	1.636	5.842	49.295
5	1.577	5.634	54.929	1.577	5.634	54.929
6	1.243	4.439	59.367	1.243	4.439	59.367
7	1.218	4.352	63.719	1.218	4.352	63.719
8	1.046	3.737	67.456	1.046	3.737	67.456
9	1.031	3.683	71.139	1.031	3.683	71.139
<b>10</b>	<b>1.008</b>	<b>3.600</b>	<b>74.738</b>	<b>1.008</b>	<b>3.600</b>	<b>74.738</b>
11	0.967	3.454	78.192			
12	0.946	3.377	81.569			
13	0.858	3.063	84.632			
14	0.738	2.635	87.268			
15	0.680	2.428	89.696			
16	0.528	1.887	91.583			
17	0.485	1.733	93.316			
18	0.357	1.276	94.592			
19	0.320	1.143	95.735			
20	0.303	1.083	96.818			

Note. List the principal components with weight ranking 1–20. The eigenvalues of the 10th principal components are greater than 1 in bold.

Friedman test, the performances of different classifiers are compared, and the best data dimensionality reduction method is selected.

The dataset is divided into the training set and test set, and the data proportions are 70% and 30%. We select the classification regression algorithm CART to classify the whole sample and set the input layer of BPNN as 20 input variables, 20 hidden layer neurons, and 3 output neurons; the learning rate value is 0.1, the training required accuracy value is 0.00001, and the maximum training time is 100. We use Newrbe function to create RBFNN and Gaussian function as kernel function, by observing the change of accuracy of KNN classifier, to set the nearest neighbor value to 8.

Based on the ranking results of the reputation evaluation index weights of crowdsourcing participants in the first stage, to avoid losing important evaluation indexes, ReliefF is used to filter the indexes with data noise and weak discrimination ability, select the first 20 evaluation indexes, and calculate the average ten-fold cross-verification accuracy of the 24 selection classifiers constructed, as shown in Table 3.

It is shown that the average classification accuracy of ReliefF-BPNN ten-fold cross-validation is the highest, which is 0.91, followed by LDA-SVM, which is 0.908, and PCA-DT, which is the lowest, which is 0.817. After using the ReliefF method for data dimensionality reduction, the evaluation index selection classifier constructed based on DT, BPNN, RBFNN, and NB algorithms has the highest accuracy. After using the LDA method for data dimensionality reduction, the evaluation index selection classifier constructed based on the SVM and KNN algorithms has the highest accuracy.

TABLE 3: Accuracy of ten-fold cross-validation of the reputation evaluation index selection classifier.

Model	DT	SVM	BPNN	RBFNN	NB	KNN	Rank
ReliefF	<b>0.899</b>	0.901	<b>0.910</b>	<b>0.856</b>	<b>0.878</b>	0.883	<b>3.67</b>
MIV	0.875	0.887	0.886	0.828	0.819	0.872	1.67
PCA	0.817	0.844	0.855	0.849	0.846	0.854	1.5
LDA	0.881	<b>0.908</b>	0.905	0.848	0.871	<b>0.893</b>	3.17

The best result for each classifier is in bold.

The effects of different dimensionality reduction methods are verified by the Friedman test. The average rank results of the four dimensionality reduction methods ReliefF, MIV, PCA, and LDA are shown in Table 3. The observed value of the Friedman test statistic is 12.2, and the asymptotic significance is 0.007, which indicates that the four feature dimensionality reduction methods have significant differences.

The average rank of the four dimensionality reduction methods ReliefF, MIV, PCA, and LDA are 3.67, 1.67, 1.5, and 3.17, respectively. The ReliefF method has the best dimensionality reduction effect, and the LDA method can also achieve a good classification effect. ReliefF evaluates the correlation and redundancy of features by calculating the correlation statistics of the corresponding features of adjacent samples between the same class and different classes, which is suitable for multiclassification problems. The LDA method removes irrelevant information in the data by projecting the dataset to a lower dimension to achieve the effect of dimension reduction.

In this case, the overall dimensionality reduction effect of the classifier is the best when using the ReliefF method. The



classification performance of MIV and PCA is significantly lower than that of ReliefF and LDA, and PCA is the worst. When using the PCA dimensionality reduction method, the evaluation index selection classifier based on DT, SVM, BPNN, and KNN shows the worst performance.

In terms of training time, the DT classifier has the fastest average training speed, followed by the NB classifier. The average training time of the RBFNN classifier is 5.13 seconds, which is the longest among the six algorithms selected. Secondly, the average training time of the BPNN classifier is 3.72, which shows that the operation cost of this method is high. Among the four dimensionality reduction methods, the average training time of LDA and ReliefF is 1.85 seconds and 2.03 seconds, respectively, and the average training time of MIV is 4.47, which takes the longest time. The average training time of PAC is 1.26 seconds, which is the shortest.

After using the ReliefF method to reduce the dimension of data, the accuracy of classifier selection based on the evaluation indexes of DT, SVM, BPNN, RBFNN, and NB algorithms is concentrated in the range of 0.86–0.93 and fluctuates gently. The highest single classification accuracy of ReliefF-BPNN is 0.922. The classification accuracy of relief RBFNN fluctuates in the range of 0.83–0.88, and the performance of the classifier is significantly lower than that of the other five classifiers. The ten-fold cross-validation results of the evaluation index selection classifier of crowdsourcing participants are shown in Figure 3.

After data dimensionality reduction using the MIV method, the accuracy of classifier selection based on evaluation indexes of DT, SVM, BPNN, and KNN algorithms is concentrated in the range of 0.86–0.92, with gentle fluctuation and high classification stability. MIV-NB has the lowest single classification accuracy of 0.775 and the highest of 0.895, with a difference of 0.12. Among all the classifiers, the classification stability is the worst.

After using the LDA method to reduce the dimension of data, the accuracy of classifier selection based on SVM and BPNN algorithms is concentrated in the range of 0.89–0.92, and the accuracy of classifier selection based on the Nb algorithm fluctuates between 0.83 – 0.87.

After data dimensionality reduction using the PCA method, the accuracy of classifier selection based on the evaluation indexes of SVM, BPNN, RBFNN, and KNN algorithms is concentrated in the range of 0.83–0.88, and the overall classification accuracy is lower than that of other dimensionality reduction methods. The fluctuation range of PCA-NB accuracy is the largest, and the maximum amplitude is 0.065, indicating that the performance of the classifier is the most unstable. The accuracy of PCA-DT fluctuates between 0.8 – 0.85, and the classification accuracy is the lowest among all methods.

Through the analysis and comparison of four data dimensionality reduction methods and six machine learning algorithms, the performance of 24 crowdsourcing participants' reputation evaluation index selection classifiers was cross constructed; the research shows that the classification effect of using the ReliefF method to select evaluation indexes is the best, which is better than the LDA, PAC, and MIV methods. Among them, the dimensionality reduction

effect of PCA and MIV methods is significantly lower than that of ReliefF and LDA dimensionality reduction methods. In terms of training time, the PCA method has the fastest training speed and the MIV dimensionality reduction method has the slowest training speed. The training speed of the LDA method is slightly higher than that of the ReliefF dimension reduction method. Compared with the LDA dimensionality reduction method, ReliefF adopts the feature selection method, which does not change the original index and is more explanatory. Therefore, we select the best data dimensionality reduction method, ReliefF.

## 6. Optimal Feature Subset Selection

Based on the empirical selection of the best data dimensionality reduction method, ReliefF, a classifier based on six machine learning algorithms, DT, BPNN, SVM, KNN, and NB, is constructed. The prediction performance of the classifier is used as the standard to evaluate the combination of indicators. The Sequential Backward Selection (SBS) strategy is used to adjust the selected evaluation indexes, and the prediction performance of the classifier is used as the standard to evaluate the combination of evaluation indexes. Through the empirical analysis of data, the selection method of the best feature subset is proposed to select the best feature subset.

*6.1. Selection Method of the Best Evaluation Index Combination Based on ReliefF.* The selection method of the best evaluation index combination based on ReliefF is shown in Figure 4.

Suppose there are  $n$  samples in dataset  $D$ , and  $m$  features are preliminarily selected after deleting redundant variables in the first stage.  $M_i$  represents the feature subset with  $i$  features in the dataset, where  $1 < i \leq m$ ;  $A_i$  represents the accuracy of classification prediction by the classifier when the dataset has  $i$  features, and  $A_{\max}$  represents the highest classification accuracy of the classifier. 70% of the samples in dataset  $D$  are divided into training sets to train classifiers; 30% are divided into test sets, which calculate the accuracy  $A_i$  of the test set when the number of features is  $i$ .

The sequential backward selection strategy (SBS) is adopted to delete a feature one by one from the set containing  $m$  features. Calculate the accuracy  $A_i$  of the six classifiers selected, compare the accuracy of classifiers with different feature numbers, and gradually remove redundant variables.  $A_{\max}$  with the highest accuracy of each classifier is selected, and then, the best feature subset  $\max$  of the classifier is determined. The steps are shown in Figure 4.

Based on the selected 20 evaluation indicators, we gradually delete an indicator that has little impact on the reputation discrimination ability of crowdsourcing participants until the number of indicators reaches the specified number. By adjusting the number of selected evaluation indicators, we evaluate the accuracy of classifiers in the different numbers of evaluation indicators, select the selection method and feature subset of evaluation indicators with the highest classification accuracy, and then determine the reputation evaluation index system.

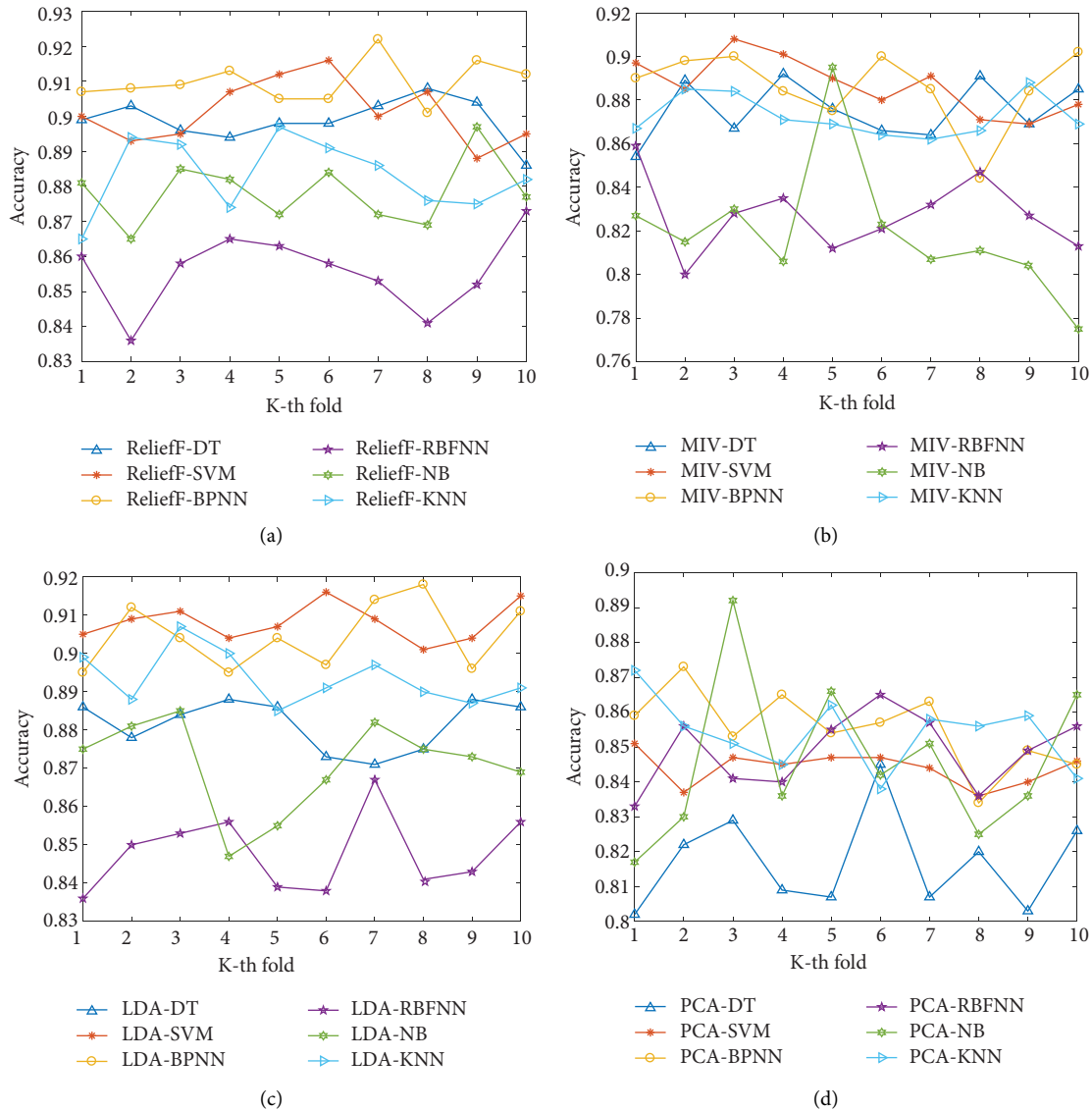


FIGURE 3: Ten-fold cross-validation diagram of the evaluation index selection classifier of crowdsourcing participants.

## 6.2. Select the Best Combination of Reputation Evaluation Indicators

### 6.2.1. Selection Results of the Best Evaluation Indicators.

The accuracy of the classifier selected by the evaluation index changes with the number of selected features. The experimental results show that, after using ReliefF dimensionality reduction, the crowdsourcing participant evaluation index based on six different machine learning algorithms selects the classifier, and the classification accuracy is the highest when selecting the best feature subset. The optimal feature number of ReliefF-KNN is 8, the optimal feature number of ReliefF-DT, ReliefF-SVM, and ReliefF-RBFNN algorithms is 9, ReliefF-BPNN is 12, and relief NB is 13. When the number of features is different, the accuracy results of ten-fold cross-validation of the classifier selected by the evaluation index are shown in Table 4 and Figure 5.

As shown in Figure 5, the accuracy of the six classifiers when selecting the best feature subset is higher than that

when selecting 20 index variables as the feature subset. When the classifier selects the best feature subset, the number of evaluation indexes decreases, but the accuracy of the classifier is improved, among which the accuracy of ReliefF-RBFNN is improved the most. The research shows that the classifier has the highest accuracy when the best feature subset is selected; adding new redundancy indicators or reducing the indicators of the best feature subset will lead to the decline of classification accuracy.

Combined with the selection results of evaluation indicators in the first and second stages, the weight ranking of the selected reputation evaluation indicators of crowdsourcing participants is shown in Table 5.

### 6.2.2. Analysis of Evaluation Index Selection Results.

The accuracy of the six classifiers was tested by the Kruskal-Wallis test when the reputation evaluation index variable was set to 20 and when the best evaluation index

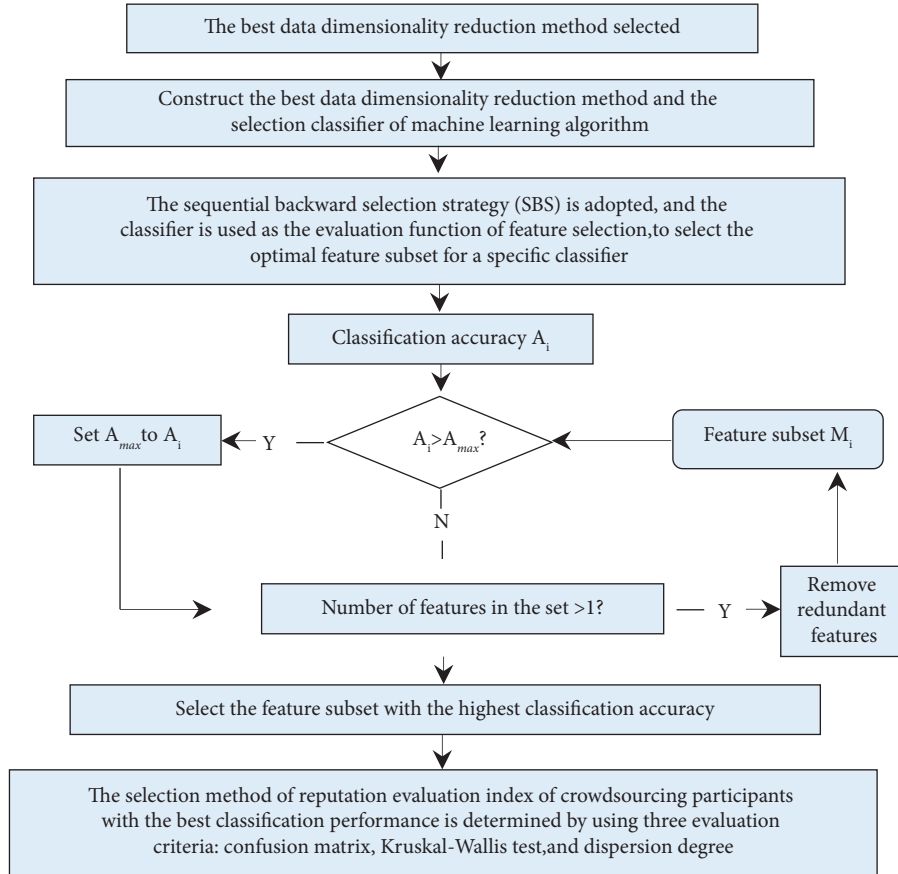


FIGURE 4: Selection method of the best evaluation index combination based on ReliefF.

variable was selected. The average rank of classifier accuracy when the best subset is represented by the blue column; the yellow column indicates the average rank of classifier accuracy when the number of selected evaluation indicators is 20. The Kruskal–Wallis test results of classifier accuracy selected by reputation evaluation indicators are shown in Figure 6.

The test results show that the K-W statistic is 92.661, and the probability  $P$  value is close to 0. After selecting the best feature subset, the performance of the six reputation evaluation index selection classifiers is improved to varying degrees, and the classification performance of ReliefF-RBFNN is the most significant. After the selection of evaluation indicators, the highest average rank of ReliefF-SVM is 105.6 and the second and third are ReliefF-BPNN and ReliefF-DT, respectively, and the average rank is 102 and 81.65, respectively. The classification performance of the ReliefF-SVM algorithm is the best.

### 6.3. Comparative Evaluation of Selection Methods

**6.3.1. Confusion Matrix Analysis.** The evaluation results of the confusion matrix of the classifier selected by the reputation evaluation index of crowdsourcing participants are shown in Table 6.

In terms of precision, the first type precision rate of ReliefF-SVM is 0.991, and the first-type precision rate of ReliefF-BPNN and ReliefF-RBFNN is also more than 0.98. The second type of ReliefF-NB has the highest precision rate of 0.534, indicating that the classifier has the strongest ability to distinguish crowdsourcing participants with a medium reputation. The third category of ReliefF-DT has the highest precision rate of 0.736 and ReliefF-NB has the lowest precision rate of 0.515, with a difference of 0.221, indicating that ReliefF-DT performs best for crowdsourcing participants with a poor reputation, while ReliefF-NB performs worst.

In terms of recall rate, it is particularly important, for it reflects the misclassification costs, especially, the recall rate of the third category. The first kind of recall rate of ReliefF-NB is the highest, which is 0.953, followed by ReliefF-DT, which is 0.95. The second kind of recall rate of ReliefF-SVM is the highest, which is 0.561, followed by ReliefF-BPNN, which is 0.488. The highest recall rate of the third category is ReliefF-SVM with 0.857, indicating that the selection method has the strongest ability to distinguish crowdsourcing participants with a poor reputation, and the misclassification cost of the third category is the lowest.

When evaluating classifiers, it is difficult to comprehensively evaluate the function of classifiers only by their precision rate or recall rate. Therefore, the F1 measure value is often used to comprehensively evaluate classifiers combined with precision rate and recall rate. The

TABLE 4: Accuracy of ten-fold cross-validation of feature subset selected by classifier.

Feature subset	ReliefF-DT	ReliefF-SVM	ReliefF-BPNN	ReliefF-RBFNN	ReliefF-NB	ReliefF-KNN
5	0.9036	0.9104	0.9050	0.8930	0.8592	0.8835
6	0.9043	0.9131	0.9087	0.8967	0.8643	0.8826
7	0.9026	0.9112	0.9085	0.8974	0.8665	0.8875
8	0.9031	0.9132	0.9088	0.8981	0.8788	<b>0.8958</b>
9	<b>0.9055</b>	<b>0.9147</b>	0.9086	<b>0.8987</b>	0.8843	0.8934
10	0.9012	0.9117	0.9099	0.8938	0.8874	0.8897
11	0.8993	0.9120	0.9126	0.8920	0.8869	0.8905
12	0.8998	0.9111	<b>0.9132</b>	0.8860	0.8856	0.8894
13	0.9034	0.9127	0.9108	0.8850	<b>0.8890</b>	0.8895
14	0.9032	0.9140	0.9109	0.8800	0.8818	0.8872
15	0.9012	0.9104	0.9100	0.8740	0.8832	0.8870
16	0.8997	0.9112	0.9130	0.8710	0.8791	0.8926
17	0.8987	0.9137	0.9100	0.8670	0.8760	0.8900
18	0.9000	0.9089	0.9110	0.8630	0.8781	0.8870
19	0.8977	0.9052	0.9130	0.8570	0.8782	0.8848
20	0.8989	0.9010	0.9100	0.8560	0.8781	0.8830

The best result for each classifier is in bold.

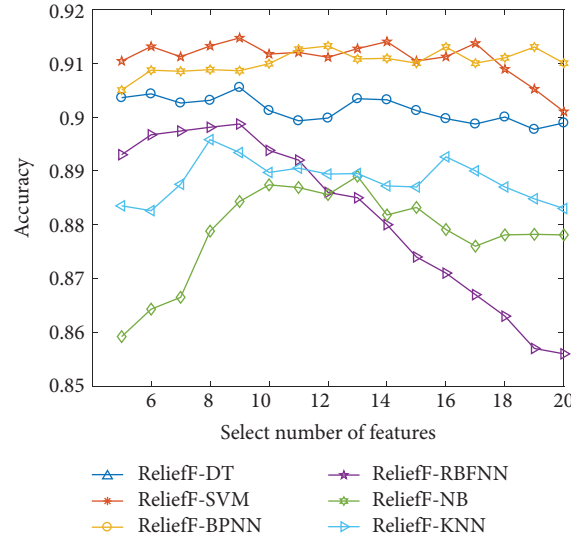


FIGURE 5: Ten-fold cross-validation of the feature subset selected by the classifier.

TABLE 5: Selection results of reputation evaluation indicators of crowdsourcing participants.

Rank	Index variable	Weight
1	$x_1$ positive feedback rate	0.054
2	$x_2$ number of punishments in three months	0.037
3	$x_3$ refund rate in three months	0.029
4	$x_4$ refund rate in this month	0.027
5	$x_5$ credibility frozen after reporting	0.023
6	$x_6$ transaction activity	0.023
7	$x_7$ work attitude	0.015
8	$x_8$ task completion quality	0.015
9	$x_9$ work speed	0.014
10	$x_{10}$ years of opening	0.014
11	$x_{11}$ comprehensive scoring	0.012
12	$x_{12}$ number of employer nonrecommend	0.008
13	$x_{13}$ margin deposit amount	0.008

maximum F1 measure of ReliefF-DT is 0.927, and the maximum F1 measure of ReliefF-SVM is 0.693 and 0.885, respectively. On the whole, ReliefF-SVM has good performance in accuracy and recall, high classification precision, and strong stability.

**6.3.2. Dispersion degree.** The robustness of the classifier is verified by the discrete degree of classification accuracy. The accuracy of the classifier is verified by ten-fold cross-validation, and the maximum value, minimum value, standard deviation, skewness coefficient, and kurtosis coefficient are calculated, as shown in Table 7.

In Table 7, the kurtosis coefficient of ReliefF-RBFNN accuracy is  $-0.014$ , indicating that the distribution form is relatively symmetrical and close to normal distribution. The kurtosis of the six evaluation index selection methods is negative, indicating that the data distribution is more

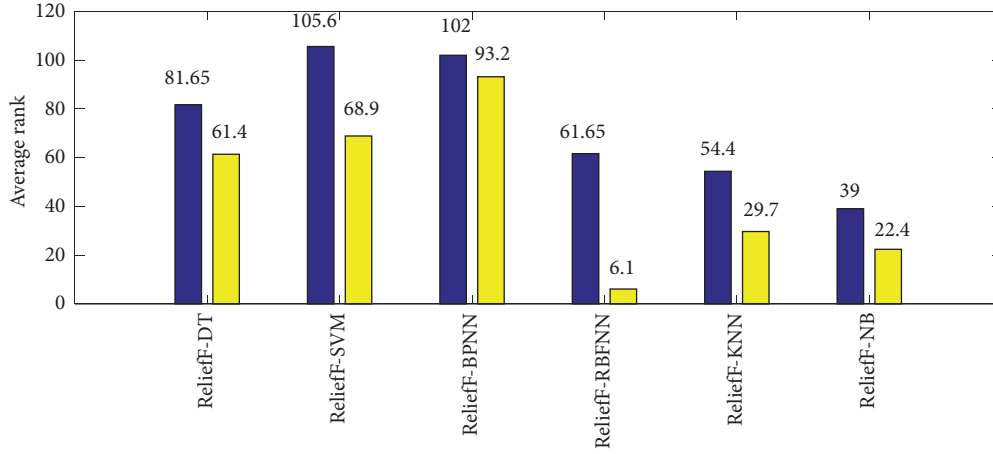


FIGURE 6: Kruskal-Wallis test results of accuracy of the classifier selected by the evaluation index.

TABLE 6: Evaluation results of crowdsourcing participants’ reputation: evaluation index selection classifier confusion matrix.

Data dimensionality reduction method	Machine learning technology	Number of indicators	Accuracy	Precision rate			Recall rate			F1-measure		
				P <sub>1</sub>	P <sub>2</sub>	P <sub>3</sub>	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	F <sub>1</sub>	F <sub>2</sub>	F <sub>3</sub>
ReliefF	DT	9	0.9055	0.964	0.431	<b>0.736</b>	0.950	0.477	0.799	<b>0.927</b>	0.624	0.849
	SVM	9	0.9147	<b>0.991</b>	0.291	0.675	0.934	<b>0.561</b>	<b>0.857</b>	0.924	<b>0.693</b>	<b>0.885</b>
	BPNN	12	0.9132	0.987	0.321	0.672	0.938	0.488	0.838	0.925	0.611	0.873
	RBFNN	9	0.8987	0.980	0.363	0.555	0.936	0.432	0.761	0.917	0.560	0.822
	KNN	8	0.8958	0.957	0.375	0.729	0.942	0.412	0.815	0.919	0.561	0.853
	NB	13	0.8890	0.951	<b>0.534</b>	0.515	<b>0.953</b>	0.416	0.756	0.920	0.567	0.814

TABLE 7: Accuracy and dispersion degree of the selection classifier of the reputation evaluation index of crowdsourcing participants.

Selection classifier	Minimum value	Maximum value	Mean value	Standard deviation	Skewness coefficient	Kurtosis coefficient
ReliefF-DT	0.898	0.915	0.906	0.006	0.406	-0.581
ReliefF-SVM	<b>0.906</b>	<b>0.922</b>	<b>0.915</b>	<b>0.005</b>	-0.485	-0.553
ReliefF-BPNN	0.905	<b>0.922</b>	0.913	0.006	0.576	<b>-0.282</b>
ReliefF-RBFNN	0.886	0.911	0.899	0.008	<b>-0.014</b>	-1.040
ReliefF-KNN	0.882	0.911	0.896	0.009	0.031	-0.335
ReliefF-NB	0.879	0.905	0.889	0.009	0.544	-0.474

The best result of the classifier is shown in bold under different measures.

gentle than the standard distribution. The minimum standard deviation of ReliefF-SVM is 0.005, and the maximum and minimum accuracy are the best levels of all methods. The discrete trend of accuracy from the center value is small, and the better the representation of mean value to data. It is shown that the ReliefF-SVM method has the best performance.

## 7. Conclusion

The credibility evaluation of crowdsourcing participants is a crucial issue for the rapid and healthy development of crowdsourcing. This study aims at the problems of a single reputation evaluation index, poor discrimination ability, and simple evaluation method of the participants of the crowdsourcing platform and studies the selection method of the reputation evaluation index of the participants of the crowdsourcing platform [37].

*7.1. Main Contributions.* This study has two main contributions.

First, a “dimension reduction-feature selection” method for selecting the optimal index combination is proposed. Firstly, the process selects the best data dimensionality reduction method from ReliefF, mean impact value (MIV), linear discriminant analysis (LDA), and principal component analysis (PCA). The sequential backward selection strategy (SBS) is adopted, and the accuracy of the classifier is used as the evaluation function of feature selection. The best feature subset is selected by evaluating the accuracy of the classifier with different feature numbers. An evaluation index selection method based on ReliefF-SVM is proposed, which has the best performance in accuracy, F1 measures, and stability.

Second, the best index combination of crowdsourcing participants’ reputation evaluation is proposed. The feature subset with the best classification performance is selected

using the proposed method (ReliefF-SVM) of selecting the best evaluation index. The research results show that the best combination of the chosen crowdsourcing participants' reputation evaluation indicators is: positive feedback rate, number of punishments, refund rate in three month, refund rate in this month, credibility frozen after reporting, transaction activity, work attitude, task completion quality, and work speed.

This study discusses the selection of reputation evaluation indicators of crowdsourcing participants for the first time. The proposed crowdsourcing participant reputation evaluation index combination makes up for the problem that the crowdsourcing platform evaluation index is single and cannot feedback the reputation status of the crowdsourcing participants. It solves the problem that a single evaluation index has a significant impact on the reputation status in the evaluation process, but the built-in index combination has a weak ability to distinguish the reputation status.

## 7.2. Main Conclusions

*7.2.1. When Using the ReliefF Method to Reduce the Dimension of Data, the Classification Performance of the Six Machine Learning Algorithms Is the Best.* Four dimensionality reduction methods of ReliefF, ReliefF, mean impact value (MIV), linear discriminant analysis (LDA), and principal component analysis (PCA), are selected, and six machine learning algorithms of decision tree (DT), BP Neural Network (BPNN), RBF neural network (RBFNN), support vector machine (SVM), K-nearest neighbor classifier (KNN), and Naive Bayes (NB) are used to cross construct 24 kinds of reputation evaluation index selection classifiers for crowdsourcing participants. Through the ten-fold cross-validation and Friedman test, the average accuracy of classifiers when different dimensionality reduction methods are used is compared. The research shows that the accuracy of selecting classifiers based on reputation evaluation indicators constructed by other dimensionality reduction methods is significantly different. When using the ReliefF method to reduce the dimension of data, the average rank of classifier accuracy is the highest, which is better than the LDA, PAC, and MIV and has the best effect on dimension reduction. The feature selection method adopted by ReliefF does not change the original index and is more explanatory than LDA and PAC feature extraction methods.

*7.2.2. The Reputation Evaluation Index Selection Method (ReliefF-SVM) of Crowdsourcing Participants Based on ReliefF Feature Selection Can Select Evaluation Indexes That Comprehensively, Objectively, and Effectively Identify the Reputation Status of Crowdsourcing Participants.* The best number of features of the ReliefF-KNN classifier is 8, the best number of features of ReliefF-DT, ReliefF-SVM, and ReliefF-RBFNN classifier is 9, and the best number of features of ReliefF-BPNN and ReliefF-NB are 12 and 13, respectively. The experimental results show that the classifier selected by the six crowdsourcing participants' reputation

evaluation indicators based on ReliefF feature selection has the highest classification accuracy when choosing the best feature subset. Adding new redundancy indicators or reducing the indicators of the best feature subset will lead to the decline of the accuracy of the classifier.

The selected ReliefF feature selection method and six machine learning algorithms are used to build a reputation evaluation index selection classifier for crowdsourcing participants. The results are analyzed and compared by the Kruskal-Wallis test, confusion matrix, and dispersion degree. The results show that ReliefF-SVM has the highest accuracy of 0.906, the highest F1 measure values of the second and third categories are 0.693 and 0.885, respectively, and the minimum standard deviation is 0.005. The experimental results show that the ReliefF-SVM classifier is excellent in accuracy; F1 measures value and stability and has stronger robustness and promotion value.

*7.3. Future Research Directions.* Future research directions include the following: how to further explore the classification problem of the new sample reputation of crowdsourcing participants by improving the usability of classifiers, reducing resource consumption, and improving prediction ability? In the machine learning algorithm, how to constantly update and adapt the new data, reduce the amount of computation required to train the classifier repeatedly to learn new and old data, and improve the performance and accuracy of the classifier.

Based on the best evaluation index combination and evaluation method research, how to combine the reputation of crowdsourcing participants with the recommendation of crowdsourcing tasks is studied. According to the reputation of crowdsourcing participants, how to achieve task recommendation under multidimensional constraints, how to recommend task selection sequences for crowdsourcing participants through matching algorithms, and how to improve the transaction success rate need further discussion.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China, under Grant 72101235, the Zhejiang Federation of Social Sciences, under Grant 2023B040, the State Scholarship Fund, under Grant 202108330330, Scientific Research Foundation of Zhejiang University of Water Resources and Electric Power, under Grant xky2022051, the Scientific Research Projects of Zhejiang Education

Department, under Grant Y202045307, the Soft Science Project of Zhejiang Provincial Science and Technology Department under Grant 2022C35057, and the General Research Project of Humanities and Social Sciences of the Ministry of Education, under Grant 21YJC790062.

## References

- [1] Y. R. Huang and M. Chen, "Multidimensional reputation evaluation model for crowdsourcing participants based on big data," in *Proceedings of the IEEE Int. Conf. on High Performance Big Data and Intelligent Systems*, pp. 41–46, Shenzhen, China, May 2019.
- [2] Y. R. Huang and M. Chen, "Improve reputation evaluation of crowdsourcing participants using multidimensional index and machine learning techniques," *IEEE Access*, vol. 7, no. 1, pp. 118055–118067, 2019.
- [3] J. Howe, "The rise of crowdsourcing," *Wired Magazine*, vol. 14, no. 6, pp. 1–5, 2006.
- [4] Y. R. Huang, R. Wang, B. Huang, B. Wei, S. L. Zheng, and M. Chen, "Sentiment classification of crowdsourcing participants' reviews text based on LDA topic model," *IEEE Access*, vol. 9, pp. 108131–108143, 2021.
- [5] L. N. Hao, "Research on enterprise marketing innovation strategy based on crowdsourcing mode information sharing under the sharing economy," *Journal of Hebei University of Economics and Trade*, vol. 18, no. 04, pp. 54–60, 2018.
- [6] J. F. Li and Y. Y. Zhou, "Information disclosure and market concentration of it crowdsourcing service providers," *Journal of Systems Management*, vol. 30, no. 1, pp. 88–93, 2021.
- [7] X. Han, M. X. Xue, and W. H. Song, "Role of reputation in sustainable performance of online crowdsourcing vendors: an explanation from transaction cost theory," *Frontiers in Psychology*, vol. 12, pp. 756134–756214, Nov. 2021.
- [8] W. J. Jiang, Y. S. Xu, and H. Guo, "Dynamic trust calculation model and reputation management mechanism of online transactions," *Chinese Science: Information Science*, vol. 44, no. 9, pp. 1084–1101, 2014.
- [9] J. Yan, S. P. Ku, and C. Yu, "Reputation model of crowdsourcing workers based on activity degree," *Computer Applications*, vol. 37, no. 7, pp. 2039–2043, 2017.
- [10] J. F. Liu and X. Y. Jiang, "Construction of evaluation system for it service providers in crowdsourcing Market," *Journal of system management*, vol. 28, no. 5, pp. 857–863, 2019.
- [11] X. F. Zhang and J. F. Su, "A combined fuzzy DEMATEL and TOPSIS approach for estimating participants in knowledge-intensive crowdsourcing," *Computers & Industrial Engineering*, vol. 137, p. 106085, 2019.
- [12] M. Wang, G. L. Wang, Y. Zhang, and Z. Li, "A high-reliability multifaceted reputation evaluation mechanism for online services," *IEEE Transactions on Services Computing*, vol. 12, no. 6, pp. 836–850, 2019.
- [13] C. Q. Jiang, Z. Wang, R. Y. Wang, and Y. Ding, "Loan default prediction by combining soft information extracted from descriptive text in online peer-to-peer lending," *Annals of Operations Research*, vol. 266, no. 1–2, pp. 511–529, 2018.
- [14] Y. J. Zhang, G. T. Chi, and Z. Zhang, "Decision tree for credit scoring and discovery of significant features: an empirical analysis based on Chinese microfinance for farmers," *Filomat*, vol. 32, no. 5, pp. 1513–1521, 2018.
- [15] T. Y. Li, H. J. Wang, J. Wu, Z. L. Zhang, and C. J. Tang, "Sparse Bayesian learning for credit risk evaluation," *Journal of Computer Applications*, vol. 33, no. 11, pp. 3094–3096, 2013.
- [16] B. Wei, X. Wang, X. Xia, M. F. Jiang, Z. H. Ding, and Y. R. Huang, "Novel self-adjusted particle swarm optimization algorithm for feature selection," *Computing*, vol. 103, pp. 1569–1597, 2021.
- [17] H. Zhang, Y. X. Shi, X. R. Yang, and R. L. Zhou, "A firefly algorithm modified support vector machine for the credit risk assessment of supply chain finance," *Research in International Business and Finance*, vol. 58, pp. 101482–101521, 2021.
- [18] Z. C. Zhao, G. T. Chi, and X. P. Bai, "Mining farmers' key default characteristics based on the least significant difference method," *System Engineering Theory and Practice*, vol. 40, pp. 2339–2351, 2020.
- [19] T. Zhang and G. T. Chi, "Default discrimination model based on optimal combination of credit characteristics -- a case study of Chinese A-share listed companies," *System engineering theory and Practice*, vol. 40, no. 10, pp. 2546–2562, 2020.
- [20] H. L. Zhan, W. Yan, and B. Song, "Online game of cross-border e-commerce platform reputation based on incomplete information," *Practice and Understanding Of Mathematics*, vol. 51, pp. 57–69, 2021.
- [21] J. Quan, C. X. Tang, Y. W. Zhou, X. J. Wang, and J. B. Yang, "Reputation evaluation with tolerance and reputation-dependent imitation on cooperation in spatial public goods game," *Chaos, Solitons & Fractals*, vol. 131, pp. 1–9, 2020.
- [22] Y. J. Wang, Z. P. Cai, X. R. Tong, Q. X. Pan, Y. Gao, and G. S. Yin, "Online incentive mechanism of reputation based mobile crowdsourcing system," *Computer application*, vol. 36, no. 8, pp. 2121–2127, 2016.
- [23] J. F. Lu, C. B. Tang, X. Li, and Q. Wu, "Designing socially-optimal rating protocols for crowdsourcing contest dilemma," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 6, pp. 1330–1344, 2017.
- [24] X. Y. Zhu, Y. Li, L. Fang, and P. Chen, "An improved proof-of-trust consensus algorithm for credible crowdsourcing blockchain services," *IEEE Access*, vol. 8, pp. 102177–102187, 2020.
- [25] S. Bhattacharjee, N. Ghosh, V. K. Shah, and S. K. Das, "QnQ: a reputation model to secure mobile crowdsourcing applications from incentive losses," in *Proceedings of the Communications and Network Security*, pp. 1–9, Las Vegas, NV, USA, October 2017.
- [26] X. Y. Lu, Q. Lu, M. M. Huang, and Z. Q. Li, "Construction of contractor reputation evaluation model under crowdsourcing mode based on emotional tendency," *Statistics & Decisions*, vol. 34, no. 17, pp. 177–180, 2018.
- [27] L. Sun, Q. Yang, X. Chen, and Z. X. Chen, "RC-chain: reputation-based crowdsourcing blockchain for vehicular networks," *Journal of Network and Computer Applications*, vol. 176, pp. 102956–103017, 2021.
- [28] A. Alqwadri, M. Azzeh, and F. Almasalha, "Application of machine learning for online reputation systems," *International Journal of Automation and Computing*, vol. 18, no. 3, pp. 492–502, 2021.
- [29] A. Rantanen, J. Salminen, F. Ginter, and B. J. Jansen, "Classifying online corporate reputation with machine learning: a study in the banking domain," *Internet Research*, vol. 30, pp. 45–66, 2020.
- [30] T. D. Yang, X. D. Fu, L. Liu, K. Yue, L. J. Liu, and Y. Feng, "Online service reputation measurement based on multi-dimensional evaluation information," *Minicomputer System*, vol. 39, no. 12, pp. 2625–2631, 2018.
- [31] Y. Yu, S. M. Liu, L. Guo et al., "CrowdR-FBC: a distributed fog-blockchains for mobile crowdsourcing reputation



- management,” *IEEE Internet of Things Journal*, vol. 7, no. 9, pp. 8722–8735, 2020.
- [32] M. Kealeboga, “Volunteer reputation determination in crowdsourcing projects using latent class analysis,” *Transactions in GIS*, vol. 25, no. 2, pp. 1–17, 2020.
- [33] G. W. Dombi, P. Nandi, J. M. Saxe, A. M. Ledgerwood, and C. E. Lucas, “Prediction of rib fracture injury outcome by an artificial neural network,” *The Journal of Trauma, Injury, Infection, and Critical Care*, vol. 39, no. 5, pp. 915–921, 1995.
- [34] T. Cover, J. Thomas, and J. Wiley, “Elements of information theory,” *Tsinghua University Press*, vol. 12, pp. 73–75, 2003.
- [35] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning representations by back-propagating errors,” *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.
- [36] D. West, “Neural network credit scoring models,” *Computers & Operations Research*, vol. 27, no. 11-12, pp. 1131–1152, 2000.
- [37] Z. Zheng, X. S. Xu, W. Zhao, D. Chen, and L. Yang, “Computational experiment paradigm of choice behavior research -- Based on complex scientific management methodology,” *Technology Economics*, vol. 37, no. 1, pp. 84–92, 2018.