

Retraction

Retracted: Enterprise Credit Security Prediction and Evaluation Based on Multimodel Fusion

Security and Communication Networks

Received 8 January 2024; Accepted 8 January 2024; Published 9 January 2024

Copyright © 2024 Security and Communication Networks. 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This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] L. Zhang, J. He, and Z. Zhao, "Enterprise Credit Security Prediction and Evaluation Based on Multimodel Fusion," *Security and Communication Networks*, vol. 2022, Article ID 2754302, 12 pages, 2022.

Research Article

Enterprise Credit Security Prediction and Evaluation Based on Multimodel Fusion

Lei Zhang ^{1,2}, Jie He ², and Zihao Zhao ²

¹School of Economics and Management, Chongqing Jiaotong University, Chongqing 400074, China

²School of Mathematics and Statistics, Chongqing Jiaotong University, Chongqing 400074, China

Correspondence should be addressed to Lei Zhang; zhangleicqjtu@163.com

Received 13 April 2022; Revised 23 July 2022; Accepted 29 July 2022; Published 29 August 2022

Academic Editor: Yuchuan Luo

Copyright © 2022 Lei Zhang 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.

Based on the industry data and enterprise data from tens of thousands of small and medium-sized enterprises, a deep learning and machine learning model of credit prediction is constructed through the division of data sets, processing, and integration of models. At first, with the help of two characteristic selection methods, several subsets separated from the dataset are analyzed based on convolutional neural network as coarse prediction. Then, combined with the tree model, the precise prediction is further made for the enterprise credit evaluation. Finally, the model fusion is carried out to obtain high-precision results. In the simulation experiment, this paper takes a data set of 14,366 small and medium-sized enterprise credit evaluations as the analysis samples to verify the results. The accuracy of the model is 97%, which is far more than 93% of single model with metadata set.

1. Introduction

According to statistics, by the end of 2018, the number of small and medium-sized enterprises in China has exceeded 30 million, and the number of individual industrial and commercial households has exceeded 70 million, taking over more than 50 percent of the country's tax revenue, 60 percent of GDP, and 70 percent of technology innovation. The achievements and employment of more than 80 percent of the workforce play an increasingly important role in supporting the national economy. The difficulty of financing and loans for small and micro-businesses is particularly acute, and the risks and challenges they face are severe and complex.

The credit evaluation of small and medium-sized enterprises uses the machine learning models and algorithms to carry on the statistics of enterprises themselves. Calculation and qualitative analysis are used to predict whether an enterprise can be included in the trust-breaking enterprises. Existing credit evaluation models and subset selection models can be roughly divided into traditional statistical models, machine learning models, subset selection of characteristic numerical attributes, and subset selection

based on models. The earliest credit evaluation models are generally based on classical statistical theory, which mainly includes logistic regression, linear regression, and Naive Bayesian classification. In recent years, with the rapid progress of computing capability and algorithm as well as the substantial increase in data scale, methods in the fields of machine learning and deep learning have been used in enterprise credit evaluation and achieved excellent results. At present, the representatives of data mining models include SVM, K means, decision tree, artificial neural network, and artificial neural network of long- and short-term memory. In this paper, deep learning, machine learning, and other methods are used to predict high-precision model results, so as to provide a good reference for real enterprise credit evaluation system.

For the credit evaluation of small and medium-sized enterprises in reality, an enterprise with a relatively high credit evaluation score will be relatively easier in bank lending, private financing, and government support in reality. On the other hand, it also provides direction and guidance for companies with low ratings to some extent. Liu et al. [1] studied the financial market in China, Yoshino and Taghizadeh-Hesary [2] studied the financial market of the

whole Asia, and the research results show that the financial center must be in Asia in the future. Hu et al. [3] proposed a hybrid integration model based on support vector machine, which randomly sampled the feature space and then trained the model with support vector machine. However, the accuracy was not good, and the random sampling and decision tree model could not make use of all the sample space. Gregori et al. [4] proposed a multi-criteria credit rating model for the financing of small and medium-sized enterprises, but the ductility of this model was not good enough. Zhang et al. [5] used the fuzzy BP neural network to predict investment risks, and the overall effect of the model was good. Some studies utilize back propagation neural network [6, 7] to study the risk of financing and credit risk of enterprise supply chain. Fonseca et al. [8] compared the effectiveness of fuzzy neural network and BP neural network in discriminating the credit rating of small and micro-enterprises, and the result showed that the fuzzy neural network is more effective. Li et al. [9], improved the method of XGBoost algorithm and proposed fuzzy XGBoost on the basis of adding fuzzy membership degree. Zhou et al. [10] also proposed some solutions on fuzzy algorithm. Zhang et al. [11, 12] put forward a large-sample mixed credit evaluation model based on similar sample merge and a three-stage mixed credit evaluation model based on multi-attribute subset selection strategy, respectively, in 2018 and 2019. Sun et al. [13] put forward a series of solutions to the problem of uneven distribution in terms of credit evaluation.

For the small and medium-sized enterprise credit evaluation models based on subset selection strategy of feature's numerical attributes above, most of them just simply remove the redundant features from metadata, and put a subset in one or more basic models to train. But they did not further pursue higher accuracy, nor did they compare and verify the results of selected subsets based on different base models. Is the accuracy of the selected subset training model higher than that of the metadata-trained model? Is there any difference among different subsets? If the data are redundant, does that mean the redundant data are completely useless? Can subsets highlight local features of the feature space? Can an algorithm be used to mine the relationships among data? Is it possible to achieve higher accuracy through model fusion? These are the emphases of this paper.

In order to carry out data mining, first, data should be preprocessed, outliers should be proposed, and the missing values should be filled. Next, even if the noise samples are processed, the metadata often contains data redundancy. The redundant data often cannot further improve the accuracy of the model and sometimes even prevent the model from reaching better accuracy. Last, train and verify the model. The speed of model training can be further accelerated by the selection of features and the reduction of training samples.

Aiming at the shortcomings of existing research, this paper proposes a multimodel hybrid integration model based on convolutional neural network. The main innovations are as follows:

- (1) This paper summarizes the methods and principles needed in models firstly and then combines different subset division methods: correlation coefficient and GBDT subset division. Random forest model is constructed to verify the subsets after division to ensure that the accuracy will not drop too much.
- (2) On the basis of dividing four subsets, the convolutional neural network is used to explore the relationship between data and the four channels of the convolutional network. Use three models including extreme gradient boosting, light gradient boosting machine, and CatBoost to train, respectively, the four subsets and the four-channel convolutional neural network.
- (3) Bayes ridge regression was used to fuse the advantages of all single models in different subsets with convolutional neural network channels, and a more robust model is trained whose prediction result is more uniformly distributed and more in line with the real credit evaluation situation with an accuracy of up to 97%, which is much higher than 93% of single model with metadata set.

2. Enterprise Credit Evaluation Methods and Technologies

This section first introduces the theory and technology that can explain the credit evaluation of small and medium-sized enterprises, the evaluation method of model performance, the volume neural network, and the machine learning tree model.

2.1. Evaluation Method. In this paper, to evaluate the effectiveness of the model, we use AUC (area under curve) as the evaluation metric. AUC is an evaluation index for binary classification models, and its actual meaning is the area under the ROC (receiver operating characteristic) curve. AUC considers the classifier's ability to classify positive and negative examples, and can still make a reasonable evaluation of the classifier in the case of unbalanced samples.

The ROC curve is generated based on the ground-truth class and predicted probability of the sample. The horizontal axis is FPR (false-positive rate), and the vertical axis is TPR (true-positive rate), as shown in equation (1). Among them, TPR represents the proportion of all samples whose true class is 1, and the predicted class is 1; and FPR represents the proportion of all samples whose class is 0, and the predicted class is 1.

$$TPR = \frac{TP}{TP + FN}, FPR = \frac{FP}{FP + TN}, \quad (1)$$

where TP (true positive), FN (false negative), FP (false positive), and TN (true negative) are generated by confusion matrix. We use the AUC_ROC_SCORE [14] function in scikit-learn to calculate the AUC value.

2.2. Credit Prediction Technologies

2.2.1. Random Forest. Bagging [15] gives datasets including m samples. We randomly take out a sample and put it into the sample set, and then put this sample back into the initial data set. It is still possible to select the sample in the next sampling, and then a sample set of m samples is obtained after m cycles.

Random forest (RF) [16, 17] is an extension of bagging. On the basis of building bagging integration based on decision learning, RF further introduces random attribute selection in the training process of decision tree. To be specific, the traditional decision tree selects the most attribute in the current node when selecting and dividing attributes, whereas in RF, for each node of the base decision tree, a subset containing k attributes is selected randomly from the attribute set of this node, and then selected an optimal attribute user to divide from this subset.

Random forest is used to test whether the error of subset division is too large. Random forest model is a model based on variance, so it is very difficult to fit the predicted results, and it is a good model to test the selection of characteristic subset.

2.2.2. Convolutional Neural Network. Convolutional neural network (CNN) is a feedforward neural network, whose artificial neurons can respond to a part of the surrounding units in the coverage range and perform well in regular data processing [18–21].

Convolutional neural network is composed of one or more convolutional layers and the top fully connected layer, and also includes associated weight and pooling layer at the same time. This structure enables convolutional neural network to utilize the two-dimensional structure of the input data. Compared with other deep learning structures, convolutional neural networks can provide better results in image and speech recognition and can also be trained using a backpropagation algorithm. Compared with other deep and feedforward neural networks, convolutional neural networks require fewer parameters to be considered, which makes it an attractive deep learning structure.

2.2.3. Decision Tree Model. Decision tree [22] and its advanced models are the kind of algorithms that divide the input space into different regions, and each region has its own weighted parameters. In machine learning, decision tree is a prediction model. It represents a mapping relationship between object attributes and object values. Each node in the tree represents an object. The join node represents a small region divided into a decomposition space, and each branch fork path represents a possible attribute value—the node weight. Each leaf node represents the value of the object represented by the path from the root node to the leaf node.

Let $\hat{y}_i^{(t)}$ represent the predicted value of the i th example in the t th tree, $\omega(f_t)$ is the size of the t th tree, $f_t(x_i)$ is the weight value of the leaf node corresponding to the i th

training sample in the t th tree, and $L^{(t)}$ is the target function of the t th tree.

$$L^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \omega(f_t). \quad (2)$$

For $f_t(x_i)$, Taylor quadratic expansion is used, where g_i is the first gradient of the loss function to $\hat{y}_i^{(t-1)}$ and h_i is the second gradient of the loss function to $\hat{y}_i^{(t-1)}$.

$$L^{(t)} = \sum_{i=1}^n \left[l(y_i, \hat{y}_i^{(t-1)}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \omega(f_t). \quad (3)$$

The loss function value of tree $t-1$ in front of the t th tree is known, and $\tilde{L}^{(t)}$ is an approximate value, so

$$\tilde{L}^{(t)} = \sum_{i=1}^n \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \omega(f_t). \quad (4)$$

$I_j = \{i | x_i = j\}$ is defined as the mapping of the i th sample to the j th cotyledon node

$$\tilde{L}^{(t)} = \sum_{i=1}^n \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (5)$$

$$= \sum_{j=1}^T \left[\left(\sum_{i \in I_j} g_i \right) w_j + \frac{1}{2} \left(\sum_{i \in I_j} h_i + \lambda \right) w_j^2 \right] + \gamma T.$$

Take the first derivative of w_j from $\tilde{L}^{(t)}$, set it to 0, and get

$$w_j^* = \frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda}. \quad (6)$$

Bring the w_j^* back into the upper form:

$$\tilde{L}^{(t)} = -\frac{1}{2} \sum_{j=1}^T \frac{\left(\sum_{i \in I_j} g_i \right)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T. \quad (7)$$

So, we can get the solution of the objective function.

3. Data Processing of Credit Assessment Model

3.1. Data Preprocessing. In this paper, the data which come from the Statistics Bureau of Shandong Province include enterprise type, the registration authority, the enterprise status, the total amount of investment, registered capital, industry code, industry, business category, the jurisdiction of the organs, the value-added tax, corporate income tax, stamp duty, education, urban construction tax and the rest in early and late, a series of registration time correlation of dynamic data, a total of 200 or so.

According to the obtained data, make feature combination as far as possible, for example, subtract all year-end values from the initial values of last year, so as to construct new features and expand the sample space of training set. However, combining as many features as possible also brings some problems. If data redundancy occurs, then based on the tree model, there are two very important features that are

very good for predicting the analytic hyperplane, but if the similarity of these two features is very high, data redundancy will occur. In a word, if the weight of split points of the tree model for these two features is very low, it will deteriorate the effect of the model. From this point of view, feature selection becomes quite important for the tree model.

3.2. Feature Selection—Using Random Forest to Test the Feasibility. Filter-based feature selection, also known as filtering method, scores every feature according to the divergence or relevance of each feature score, and set a threshold or the number of thresholds that are about to be selected and then select features. The steps for using a filter in this article are as follows:

The correlation coefficient method: Calculate the correlation sparsity among the features, and the features with larger Pearson coefficient are selected according to the threshold.

Directly delete features whose variance does not meet the threshold; directly delete one of the two corresponding features with a large Pearson coefficient.

3.2.1. The Correlation Coefficient Method. For the correlation between the two characteristics of A and B , we can obtain the correlation degree $r_{A,B}$ by calculating the correlation coefficient of A and B (also called Pearson product moment coefficient, Pearson's product moment coefficient).

$$r_{A,B} = \frac{\sum_{i=1}^n (a_i - \bar{A})(b_i - \bar{B})}{n\sigma_A\sigma_B} \quad (8)$$

$$= \frac{\sum_{i=1}^n (a_i b_i) - n\bar{A}\bar{B}}{n\sigma_A\sigma_B},$$

where n is the number of tuples; a_i and b_i are the values of tuples i on A and B , respectively; \bar{A} and \bar{B} are the mean values of A and B , respectively; σ_A, σ_B are the standard deviations of A and B , respectively; and $\sum (a_i b_i)$ is the cross product of A and B (for each tuple, the product within the element is used), Pearson greater than zero means a positive correlation, indicating that one value increases as another value increases; A negative value of Pearson indicates that one value decreases as the other increases. The larger the value is, the stronger the correlation is, and the stronger the correlation is, the more obvious the data redundancy is.

Based on the analysis of enterprise type, registration authority, enterprise status, registered capital, industry code, industry category, jurisdiction authority, and other category characteristics, only part of the data with high correlation is retained here, and it can be found that there is a certain correlation between industry category and industry code, jurisdiction authority, and registration authority. As shown in Figure 1, especially for tax characteristics, there is an obvious correlation between value-added tax and education expenses, and between value-added tax and urban construction tax.

Based on Pearson coefficient, a thermal diagram can be made to further determine the specific relevant values. As shown in Figure 2, there are three features with high correlation: value-added tax, urban construction tax, and education fee. Taxpayers in the urban construction tax are 7%. The construction tax in county towns is 5%. Elsewhere, the city construction tax is 1%. The surcharge of education expenses is (VAT + consumption tax actually paid)×3%. Based on this, further feature engineering is carried out and further data relationship is mined.

Cross-features can be made for features with medium or low correlation, such as Pearson coefficient interval [0.2, 0.5], so as to obtain certain cross-new features, so as to achieve the best effect of the model.

In order to avoid data redundancy caused by highly correlated features, we divide the data set with high correlation coefficient into m subsets, and the union of all subsets is N -dimensional data. Then, the model is trained in turn, and finally the model is fused to get the best data effect. The features with high correlation coefficient in the metadata are split into two subsets, where the metadata is divided into subsets 1 and 2. The results of random forest test are shown in Table 1.

According to the comparison of the results in Table 1, although the accuracy of both subset 1 and subset 2 is lower than that of metadata set, they avoid cleverly data redundancy and have only a small difference in accuracy. Therefore, the division of data set can be considered reasonable.

3.2.2. Tree Model Selection. Tree model selection is a model-based selection method and not based on the basic attributes of numerical values. The tree model divides the nodes based on the Gini index of information theory to get the index value of the features.

Assuming that the proportion of class K samples in the current sample set D is p_k ($k = 1, 2, \dots, |Y|$), the Gini index of D is defined as shown in

$$\text{Gini}(D) = \sum_{k=1}^{|Y|} \sum_{k' \neq k} p_k p_{k'} = 1 - \sum_{k=1}^{|Y|} p_k^2. \quad (9)$$

Intuitively, $\text{Gini}(D)$ reflects the probability that two samples are randomly drawn from data set D with different category markers. So, the smaller the $\text{Gini}(D)$ is, the higher the improvement in data set D is made. In the set of candidate attributes, the attribute with the lowest Keeny index is selected as the dividing attribute.

The metadata is divided into a subset 3 and a subset 4, where subset 3 is considered as an important feature in the GBDT model, and subset 4 is not considered important to the GBDT model. Using random forest to test the subsets, the test results are shown in Table 2.

As shown in Table 2, the accuracy deviation between the two subsets is not significant and can be regarded as reasonable. The above two different methods are as follows: the first is based on mathematical relations, and the second is based on model selection. Due to certain deviations of

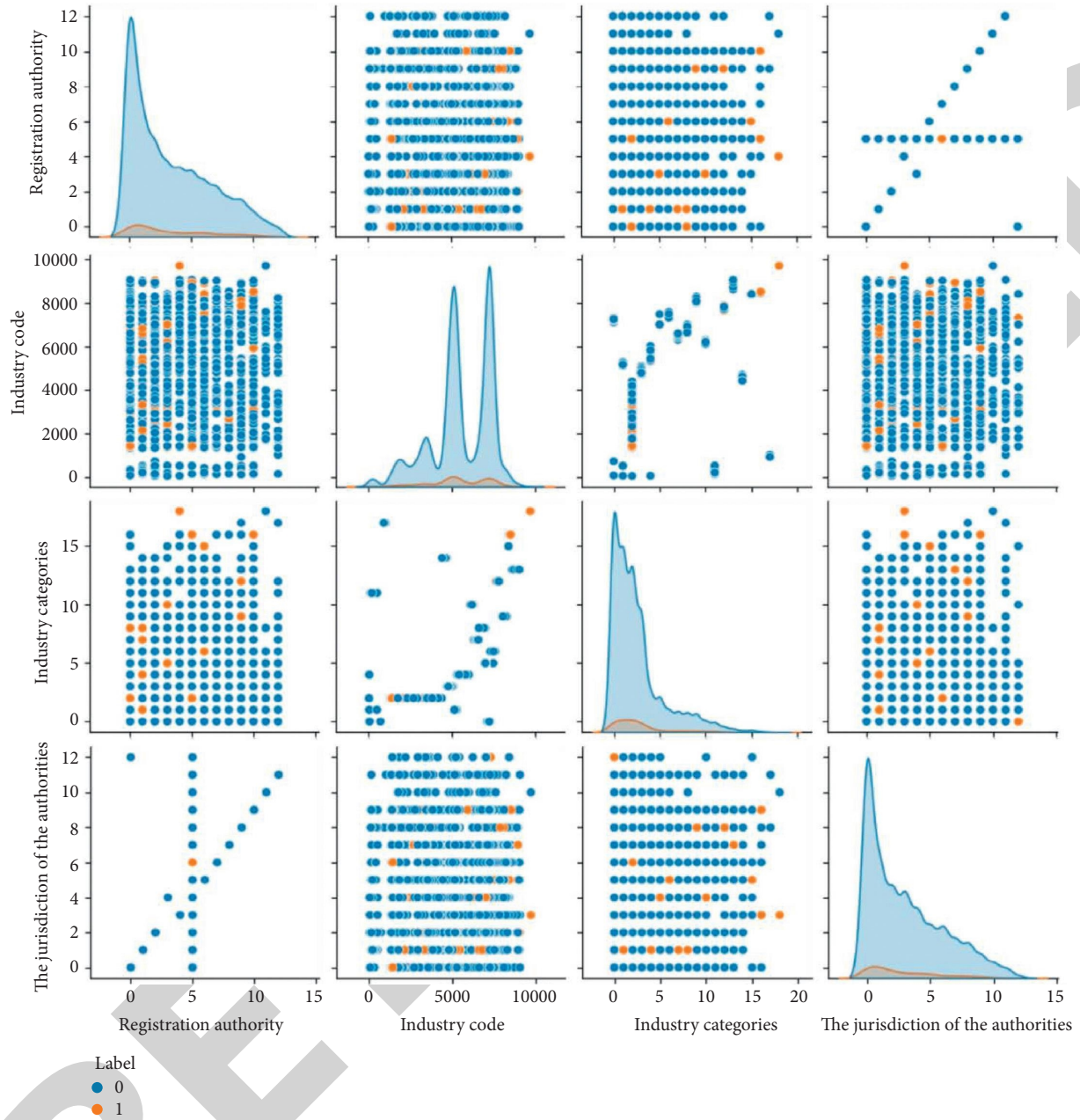


FIGURE 1: Feature space distribution.

results, the uncertainty of performance of feature selection method exists. Therefore, this paper will try to achieve higher scores by comparing different methods and different subsets longitudinally, and, at the same time, with undivided subsets. And use only the correlation coefficient method and GBDT selection method, after which there will be subset 1, subset 2, subset 3, and subset 4. When the sample space is segmented, the segmented sample space pays more attention to the local sample space of the features than the global sample space. So, the precision of the single model will decrease, but the precision of the whole model should be guaranteed not to be lost simultaneously. In the following research, we use stacking model fusion method to make full

use of each data and construct the model to take better care of the global sample.

4. Design of Credit Evaluation Model

After processing the outlier data and the feature, we divide the metadata into four channels by using the correlation coefficient and GBDT, and get the sub-data set of the four channels, using convolutional neural network to capture the interaction among the data as a prediction, and different channels correspond to different rough predictions. The corresponding rough predictions of four channel results are taken as the characteristics of XGB, LGB, and CB. Make

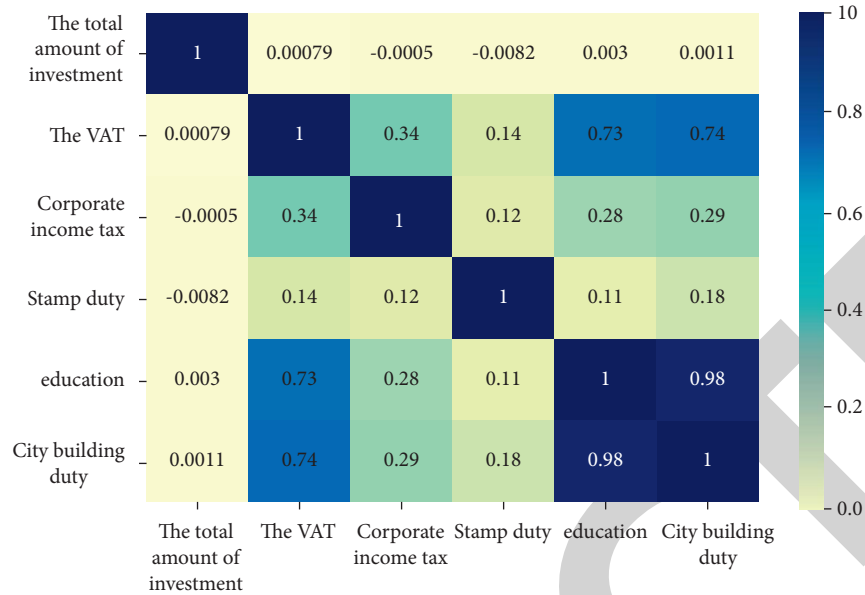


FIGURE 2: Correlation coefficient of tax characteristics.

TABLE 1: The correlation coefficient method.

Model accuracy	Metadata	Subsets 1	Subsets 2
Random forest	91.139%	89.255%	91.051%

TABLE 2: Tree model selection.

Model accuracy	Metadata	Subset 3	Subset 4
Random forest	91.139%	91.139%	88.911%

predictions correspondingly and respectively, and 12 fine prediction results are obtained after model fusion. The specific operation process is shown in Figure 3.

4.1. Convolutional Neural Network Seeks Data Representation. Combined with the actual data, the data sets collected by enterprises often have a lot of missing data. So, missing values must be filled in before using neural network model. Encode the data to meet the input of the convolutional neural network, and attempt to find the data mining point from the angle of the relationship between the numerical representations of the data.

Due to the small amount of data and the low dimension of two-dimensional matrix, a relatively simple convolutional neural network structure is used. The parameters of convolutional neural network are shown in Table 3.

The results of the four-channel experiment were obtained by plotting, respectively, showing the loss and AUC of binary cross-entropy and the loss and AUC of the verification set.

As shown in Figure 4, in the case of constant iteration of convolutional neural network, the loss function value of training set and verification set is close to the value of AUC, so it can be considered that the convolutional neural network has learned the preliminary representation of data.

4.2. Multi-Subset Comparison—Tree Model Prediction Results. In the algorithm of the decision tree model, the tree model can use the missing value to calculate the

missing value direction. Even in the face of the missing value, the tree model still has good accuracy, so three models are used, respectively, XGBoost, LightGBM, and CatBoost.

The three tree models are all based on boosting idea. The tree growth strategy of XGBoost is width-first growth, which has the advantage of accurate node segmentation of the model, but the model often grows nodes that are not beneficial to the accuracy of the model. LightGBM's tree growth strategy that is depth-first, which has the advantage of fast model speed and will not grow for nodes that do not benefit much from model accuracy. CatBoost's tree growth strategy is to generate symmetric binary trees. The model is insensitive to parameter adjustment and has the highest robustness, but consumes the most space.

According to the model accuracy, as shown in Table 4, the data accuracy of the LightGBM model is the best. It can be seen that even subsets 2 and 4 have certain effects, and they are not useless data that completely lose characteristic expression.

In order to verify the improved effect of adding CNN channel, the data distribution without CNN channel is shown in Figure 5.

Figure 5 shows the accuracy broken line graph and numerical distribution graph of prediction results of tree model. Three broken lines of Figure 5(a) represent different tree models, and Figures 5(b)–5(d) represent different distribution maps of four subsets, respectively. This step is to highlight that different subsets corresponding to different models have different model results, to show that the discrimination degree of feature selection method is effective, and to show that

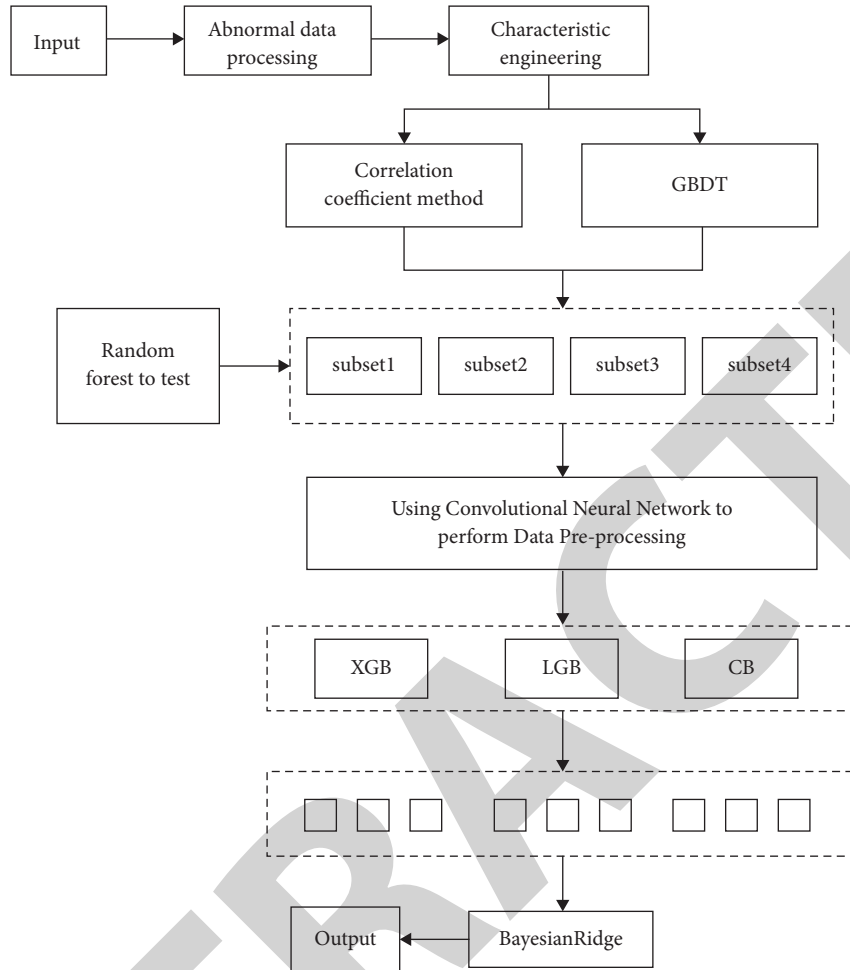


FIGURE 3: Model flow diagram.

TABLE 3: Convolutional neural network parameters.

Model	Input layer	Hidden layer	Output layer	Learning rate	Number of times
CNN	$(k \times m)$; sigmoid	$2 \times, 2 \times \text{conv}; 2 \times \text{MaxPooling}$	1; sigmoid	0.01	50

the model can learn different advantages and disadvantages of data according to different feature spaces.

- (1) As shown in Figure 5, the distribution of the LGB model has the best acceptability, but the prediction score of the LGB model is mainly in [50, 100], and there is no uniform distribution of models such as XGB and CB; also, the high frequency ratio is too high. However, the advantage of the LGB model is that its data distribution is relatively close to the real distribution, and the companies with high scores are relatively concentrated, while the scores of small and medium-sized enterprises with low scores are relatively scattered.
- (2) XGB is in a state of serious two-level differentiation. The advantage of XGBoost is the data with high scores are concentrated at about 80 points instead of 90. The area of [60, 100] is close to the normal distribution; that is, XGBoost can be well-

distributed A company with a predicted score of [60, 100].

- (3) CB is also in a state of serious polarization. CatBoost has the same advantages in subsets 1 and 3 as XGBoost, but the distribution of subset 3 and subset 4 is similar to that of neural network, and the data are very discrete.

4.3. After Adding the CNN Channel—Tree Model Prediction Results. Now add the CNN channel to try to achieve better accuracy. The distribution is shown in Figure 6.

The model fusion used here is to merge the advantages of a single model while discarding the shortcomings, so as to obtain a better model to evaluate the credit score of SMEs.

- (1) After applying the CNN channel, it can be found that the distribution is significantly more concentrated

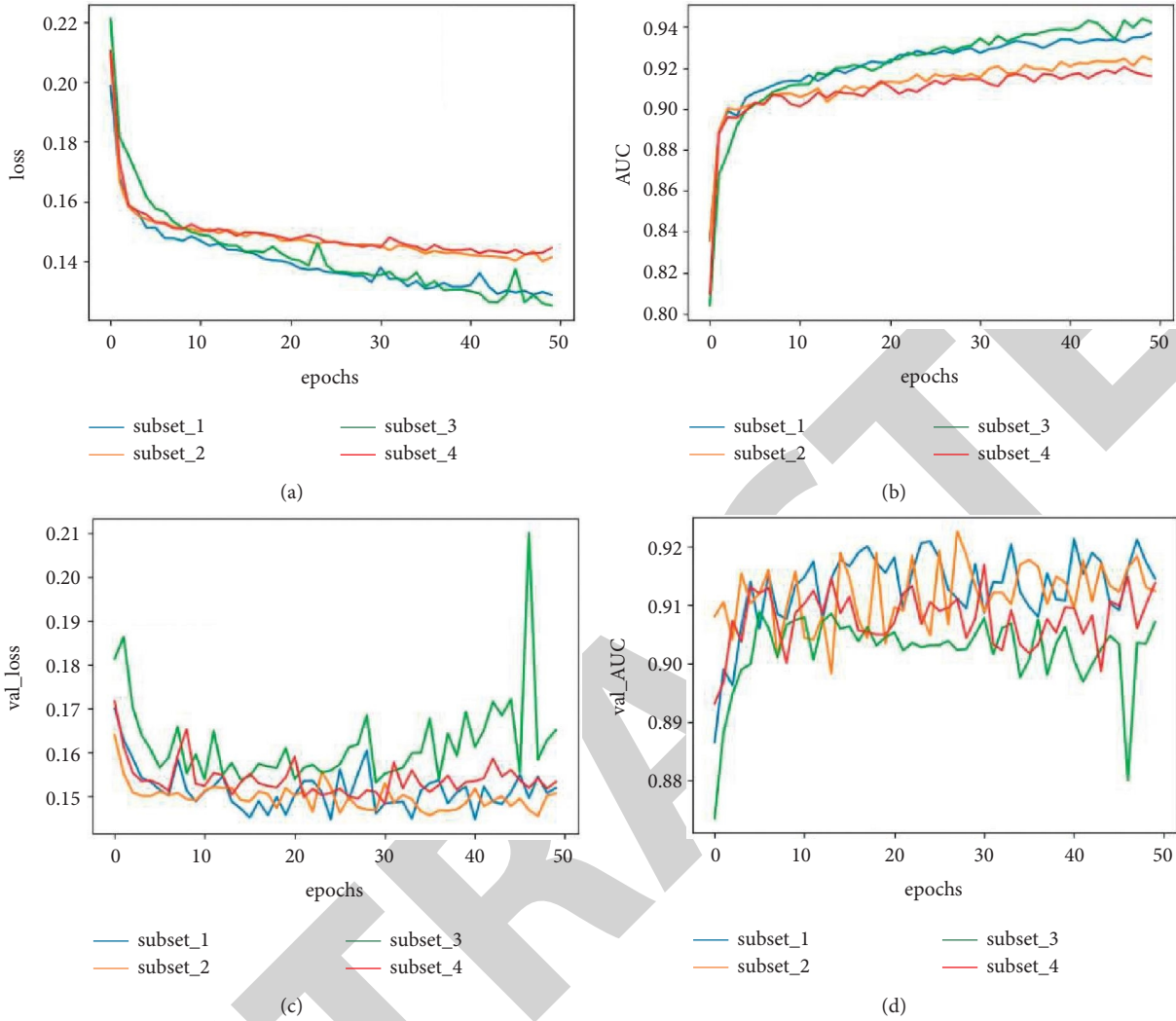


FIGURE 4: Convolutional neural network learning process. (a) Loss. (b) AUC. (c) Val_loss. (d) Val_AUC.

TABLE 4: Tree model accuracy.

Accuracy	Metadata (%)	Sub 1 (%)	Sub 2 (%)	Sub 3 (%)	Sub 4 (%)
LGB	93.723	92.933	91.552	92.310	92.091
XGB	93.225	92.703	92.625	92.760	92.493
CB	92.416	86.295	89.814	92.359	83.787
<i>With CNN channel</i>					
CNN-LGB	92.649	91.896	90.907	91.146	92.325
CNN-XGB	92.188	92.328	92.384	92.570	91.858
CNN-CB	92.334	92.040	88.464	91.960	86.718

and more stable, and the distribution of prediction is more reasonable.

- (2) For the utilization of metadata, it can be seen that the distribution of each model of metadata is closer to each model of subset 1. Although other local details on the data set cannot be highlighted, the overall distribution is taken good care of relatively.

4.4. Credit Assessment Model Integration

4.4.1. Stacking Ensemble, Multi-Model Comparison. Stacking absorbs the idea of neural networks. In a neural network, n units of the last hidden layer of the neural network are mapped to a unit on the output layer. The activation function may be sigmoid or ReLU. In stacking, n neural network units correspond to n base models (such as NN, LightGBM, XGBoost, and CatBoost), where the activation functions correspond to linear regression, logistic regression, or Bayesian ridge. And finally it outputs the result.

Stacking algorithm framework is as follows.

- (1) Select the base model, for example, XGBoost, LGB, random forest, SVM, KNN and other basic algorithm models, assuming that n base models are used.
- (2) Use K fold to divide the training set into m folds, and they are marked as Train_1 to Train_m .
- (3) Suppose training the k -th model. Start from Train_1 , and divide this into a validation set. Use Train_2 to

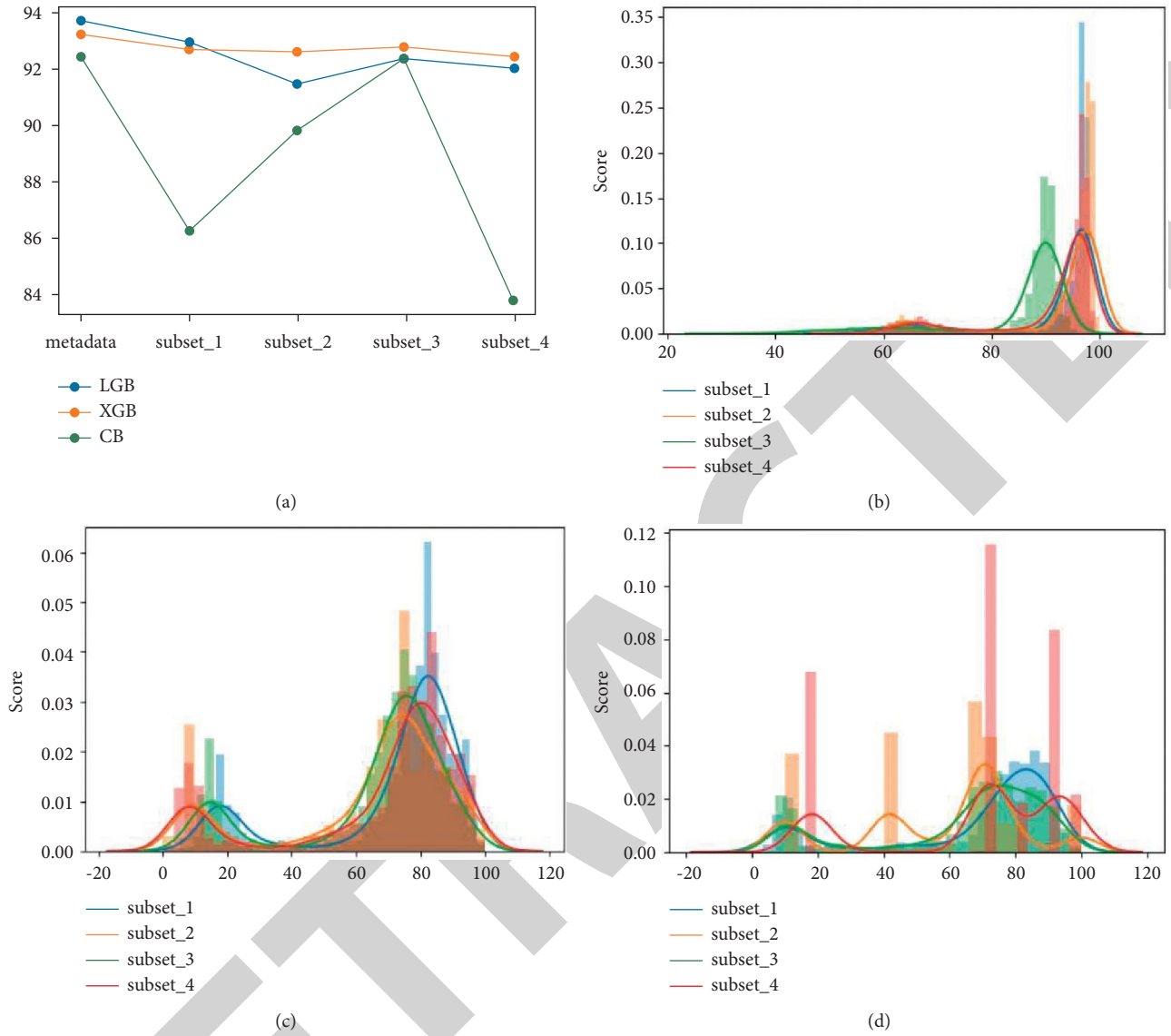


FIGURE 5: Distribution of tree models without CNN channels. (a) Tree model accuracy. (b) LGB result distribution. (c) XGB result distribution. (d) CB result distribution.

Train_m to train the model, then predict Train₁, and keep the prediction results. Then, take Train₂ as the validation set, use Train₁ and Train₃ to Train_m for modeling, and predict Train₂, then keep the results, and so on, until all Train₁ to Train_m are predicted once the establishment of the k-th base model is completed.

- (4) In the base model established in step 3, each model separately predicts the data set and retains the results, and then averages the prediction results m times as the base model prediction result of the k-th column.
- (5) Select the $(k + 1)$ -th base model and repeat the above operations from equations (2)–(4). Get the result of the $(k + 1)$ -th model until the training of n base models is completed.
- (6) If there are n base models, n columns of new feature expressions will be generated, that is, n single model

prediction results. Similarly, the prediction data set also has n columns of new feature expressions. There are totally $n \times m$ prediction results.

- (7) The above 6 steps are used as the input unit of stacking, and the set model is used to predict and output all the units of the first layer.

In this article, NN, LGB, XGB, and CB are the base model of stacking, and K fold has 5 fold. The selection of output model in stacking is shown in Table 5.

According to the results in Table 5, the accuracy of Bayesian regression is generally higher than that of linear regression or logistic regression. Therefore, the stacking output model chosen here is Bayesian ridge regression instead of linear regression or logistic regression. The distribution of the final model is shown in Figure 7.

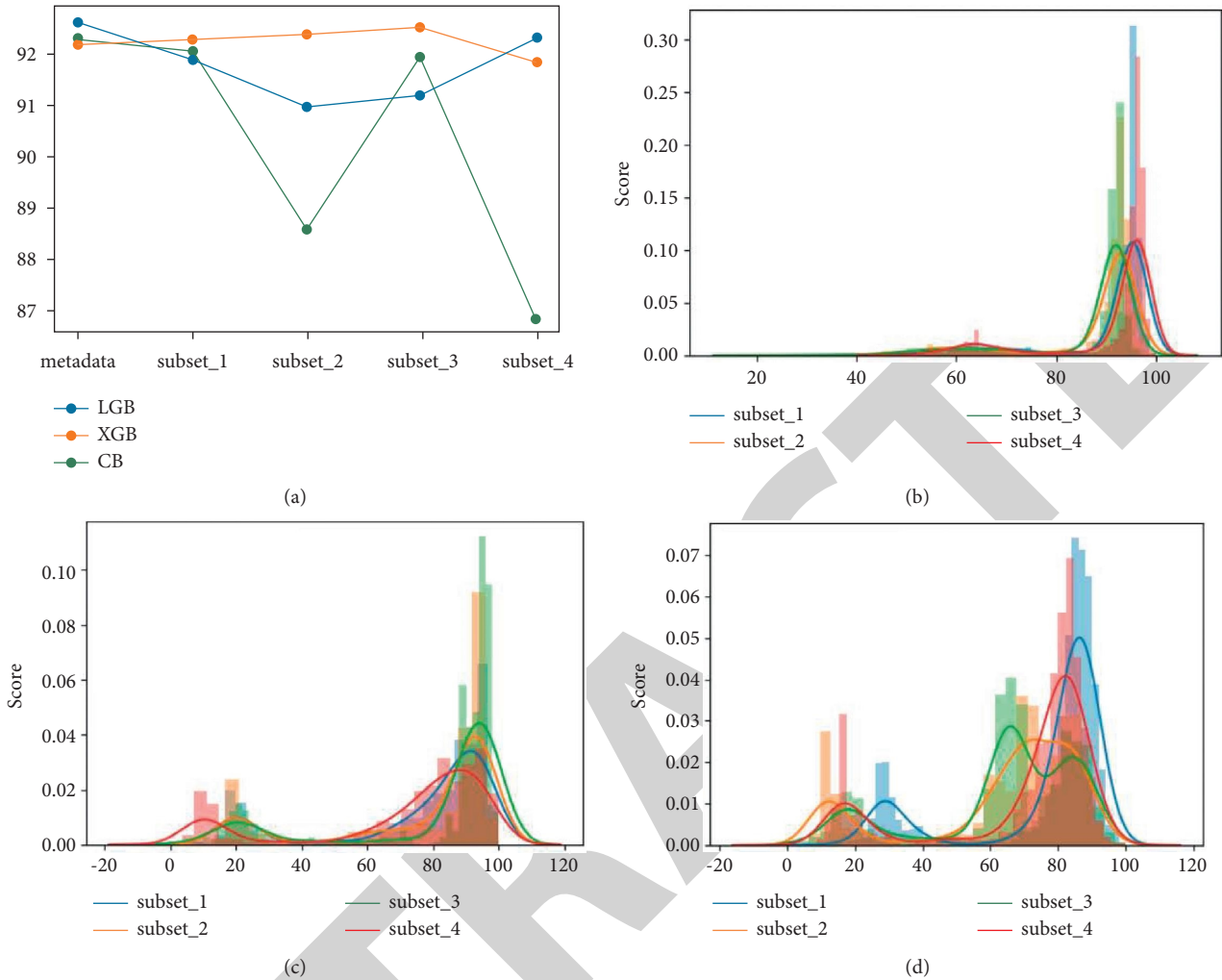


FIGURE 6: Distribution of tree models with CNN channels. (a) Tree model accuracy. (b) CNN-LGB result distribution. (c) CNN-XGB result distribution. (d) CNN-CB result distribution.

TABLE 5: Stacking model accuracy.

Stacking output	Linear regression (%)	Logistic regression (%)	Bayesian ridge (%)
Metadata	95.245	69.477	95.267
Subset 1	94.831	67.437	94.838
Subset 2	93.813	59.437	93.812
Subset 3	96.500	78.353	96.503
Subset 4	93.675	59.138	93.675
All subsets	96.746	77.884	96.749
With CNN channel	97.122	77.221	97.130

It can be seen from Figure 7 that stacking improves the problem that the distribution range of the LGB single model is [50, 100], there are no data that are too concentrated around 90 points, and it will not be like LGB model that the evaluation scores are too concentrated around 95. Stacking makes the distribution of the model more uniform and more close to the actual distribution of business forecasts.

Moreover, the distribution of the model fused by four subsets is better than that of the model fused by metadata, and the accuracy of the verification set reaches 96.749%.

After applying the CNN channel, the accuracy is further improved to 97.130%, which is a great increase over metadata. Every 1% increase up to 95% accuracy is difficult. This shows that although the model is divided into four subsets, the accuracy of a single model is not good enough because a different feature space has different targeted regions; that is to say, the divided subsets can highlight local details instead of global accuracy. Using the stacking method, the advantages of the model trained on the subset can be well combined and disadvantages can be discarded, which

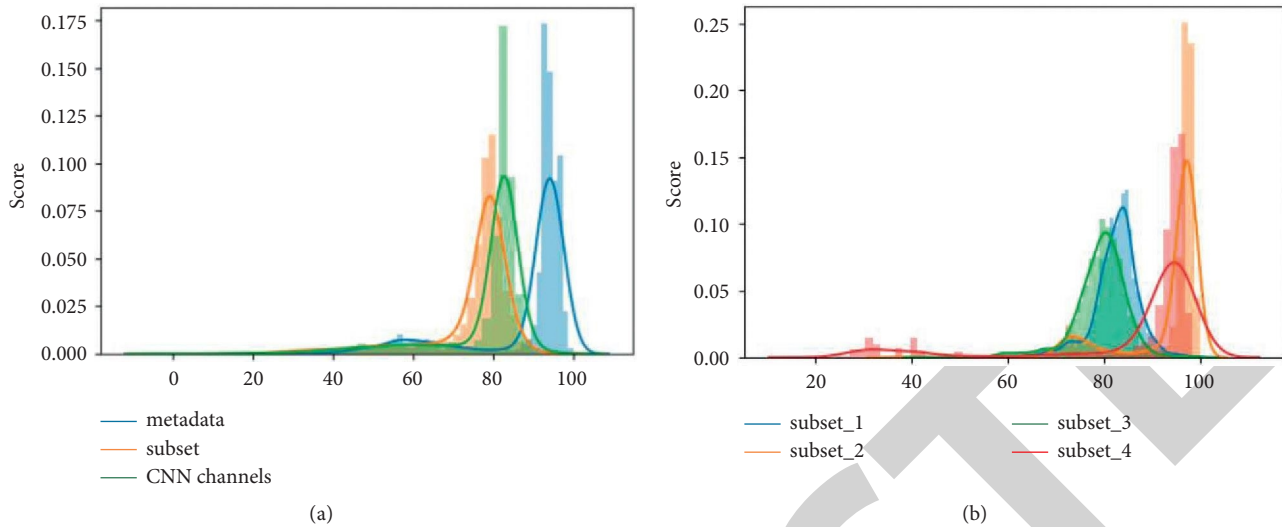


FIGURE 7: Weighted fusion comparison. (a) Weighted metadata. (b) Weighted subset.

contributes to an excellent result. This is the so-called removing the dross and choosing the essence. It is believed that the prediction effect of enterprise credit evaluation based on the integration of multiple models is good, and it provides a good reference for enterprise credit evaluation.

5. Conclusion

The data used in this experiment contain the data of 14366 enterprises. Each data item records the features of 274 dimensions including enterprise type, registration authority, enterprise status, total investment, registered capital, industry code, industry category, enterprise category, value-added tax, enterprise income tax, and stamp tax. The credit evaluation of SMEs has very important practical value and theoretical significance for commercial banks and credit institutions.

Existing researches often screen out subsets with good quality based on related theoretical results and then conduct training and verification on a single model. Although it can play a certain role, it does not take into account the differences among the individual models and the advantages of the models. This paper uses the method based on numerical selection and model selection to construct the subset, and makes full use of the insufficient data, then puts them in another subset, and verifies all the subsets with the base model respectively, and finally merges all the models to obtain very accurate high SME credit evaluation results.

Compared with existing related researches, this article has the following characteristics: observe the distribution of data; find outliers in the data; use metadata set data as much as possible; carry out feature engineering; and analyze the relationship among data. Simulate as much as possible the situation of credit evaluation in real enterprises, avoid data redundancy, and use separate metadata sets. Because data redundancy is not completely redundant, in order to mine the data features as many as possible, the correlation

coefficient and GBDT are used to divide the data, and then the random forest test model is used. Encode the data so that it conforms to the input mode of the convolutional neural network. Through the convolutional neural network, find the relationship among the data. The four subsets are input into four different convolutional neural networks to obtain four coarse prediction channels. Aiming at the 12 training results of all models, the Bayesian regression is used to fuse the training results of single models, and a particularly high precision result is obtained, with an accurate value of 97.13%.

From the point of empirical results, the effect of multimodel fusion model established by this paper is better, compared with other algorithms and traditional methods, it shows that it can help more creditworthy enterprises to more effectively and more fairly get loans, and it can better help bank enterprise risk assessment, reduce the rate of bad debts, and promote whole credit business benign development thereby.

The simulation results show that the model has achieved good results in SMEs credit risk assessment, improving the efficiency and accuracy of assessment, and can make accurate and reliable assessment in SMEs financing, loan, and other scenarios. It is of great significance to the financing, risk management, and financial service supervision of small and medium-sized enterprises. Its application prospect is very broad, and it has certain practical application significance and theoretical research value.

Data Availability

The data used to support the findings of this study can be obtained from the corresponding author (zhangleicqjtu@163.com).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was partially supported by Group Building Scientific Innovation Project for Universities in Chongqing (CXQT21021), Science and technology research project of Chongqing Education Commission (KJQN202100712), and Joint Training Base Construction Project for Graduate Students in Chongqing (JDLHPYJD2021016).

References

- [1] Y. D. Liu, Y. N. Song, and J. Y. Yu, "Big-data-driven model construction and empirical analysis of SMEs," *Credit Assessment in China Procedia computer science*, vol. 147, pp. 613–619, 2019.
- [2] N. Yoshino and F. Taghizadeh-Hesary, "Optimal credit guarantee ratio for small and medium-sized enterprises' financing: evidence from Asia," *Economic Analysis and Policy*, vol. 62, pp. 342–356, 2019.
- [3] X. Hu, J. Hu, and L. Chen, "Credit risk assessment model for small, medium and micro enterprises based on RS-PSO-SVM integration[C]," *Proceedings of the 2021 IEEE 6th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA)*, pp. 342–345, IEEE, Chengdu, China, 2021.
- [4] T. Gregori, S. Montesor, and S. P. Rossi, "External financing of innovative small and medium enterprises (SMEs): unpacking bank credit with respect to innovation typologies and combinations," *Industrial and Corporate Change*, vol. 31, no. 1, pp. 234–267, 2022.
- [5] J. Zhang, "Investment risk model based on intelligent fuzzy neural network and VAR," *Journal of Computational and Applied Mathematics*, vol. 371, Article ID 112707, 2020.
- [6] X. Cai, Y. Qian, Q. Bai, and W. Liu, "Exploration on the financing risks of enterprise supply chain using Back Propagation neural network," *Journal of Computational and Applied Mathematics*, vol. 367, Article ID 112457, 2020.
- [7] X. Huang, X. Liu, and Y. Ren, "Enterprise credit risk evaluation based on neural network algorithm," *Cognitive Systems Research*, vol. 52, pp. 317–324, 2018.
- [8] D. P. Fonseca, P. F. Wanke, and H. L. Correa, "A two-stage fuzzy neural approach for credit risk assessment in a brazilian credit card company," *Applied Soft Computing*, vol. 92, Article ID 106329, 2020.
- [9] H. Li, Y. Cao, S. Li, J. Zhao, and Y. Sun, "XGBoost model and its application to personal credit evaluation," *IEEE Intelligent Systems*, vol. 35, no. 3, pp. 52–61, 2020.
- [10] L. Zhou, B. Liu, Y. Zhao, and Z. Jiang, "Application research of grey fuzzy evaluation method in enterprise product reputation evaluation," *Procedia CIRP*, vol. 83, pp. 759–766, 2019.
- [11] R. C. Zhang, Y.-H. Zhang, and H. Zhao, "Quantitative Comparison of the effects of bank credit evaluation Models," *Financial Supervision research*, vol. 1, pp. 66–85, 2021.
- [12] S. Moradi and F. Mokhatab Rafiei, "A dynamic credit risk assessment model with data mining techniques: evidence from Iranian banks," *Financial Innovation*, vol. 5, no. 1, pp. 15–27, 2019.
- [13] J. Sun, J. Lang, H. Fujita, and H. Li, "Imbalanced enterprise credit evaluation with DTE-SBD: decision tree ensemble based on SMOTE and bagging with differentiated sampling rates," *Information Sciences*, vol. 425, pp. 76–91, 2018.
- [14] J. Fan, S. Upadhye, and A. Worster, "Understanding receiver operating characteristic (ROC) curves," *Canadian Journal of Emergency Medicine*, vol. 8, no. 01, pp. 19–20, 2006.
- [15] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.
- [16] M. Mohammady, H. R. Pourghasemi, and M. Amiri, "Land subsidence susceptibility assessment using random forest machine learning algorithm," *Environmental Earth Sciences*, vol. 78, no. 16, pp. 503–512, 2019.
- [17] N. Arora and P. D. Kaur, "A Bolasso based consistent feature selection enabled random forest classification algorithm: an application to credit risk assessment," *Applied Soft Computing*, vol. 86, Article ID 105936, 2020.
- [18] D. Cheng, Z. Niu, and L. Zhang, "Delinquent events prediction in temporal networked-guarantee loans," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–13, Article ID 33048760, 2020.
- [19] L. Xu, L. Cui, T. Weise et al., "Semi-supervised multi-Layer convolution kernel learning in credit evaluation," *Pattern Recognition*, vol. 120, Article ID 108125, 2021.
- [20] L. Zhang and Q. Song, "Multimodel integrated enterprise credit evaluation method based on attention mechanism," *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–12, 2022.
- [21] L. Zhang and Q. Song, "Credit Evaluation of SMEs Based on GBDT-CNN-LR Hybrid Integrated Model Wireless Communications and Mobile Computing," *Service-Oriented Management and Computing in Edge-Cloud IoT*, vol. 2022, Article ID 5251228, 2022.
- [22] T. Chen and C. X. Guestrin, "A scalable tree boosting system," in *Proceedings of the DSLeBdings of the 22nd AcM Sigkdd International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, ACM, New York, NY, USA, 2016.