WILEY | Hindawi

Research Article

Risk Assessment of Biological Asset Mortgage Loans of China's New Agricultural Business Entities

Shuzhen Zhu (iD)**, Yutao Chen** (iD)**, and Wenwen Wang** (iD)

Glorious Sun School of Business and Management, Donghua University, Shanghai 20051, China

Correspondence should be addressed to Yutao Chen; 1054580867@qq.com

The large-scale proliferation of China's new type of agricultural entities has given rise to a higher demand for funds. Farmers have insufficient effective collateral, which makes it difficult for them to obtain sufficient loans. Chinese financial institutions have developed a biological asset mortgage loan business to cope with this situation. China has not considered biological mortgages but has been using real estate and asset mortgage models with strong realizability. This innovative financial business has achieved positive results since it was attempted, but it also faces many risks. It is very important to comprehensively and accurately consider the risk factors of biological asset mortgage loans. Based on 1249 production and operation data samples of new agricultural entities in Zhejiang, Henan, and Shandong provinces, this study constructs an XGBoost model for empirical analysis and compares it with logical regression, support vector machine, and random forest algorithms to obtain the optimal model and feature importance value. According to the characteristic importance value, a biological asset mortgage loan risk assessment system with 4 primary indicators and 20 secondary indicators is established, which can effectively identify the biological asset mortgage loan risk of new agricultural entities.

1. Introduction

As China's agricultural production increases in scale, intensification, and specialization, the process of agricultural modernization is gradually accelerating, and the ranks of new agricultural management entities are gradually expanding. Large-scale production requires large amounts of manpower and material resources; consequently, the demand for funds is also gradually increasing. In China, the new type of agricultural management body is defined as an agricultural management organization with a relatively large scale of operation, good material and equipment conditions and management capabilities, high labor production, resource utilization, and land production rate, and includes commercial production as its main objective. According the People's Bank of China, by the end of March 2019, the balance of various RMB (RMB: Renminbi, currency in circulation in China) loans from financial institutions was 142.11 trillion yuan, up 17.3% year-on-year. The balance of loans from agriculture, forestry, animal husbandry, fishery,

and agriculture in China was 14.68 trillion yuan, up 3.04% from the beginning of the year. Financial institutions' loans to agriculture accounted for 23.72% of total loans, of which loans from pure agriculture, forestry, animal husbandry, and fishery accounted for only 9.68%; this is far from the actual demand for loans from these sectors.

In order to solve the issue of limited funds of new agricultural entities, financial institutions in areas with more developed agricultural industrialization have started to experiment with the biological asset mortgage loan business. Some biological assets have become effective as collateral assets, because of their good liquidity. However, agriculture is a weak industry, and its production and operations are facing great uncertainty, which makes it easier to be affected by natural disasters. Moreover, China's agriculture is undergoing a transition from a small-scale peasant economy to an intensive one. Under the environment of a small-scale peasant economy, the personal credit quality of some farmers is uneven. Coupled with serious information asymmetry and imperfect laws and regulations related to

biological asset mortgage loans, there are numerous potential risks involved in the implementation of these loans. Only some agricultural or financially developed areas carry out this form of business.

This study examines the risk factors involved with biological asset mortgage loans and the main body of new agricultural management, in order to build a risk assessment system to reasonably assess the risk of biological asset mortgage loans. As the magnitude and nature of Chinese agriculture are different from those of other countries, there is less research on establishing a risk assessment system for biological asset mortgage loans in at present, and Chinese scholars often use analytic hierarchy process and fuzzy comprehensive evaluation methods for analysis. These methods are subjective and rely on experts' knowledge and experience, and some risk factors are less available and difficult to apply in practice. Therefore, this study obtains sample data through a questionnaire survey and uses mainstream machine learning methods such as XGBoost, random forest (RF), support vector machine (SVM), and logical regression (LR) to carry out empirical research. According to the value of characteristic importance, this study builds a risk assessment system for biological asset mortgage loans of new agricultural business entities, provides reference for financial institutions to carry out biological asset mortgage loans, and presents several suggestions to promote the development of this innovative financial business and improve the current shortage of funds for new agricultural business entities.

1.1. Literature Review and Related Research. After 5 years of development, China's new agricultural business entities have achieved great progress; however, the overall level of development is low, and there are large differences across regions, which is generally consistent with the development degree of regional economy [1]. Capital, land circulation, market, and talents are the main restricting factors in the development process of new types of agricultural business entities [2–4]. In order to promote the development of new agricultural business entities, financial institutions should innovate financial products and service methods and actively carry out pilot projects such as forest right mortgage, land contractual management right mortgage, and biological asset mortgage loan, in order to activate farmers' assets and form a multichannel financial support system [5, 6].

There are usually basic and agricultural-related parameters in the analysis of agricultural basic credit parameters. The traditional basic credit parameters are still used in general low-level financial institutions and financial underdeveloped countries [7, 8]. Dobbie et al. [9] studied financial and labor market impacts of bad credit reports. They think labor is the critical factor. Sossou [10] found the evidence that farm revenue is positively correlated with land acreage, quantity of labor, and costs of fertilizers and insecticides. In Nigeria, the authors in [11] classified farmers into beneficiaries and nonbeneficiaries. These studies are relatively perfect, but the rapid development of agricultural technology, so that some evaluation criteria have not been effectively analyzed.

Biological asset mortgage loans, a new financial product, has been widely studied by scholars, including their feasibility and necessity, the influencing factors of biological asset mortgage loan willingness, biological asset value evaluation, and biological asset mortgage loan risk research. Biological assets have a clear property right relationship and market value, and do not need to be transferred during the mortgage period, thus resulting in good feasibility [12, 13]. American banks and regulatory agencies think that animal husbandry is the most appropriate collateral [14]. China adopted biological asset mortgage loans relatively late, and only a few regions have attempted it. The research shows that parameters, such as age, total family population, loan experience, production scale, policy cognition, and mortgage financing significance cognition, have significant influence on farmers' willingness to obtain mortgage loans [15, 16]. The value evaluation of biological assets is a necessary process for the implementation of the biological asset mortgage loan business. However, the value of biological assets has always undergone dynamic changes. In addition, the infrastructure of biological asset value evaluation in China is poor, which makes the evaluation more difficult. At present, the international authoritative evaluation methods include the market method, income method, and cost method. Wenjuan et al. [17] compared and analyzed various evaluation methods and considered that the valuation methods of biological assets of different types and growth stages are different. Among crop assets, flower assets should be evaluated through the market method, while other crop assets should be evaluated through the cost method; forest assets with trading market should be evaluated through the market method, as economic forests will bring continuous income. The income method should be used to evaluate the benefits, and the cost method should be used for the evaluation of the young forest assets; the market method should be used to evaluate the livestock assets within an active trading market, and the income method should be used for biological assets that are produced repeatedly. The cost method should be used for the fishery biological assets that are attached to the river water resources and do not exist in the active trading market. However, in different countries and different environments, the scheme of biological asset mortgage needs to be further analyzed and formulated in different regions.

There are many basic approaches to credit risk. Baltensperger [18] first proposed the theory of "credit rationing." The theory holds that compared with applicants with lower credit rating, commercial banks tend to choose applicants with higher credit rating in order to avoid risks and achieve their own business objectives. Altman et al. [19] believe that the logistic model can achieve the best effect in measuring the credit risk of SMEs (small and medium-sized enterprises). Weisen et al. [20] established a credit default risk assessment model for agricultural small and medium-sized enterprises by using the method of logistic combined with factor analysis, and found that the prediction effect of the model is good and has universality. Samad [21] used the probit model to analyze the reasons for the failure of banks, and conducted discriminant analysis on risk factors. It was

proved that the model established by probit can achieve 80% prediction accuracy. Hsieh and Lee [22] used the benchmark model to measure the internal risk of banks. Kamalloo [23] proposed a classifier using immune principles and fuzzy rules to predict quality factors of individuals in banks and other financial institutions. These methods are also commonly used by banks and other financial institutions.

There are few studies on the risk assessment of biological asset mortgage loans, and most of them use an analytic hierarchy process, combined with a fuzzy comprehensive evaluation method [24], to build a risk assessment index system and set weights for the indicators. This type of method is more subjective. Teles [25] compares the SVM and RF algorithms to forecast the recovered value in a credit task. The execution of the projected intelligent systems uses tests and algorithms for authentication of the projected model. Westland [26] developed and tested machine learning models to predict significant credit card fraud in corporate systems using Sarbanes-Oxley (SOX) reports, news reports of breaches, and Fama-French (FF) risk factors. A biological asset mortgage loan risk assessment is similar to a credit risk assessment, which is a binary classification problem. At present, the mainstream method for the study of a dichotomy problem is the machine learning method. Both traditional machine learning methods (such as logical regression and decision tree) or frontier machine learning methods (such as random forest and XGBoost model) can show excellent early warning effects [27–29]. The XGBoost algorithm, with its high classification accuracy, has become the mainstream classification prediction method, and is widely used in various fields such as forecasting and credit rating [30–33]. For unbalanced samples, the XGBoost algorithm can also show better classification performance, which is better than other algorithms.

Considering the imbalance of sample data, this study uses the XGBoost algorithm to fit the possible risk points of biological asset mortgage loans, and compares the classification performance with other algorithms. This study further establishes an evaluation system that can accurately identify the risk of biological asset mortgage loans of new agricultural business entities.

2. Index System Establishment and Research Methods

2.1. Initial Selection of an Indicator System. Considering the risk factors involved before, during, and after the establishment of China's new agricultural business entities, combined with the characteristics of biological assets, and using the 5C factor analysis method of credit analysis, the risks are divided into four categories. The categories are as follows: personal credit risk, specific indicators included in operating risks, biological assets' own risks, and policies and market risks. They are shown in Table 1.

The first category covers the individual credit risk of farmers. The basic characteristics of farmers are used to measure credit risk, and the availability of data is considered. Finally, five secondary indicators are selected to measure gender, age, health status, education level, and asset-liability ratio. The older the participants, the lower is their risk-taking ability. The more educated the participants, the higher is their willingness to repay their credit. The asset-liability ratio measures the household's repayment ability; the lower the debt to asset ratio, the higher is the repayment ability.

The second category covers the operational risk of farmers. The production and operation conditions determine the output level of biological assets. The output level of biological assets directly affects the income level, which has an impact on the repayment ability and increases the risks of the biological asset mortgage loans. The first-level indicators of business conditions include the following: years of employment, guarantees for others, foreign investment, professional and technical personnel, mechanical automation equipment ownership, simple electronic management, stable and reliable sales channels, stable and high-quality channels of purchase of production materials, grasp of market conditions, agricultural insurance purchases, sales profit margins, and land circulation years.

The third category covers the risk of biological assets, including the difficulty of assessing the value of biological assets, the difficulty of realizing biological assets, the degree of loss caused by natural disasters, and whether the products are high-quality varieties. If the value of biological assets is difficult to determine, the accuracy of the valuation could be reduced; the value of the loans obtained by farmers then exceeds the value of their biological assets, and financial institutions are therefore faced with greater risks. The liquidity of biological assets and the quality of the varieties determine the ability of the product to generate income via sales which will have an impact on the farmers' repayment ability. Natural disasters will affect the growth of biological assets, thus affecting the farmers' income.

The fourth category is policy and market risk, including the range of price fluctuation in the biological asset market, the number of regional asset assessment institutions, and environmental policy risk. The stricter the government's environmental protection requirements, the greater the risk of environmental protection policies affecting farmers. Market conditions are constantly changing due to the influence of factors such as supply and demand, the product cycle, and natural factors; the greater the fluctuation of market prices, the higher is the risk of mortgage loans. The number of institutions in each region determines their asset valuation environment. The better the asset evaluation environment, the more reasonable is the valuation of biological assets, and the lower is the risk associated with biological asset mortgage loans.

2.2. Index System Screening. Through onsite investigations of new agricultural management entities in Zhejiang Province and visiting their production sites, the established index system is screened. Although the term of the land transfer is related to the continuity of operation, the duration of the land transfer contracts signed by the interviewed farmers is generally longer, with them signing these contracts once every few years. Moreover, farmers think that there is almost no sudden interruption of the contract, which will not lead

TABLE 1: Preliminary selection of the risk assessment index system of biological asset mortgage loans.

Primary indicators	Secondary indicators
Basic information relating to the subject	Gender, age, education level, asset-liability ratio, and health status
Operating risk	Working years, guarantee for others, foreign investment (part-time), technical personnel, mechanical automation equipment, sales profit margin, stable sales channels, stable source of high-quality means of production, whether there is simple electronic management, the degree of market control, whether to buy agricultural insurance, the number of years of land transfer (In China, land is not owned by individuals. Farmers can get the right to use land from the state. The right of use can be transferred to the farmers through the circulation mode (similar to leasing). However, farmers have the right to recycle), and the expenditure of land transfer
Risks of biological assets	Whether the products are high-quality products, the degree of loss caused by natural risks, whether the value of biological assets is easy to evaluate, and the degree of difficulty in realizing biological assets
Policy and market risk	The fluctuation range of the product market price, the soundness of asset appraisal institutions, and whether there are environmental risks

to production interruption. Therefore, this paper will delete the index of "land circulation period"; land transfer expenditure causes production and operation costs. In the survey, it is found that the expenditure of land circulation per mu in almost the same area is the same, basically maintained between 500 and 800 yuan/mu (mu, a Chinese unit of area (=0.0667 hectares), which is not comparable, and the sales profit rate index has already reflected the cost situation. Therefore, the indicator of "land transfer expenditure" is deleted; the index of "the degree of loss caused by natural risk" can reflect the degree of loss caused by natural disasters, but the research shows that the cost of land circulation is not the same. It is found that vegetables, fruit trees, and other biological assets are planted in greenhouses and are not affected by bad weather. This indicator is difficult to measure and is therefore deleted; in addition, it is difficult to obtain true information relating to health status due to some farmers providing false information; therefore, this indicator is deleted after comprehensive consideration.

The continuous variables were tested using a person test, and the redundant indexes were deleted. There are four continuous variables in the above indicators: age, years of employment in the current industry, profit margin of sales, and asset-liability ratio. After testing the above four continuous variables, it was found that the correlation is not obvious. The correlation coefficient between age and current industry working years is 0.41, the value is less than 0.5, and thus there is no obvious correlation between continuous variables. After screening, a two-tier index system with 21 risk factors is obtained, as shown in Table 2.

Using the experience of previous scholars to measure the risk of biological asset mortgage loans, where farmers have obtained loans from the bank in the past, or there are still outstanding loans, it shows that the bank has conducted an in-depth investigation into these farmers and determined that they have the ability to repay the loans. When these farmers apply for biological asset mortgage loans, the bank determines the farmers' biology. The risk of a mortgage loan is small; if farmers have not successfully obtained loans or default, it is considered that there is a greater risk of biological asset mortgage loans. Some may not have obtained loans successfully, although the banks have mastered

sufficient information in terms of the main applicant because their production and operation levels and other basic conditions do not meet the requirements of bank lending, as there will be greater risks when applying for biological asset mortgage loans.

2.3. Research Method. There are many related research results on risk assessment methods, such as using the analytic hierarchy process (AHP) and fuzzy comprehensive evaluation methods to set the weight of risk indicators, so as to comprehensively evaluate the risk size, probit regression, and widely used machine learning methods such as support vector machine, random forest, gradient boosting decision tree (GBDT), logistic regression, and XGBoost (extreme gradient boosting). However, existing scholars only use the analytic hierarchy process (AHP) and fuzzy comprehensive evaluation method to evaluate the risk of biological asset mortgage loans. After considering the advantages and disadvantages of various methods, this study uses the current mainstream classification method, namely, XGBoost. The algorithm evaluates the risk of biological asset mortgage loans and compares it with the classification effect of the support vector machine, random forest, and logical regression, and finally obtains an index system that can effectively evaluate the risk of biological asset mortgage loans.

The XGBoost algorithm, proposed by Dr. Chen Tianqi in 2016, is an integrated learning method based on GBDT. It not only retains the original characteristics of GBDT but also greatly improves the training speed and prediction accuracy of the model through a lot of optimization work. It has achieved good results in terms of recommendation, search ranking, user behavior prediction, click-through rate prediction, and product classification. The working principle of the XGBoost algorithm is to establish K cart classification trees. In the running process, the boosting ensemble learning method is used to fit the error of the previous tree with the next tree to reduce the gap between the real value and the predicted value, and shape the model to have the ability to generalize as much as possible. The main feature is to carry out a parallel operation by using a Taylor expansion of the objective function and adding a penalty term, that is, to limit

TABLE 2: Variable definition and descriptive statistics.

Primary indicators	Secondary indicators	Index description
Subject's personal credit risk	Age	Continuous variable
	Gender	Male = 1, female = 0
	Education level	Junior high school and below = 1, high school and technical secondary school = 2, junior college and above = 3
Main operating risk	Assets and liabilities	Continuous variable
	Whether to buy agricultural insurance	Yes = 0, no = 1
	Degree of grasp of market conditions	Understand = 1, sure = 2, more sure = 3
	Stable and reliable sales channels	With = 0, without = 1
	Stable and high-quality production material purchase channels	With = 0, without = 1
	Is there simple electronic management?	With = 0, without = 1
	Whether there are professional and technical personnel worker	With = 0, without = 1
	Revenues	Continuous variable
	Whether to guarantee others	Yes = 0, no = 1
	Years of agricultural employment	Continuous variable
	Invest outward	Yes = 0, no = 1
	Is there mechanized automatic equipment?	With = 0, without = 1
Risks of biological assets	Difficulty of realization of biological assets	Easier = 1, medium = 2, hard = 3
	Is the product of high-quality?	Yes = 0, no = 1
	Is the value of biological assets easy to valuate?	Yes = 0, no = 1
Policy and market risk	Number of regional asset appraisal institutions	Less = 1, more = 0
	Biological asset market price fluctuation range	Small = 1, medium = 2, big = 3
	Is there any environmental policy risk?	With = 0, without = 1
Result	Is there a greater risk?	Yes = 0, no = 1

the number of leaf nodes to prevent the model from being too complex. The objective function of the model is composed of the loss function and regularization term:

$$\text{Obj}^{(t)} = \sum_{i=1}^{n} l\left(y_i, \widehat{y}_i^{(t)}\right) + \sum_{i=1}^{t} \Omega(f_t), \quad (1)$$

where $l(y_i, \widehat{y}_i^{(t)})$ is the difference between the real value and the predicted value and $\Omega(f_t)$ is the regularization term, which limits the number of leaf nodes.

2.4. Data Sources. The samples in this study come from new agricultural business entities in 14 regions in the three provinces of Zhejiang, Shandong, and Henan, including Hangzhou, Ningbo, Wenzhou, Shaoxing, Jiaxing, Taizhou, Jinhua, Huzhou, Quzhou, Lishui, Zhoushan, Zaozhuang, Weifang, and Xinyang. According to the established biological asset mortgage loan risk indicators, the questionnaire was designed and conducted during on-site investigations. Accompanied by relevant personnel of the regional banks, the new agricultural business entities were interviewed, and the questionnaires were issued to obtain first-hand information. The investigation obtained 1498 data points.

3. Analysis and Results

3.1. Data Preprocessing and Data Descriptive Statistical Analysis. The data in this article come from field surveys and contain some problems such as missing and distorted data.

Subtype variables should be numerically processed when using machine learning modeling, and continuous variables should be normalized when using logistic regression and support vector machine modeling to avoid inaccurate results caused by large data values. Therefore, this study performs one-hot encoding on the categorical data and performs missing values, outliers, and normalization at the same time.

3.1.1. Outlier Handling. In terms of continuous variables, it is impossible to judge whether the data are caused by human error or whether the data themself are true; therefore, outliers cannot be deleted directly. An excel filtering operation is used to view the distribution of data and subjectively determine whether to delete or keep unreasonable data. For example, for the sales profit rate feature, the maximum value of the sales profit rate itself is 1; if the data value exceeds 1, there is the possibility of incorrect filling. Deleting such sample data has little effect on the training of the model and can therefore be deleted directly.

3.1.2. Missing Value Processing. In order to ensure the integrity of the data, the missing data need to be processed, and the processing of missing values is determined according to the degree of missing data. This study uses 20% as the limit, deleting data with missing values over 20%, and fills in data with missing values less than or equal to 20%. Among them, continuous data are filled with averages, and subtype data are filled with modes.

3.1.3. One-Hot Encoding.

The working principle of one-hot encoding is to use N-bit status registers to encode N states. Each state has its own independent register bit, and at any time, only one of them is valid. One-hot encoding can, to a certain extent, expand the features to prevent overfitting of the model. If the sample contains categorical data, such as gender, it usually uses 0,1 as a label to mark the value of the variable, where there is no logical relationship in itself. However, when training the model, the computer will default to a numerical value and make logical judgments based on the size of the data. Therefore, in order to facilitate the input of the model, discrete variables need to be quantified into numerical values. This article performs one-hot coding on 17 variables including gender, education level, whether to invest abroad, whether to guarantee others, and whether there are professional and technical personnel. Finally, model training and prediction data can be carried out directly.

3.1.4. Normalization.

Normalization processing, refers to the processing of numerical data to limit it to a certain range, so as to prevent certain characteristics from being "eaten" by big data due to their own expression problems. This results in each feature being treated fairly by the classifier. When using algorithms to test data, the tree-based algorithm model does not need to normalize the data, while the support vector machine and logistic regression need to normalize the data before the experiment is conducted, placing the data on a different dimension. In this study, the maximum-minimum method is used to process the data so that the numerical data are in the interval [0, 1]. The specific formula is as follows:

$$x = \frac{x - x_{min}}{x_{max} - x_{min}}, \tag{2}$$

where X is the specific value of the selected variable, x_{max} is the maximum value that the variable can obtain, and x_{min} is the minimum value that the variable can obtain.

3.1.5. Data Descriptive Statistical Analysis.

After processing the data, a total of 1249 samples for modeling were finally obtained, with 21 feature numbers, including 1,038 samples from Zhejiang Province, 116 from Henan Province, and 95 from Shandong Province. The biological assets involved include field crops, live animals, and economic crops, a total of 39 products. The specific products are listed in Table 3.

Python is used to perform statistics on each feature to obtain the distribution of each feature value, such as the average value, standard deviation, maximum value, and minimum value. See Table 4 for details. From the statistical results, the majority of the sample is male, accounting for about 86% of the total sample; the average age of the new agricultural business entities is 46.8 years old, the oldest in the sample is 67 years old, while the youngest is only 23 years old. Some young people have joined large-scale agricultural productions, but their age is generally high at present. The average education level is 1.6, indicating that the education level of farmers is generally not high, but there are also some high-intellectuals engaged in agricultural production. The

average working life of farmers is 12–13 years. Long-term agricultural production has allowed it to accumulate a lot of experience; the average asset-liability ratio is 36.23%; 28.87% of farmers provide guarantees for others; 21.22% of farmers also invest in other industries while engaged in agriculture; nearly half of them are equipped with professional and technical personnel. As they have been engaged in agriculture for many years, some business entities are more skilled in agricultural production and can be used as professional and technical personnel. Most farmers have introduced mechanized production, and a small number of farmers have simple electronic management. Most entities have a good grasp of market conditions and have more accurate judgments in terms of their environment, resulting in an average sales profit rate of 29%. In order to avoid natural risks, approximately half of the entities have purchased agricultural insurance, and most farmers produce ordinary products. Approximately, 13% of the main products produce good varieties, and the price fluctuations of the products produced are small. The balance of supply and demand can be achieved in the domestic market. Evaluation of the value of some biological assets is difficult. At the same time, the environmental protection policy risk faced by the operating entity, and the risk of production interruption are relatively small.

3.2. Construction of Risk Assessment Model of Biological Asset Mortgage Loan Based on XGBoost Algorithm.

The XGBoost algorithm is used as a modeling tool and compared with the classification performance of LR, SVM, and RF algorithms. After optimizing the parameters of the model, the best model and importance value of each feature of the biological asset mortgage loan risk assessment are obtained.

3.2.1. Model Performance Evaluation Index.

The confusion matrix (Table 5) is the most basic, intuitive, and easiest way to measure the accuracy of model classification. From the confusion matrix results, the accuracy, recall, F1-Score, classification accuracy of positive samples, and classification accuracy of negative samples are obtained. Rate and other evaluation indicators are as follows.

In addition to the results of the confusion matrix, a receiver operating characteristic (ROC) curve and area under the curve (AUC) values are also important evaluation indicators. The ROC curve describes the relationship between the model's TPR and FPR, where TPR is the proportion of the number of positive samples that are correctly classified by the model to the total number of positive samples, and FPR is the number of negative samples that are incorrectly classified by the model to the total number of negative samples. The horizontal axis of the ROC curve represents the FPR, and the vertical axis represents the TPR. For classification problems, one predicts a positive sample probability for each sample and then compares it with the set threshold to determine whether it is a positive sample or a negative sample. The AUC value is the area under the ROC curve, and the value range is [0,1]. The larger the AUC value, the better the model classification effect. The purpose of the

TABLE 3: Classification of sample products.

Livestock products	Fruits	Vegetables	Food crops	Seedlings and flowers
Pig	Pomelo	Toona		Ginkgo
Chicken	Tangerine	Soy		Flowers
Duck	Grape	Eggplant		Bonsai
Beef cattle	Pitaya	Tomato		Succulent plants
Mutton	Loquat	Mushroom		Chrysanthemum
Goose	Peach	Zizania	Rice	Camphor tree
	Pear	Black fungus		Chinese herbal medicine crops
	Persimmon	Lily		Red plum
Mink	Kiwi			Pecans
	Watermelon	Bamboo shoots		
	Fig			Green maple
	Strawberry			

TABLE 4: Data descriptive statistical analysis.

Index	Mean	Standard deviation	Minimum	Maximum
Age	46.82	8.08	23	67
Gender	0.14	0.35	0	1
Education level	1.61	0.75	1	3
Assets and liabilities	0.36	0.65	0	12.5
Whether to buy agricultural insurance	0.54	0.50	0	1
Degree of grasp of market conditions	2.34	0.58	1	3
Stable and reliable sales channels	0.20	0.40	0	1
Stable and high-quality production material purchase channels	0.32	0.47	0	1
Whether there is simple electronic management	0.67	0.47	0	1
Whether there are professional and technical personnel worker	0.49	0.50	0	1
Revenues	0.29	0.21	−1	0.9
Whether to provide guarantee for others	0.71	0.45	0	1
Years of agricultural employment	12.13	8.11	1	43
Invested outward	0.79	0.41	0	1
Is there mechanized automatic equipment?	0.61	0.49	0	1
Difficulty of realization of biological assets	2.21	0.77	1	3
Is the product of high-quality?	0.87	0.33	0	1
Is the value of the biological assets easy to evaluate?	0.32	0.47	0	1
Number of regional asset appraisal institutions	0.37	0.48	0	1
Biological asset market price fluctuation range	1.40	0.75	1	3
Whether there is any environmental policy risk	0.70	0.46	0	1
Whether there is a risk	0.25	0.43	0	1

risk assessment model is to be able to identify users with higher risks; thus, the correct rate of the negative sample classification is very important. The positive and negative samples in this paper are not balanced; therefore, the AUC value and the classification accuracy of negative samples are mainly used as the model performance evaluation indicators.

3.2.2. Model Construction and Parameter Optimization.
Using the Python 3.8 series as the modeling tool, install the XGBoost library and import the data, divide the data by the setout method, and randomly select 75% of the data as the training set through the frac in the df.sample function, and the remaining 25% of the test set. The distribution of the positive and negative samples of the training set and the test set is relatively balanced, as shown in Table 6.

The selection of XGBoost model parameters has a great impact on the performance of the model. The parameters included in XGBoost are divided into three types: general parameters, tree booster parameters, and learning task parameters. The learning rate parameter eta and row-sampling parameters in the tree booster parameters, such as subsample, column-sampling parameter colsample_bytree, and L1 and L2 regularization item weights alpha and lambda play a key role in preventing model overfitting. For the binary classification problem, the objective function parameter will be set to binary: logistic and the scale_pos_weight parameter will be set to 0.34 according to the positive and negative ratios of the sample.

We train the model; except for the fact that the maximum tree depth is selected as 3, the other parameters are all the default values of the XGBoost algorithm. The trained model is tested on the test set, and the test set evaluation

TABLE 5: Confusion matrix.

		Predictive value	
		0	1
Actual value	0	True negative (TN)	False positive (FP)
	1	False negative (FN)	True positive (TP)

TABLE 6: Training set and test set sample distribution.

	0 (risky sample)	1 (low-risk sample)
Training set	237	700
Test set	79	233

indicators are obtained. The model accuracy rate is 93.81%, the recall rate is 90.99%, the F1-score is 92.37%, the accuracy rate of negative samples reaches 82.28%, and the accuracy rate is 90.99% in terms of classification of positive samples. The accuracy rate is 90.99%, the AUC value is 0.9363, and the classification performance of the model is better.

Thereafter, the parameters are optimized. Hyperparameter optimization methods include grid search, random search, and Bayesian optimization. Grid search first specifies a subset of the hyperparameter space, exhausts all combinations of the given hyperparameters, and tries to find a set of optimal hyperparameters. In scikit-learn, the grid search can start from the parameter param_grid. The specified parameter grid is exhaustively obtained to obtain the optimal parameters to realize the grid search, A random search of the parameter space is conducted, and the value of the parameter is obtained through a probability distribution extraction. A grid search is suitable for small datasets and is currently the most widely used method in hyperparameter optimization. Therefore, this study selects a grid search as the parameter optimization method, a nested cross-validation in the grid search, and uses a three-fold cross-validation to find the optimal number of decision trees, and then grid search for other parameters step by step. There are eight parameters selected for optimization: eta, max_depth, min_child_weight, gamma, subsample, colsample_bytree, reg_alpha, and reg_lambda.

When tuning parameters, if the learning rate eta is too large, it will easily fall into overfitting. Being a too small value will increase the number of trees and increase the amount calculated for model training. Therefore, we first select a higher eta and set it to 0.1. Then, the cross-validation function in the XGBoost function is used to select the number of decision trees with the optimal eta value, tuning each parameter in turn, then reselect the eta value, and use the AUC value as the model performance evaluation index to finally obtain the best XGBoost model and optimal parameter combination. The optimal parameter combination is shown in Table 7.

Under the optimal parameter combination, the AUC value on the test set is 0.9493, and the ROC curve (Figure 1) is steep. The results of the confusion matrix are shown in Table 8. According to the confusion matrix, the accuracy of the negative sample classification of the model is 83.54%, the accuracy of the model is 94.17%. The recall rate is 90.13%,

and the f1 score is 92.11%, which is generally better than that of the model before tuning. The classification performance is better.

3.2.3. A Comparison of the Classification Performance of Different Machine Learning Methods.
Select LR, SVM, and RF algorithms for comparison, use the default parameters of each algorithm to train on the same dataset, and compare the evaluation index values after testing on the same test set. Then, optimize the parameters of each algorithm. The grid search method based on a three-fold cross-validation is still used in parameter optimization, and the model is retrained after finding the optimal parameter combination. The prediction is made on the test set, and the classification performance indicators such as the classification accuracy of negative samples and the AUC value are compared. Finally, the classification index values before and after optimization of each model are shown in Table 9.

From the comparison of the above four models, it can be seen that the classification performance of the XGBoost is better than that of other models before parameter tuning, which also confirms the desirability of choosing the XGBoost model to evaluate the risk of biological asset mortgage loans in this study. After parameter optimization, the negative sample classification accuracy rate of the XGBoost model is as high as 83.54%, which is much higher than that of the other models. Therefore, the XGBoost model is more suitable for biological asset mortgage risk assessment because of its higher classification accuracy and AUC value.

3.2.4. Analysis of Empirical Results.
Based on the above empirical analysis, this study establishes a biological asset mortgage loan risk assessment model based on the XGBoost, which has a good classification effect. According to the established XGBoost model, the importance value of each feature is obtained with the xgb.feature_importances_ command. There are three calculation methods for feature importance: "weight," "gain," and "cover." In this paper, the default method "weight," is the number of times that features are divided into attributes in all the trees. This is calculated as the standard, and the importance of each feature is shown in Figure 2.

It can be seen from the feature importance value that, except for the feature "whether the product is a high-quality product," where the importance value is 0, other features have a certain impact on the risk of biological asset mortgage loans. This shows that there is no significant relationship between product variety and loan risk. The difference between high-quality products and ordinary varieties is that the final circulation market is different, and the products can sell well in their respective circulation markets. Summarizing the above characteristic importance values, the established biological asset mortgage risk assessment system for new agricultural business entities is shown in Table 10.

TABLE 7: Optimal parameter combination.

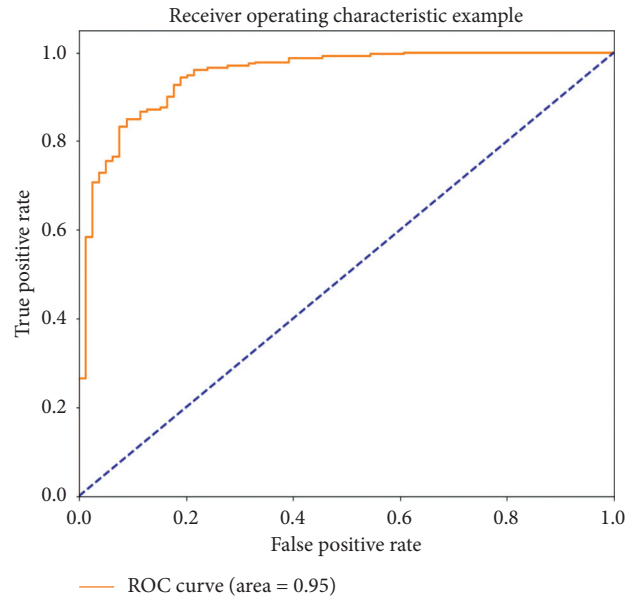| Hyperparameter	The optimal value
Learning_rate	0.1
Max_depth	3
Gamma	0.06
Min_child_weight	0.5
Subsample	0.8
Colsample_bytree	0.8
Reg_alpha	0.9
Reg_lambda	1.5

FIGURE 1: ROC curve of the XGBoost model after optimization.

TABLE 8: Confusion matrix results of the XGBoost model after tuning.

		Predictive value	
		0 (risky farmers)	1 (low-risk farmers)
Actual value	0 (risky farmers)	66	23
	1 (low-risk farmers)	13	210

TABLE 9: Classification index value of each model before and after tuning.

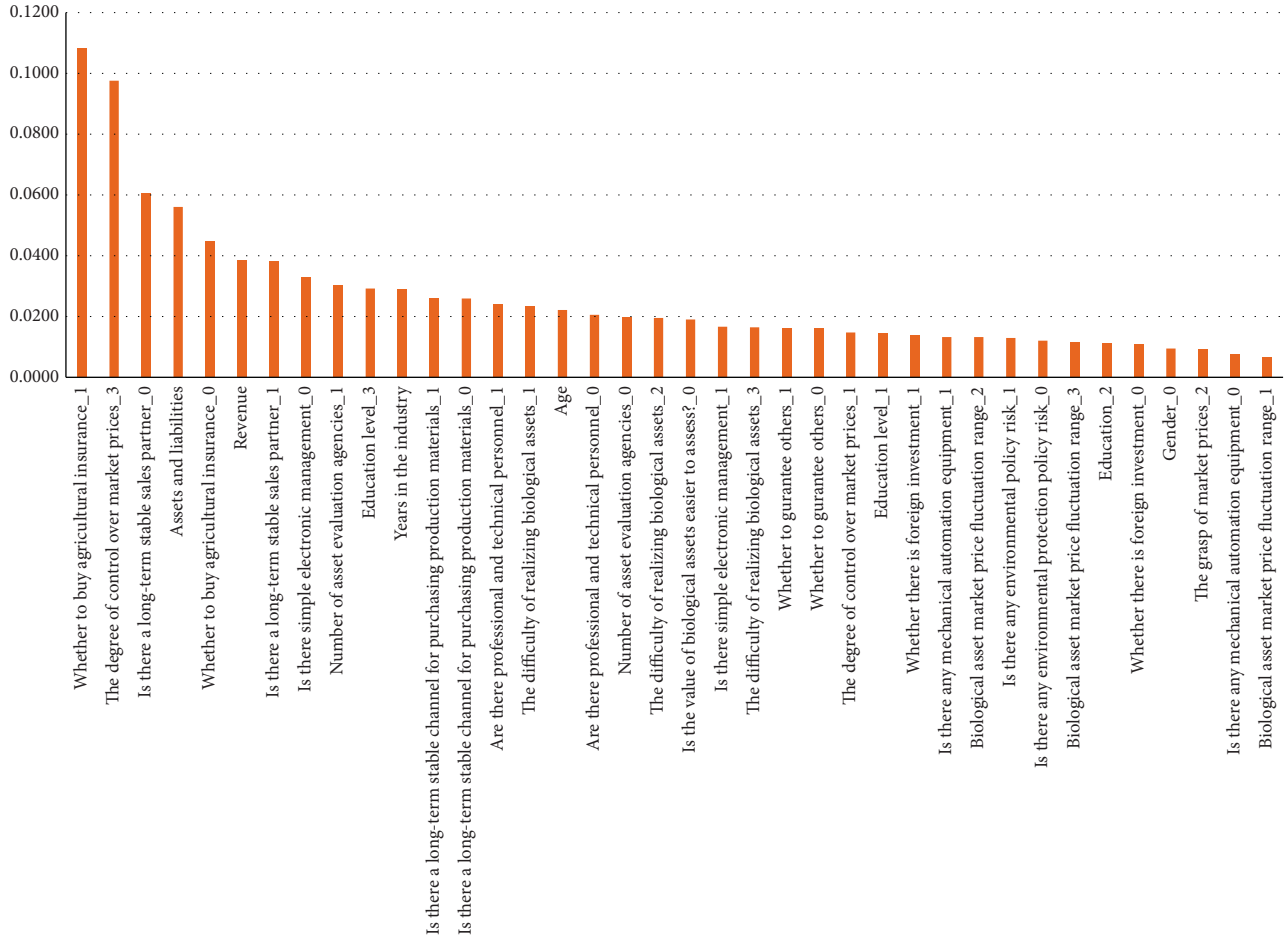| Model	AUC (%)	Negative class accuracy (%)	Positive accuracy (%)	F1-score (%)
Before parameter optimization				
XGBoost	93.63	82.28	90.99	92.37
LR	82.84	54.43	92.27	88.84
SVM		16.46	98.71	86.96
RF	92.88	60.76	96.14	91.80
After parameter optimization				
XGBoost	94.93	83.54	90.13	92.11
LR	82.82	50.63	93.13	88.75
SVM		67.09	93.13	91.18
RF	95.31	68.35	97.42	93.61

Figure 2: The importance of characteristics of biological asset mortgage risk assessment.

Table 10: The risk assessment index system of biological asset mortgage loans for new agricultural business entities.

Primary indicators	Secondary indicators	Index importance value
	Assets and liabilities	0.0561
Subject's personal credit risk	Education level	0.0559
	Age	0.0222
	Gender	0.0097
	Whether to buy agricultural insurance	0.1537
	Degree of grasp of market conditions	0.1224
	Stable and reliable sales channels	0.0990
	Stable and high-quality production material purchase channels	0.0525
	Whether there is simple electronic management	0.0501
Main operating risk	Whether there are professional and technical personnel worker	0.0450
	Revenues	0.0386
	Whether to guarantee others	0.0329
	Years of agricultural employment	0.0291
	Invest outward	0.0255
	Whether there is mechanized automatic equipment	0.0212
Risks of biological assets	Difficulty of realization of biological assets	0.0600
	Is the value of biological assets easy to evaluate?	0.0192
	Number of regional asset appraisal institutions	0.0503
Policy risk and market risk	Biological asset market price fluctuation range	0.0316
	Whether there is any environmental policy risk	0.0246

4. Conclusions

4.1. Basic Conclusion. The operating status of the lender is the most important type of risk. Among them, "whether to buy insurance" is an important indicator to measure the risk of biological asset mortgage loans for new agricultural operators, and this indicator can be used as the "threshold" for financial institutions to issue biological asset mortgage loans. Furthermore, the new agricultural business entities' grasp of market conditions and whether the products have reliable sales channels are also key indicators that will trigger the risk of biological asset mortgage loans. The asset-liability ratio and educational level of the new agricultural business entities are important credit risk indicators. The difficulty of realizing biological assets and whether it is easy to evaluate the value of biological assets have a certain impact on the risk of mortgage loans of biological assets. The number of asset appraisal agencies determines the quality of the regional asset appraisal environment. A good asset appraisal environment can make biological asset valuations more reasonable, thereby reducing the risk of unreasonable biological asset valuations.

4.2. Related Suggestions. Relevant departments of the Chinese government should keep up with the pace of financial innovation and create good conditions for financial institutions to carry out biological asset mortgage loans. First, the "Guarantee Law" and "Property Law" should improve the relevant provisions of biological asset mortgage loans to provide a good legal environment for financial institutions to conduct business. Second, a mechanism for the government, insurance companies, and financial institutions to share the burden should be established. Third, to promote the development of biological asset mortgage loans, government departments should increase financial support to achieve full coverage in terms of policy insurance for farmers, reducing the financial pressure of farmers. Fourth, corresponding guarantee companies can be established to serve Chinese farmers. The biological asset mortgage loans are used as guarantees to reduce the risks borne by financial institutions. At the same time, insurance companies should develop agricultural insurance businesses specifically for biological assets, and for farmers who have purchased policy-based agricultural insurance, insurance costs can be appropriately reduced. This will not only increase the business volume of insurance companies but also bring benefits to farmers. When assets are damaged, insurance companies compensate farmers for their losses in a timely manner so that they have funds to repay financial institutions.

Biological asset valuation is a major problem. Asset valuation industry associations and relevant government departments should help integrate existing biological asset valuation resources, standardize and improve biological asset valuation technology, increase investment, and form a professional biological asset valuation framework. The team conducts targeted assessments of the different growth stages of various assets in the biological asset category, and forms a unified standard to improve the accuracy of the assessment of their value and continuously reduce the operational risk caused by inaccurate assessment of collateral value, thereby reducing the default risk of lenders and reducing the risk loss of financial institutions. Therefore, in the face of a wide range of mortgage loan needs, we should actively seek cooperation with external valuation agencies, learn from each other, learn from the valuation methods of biological asset valuation institutions, and combine the principle of prudence of financial institutions to rationally valuate the value of biological assets.

After financial institutions provide loans to rural households, they still need to monitor the biological assets of rural households after the loan and continue to track and analyze them. For animal biological assets, the lender should prevent them from handling or transferring it without authorization; for plant biological assets, attention should be paid to their growth cycle and the impact of natural disasters. Financial institutions should interface with agricultural product information platforms established by relevant departments of the Ministry of Agriculture of various regions, and use remote video monitoring to dynamically grasp the status of biological assets. Monitoring of the mandatory inspection and quarantine of livestock slaughter transactions can effectively prevent the risk of private disposal of live collateral. This "cloud monitoring" model can solve the problem of difficult verification and control of collateral to the greatest extent. In the future, relevant departments will publicize information and update data in a timely manner to achieve data interconnection, reduce information asymmetry, effectively control postloan risks, and continuously improve credit risk prevention and control capabilities.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon reasonable request (email: 1179188@mail.dhu.edu.cn).

Conflicts of Interest

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

Acknowledgments

References

[1] X. Gao, "Evaluation of the development level of my country's new agricultural management system," *Economist (China)*, vol. 9, pp. 73–80, 2017.

[2] X. Chen and W. Zhu, "Research on the development of new agricultural business entities under the agricultural supply-side reform—based on the investigation and analysis of Sichuan," *Rural Economy (China)*, vol. 8, pp. 108–114, 2017.

[3] X. Hu, "Research on the cultivation of new agricultural management subjects under the background of rural revitalization," *Agriculture of Henan (China)*, vol. 26, p. 50, 2019.

[4] H. Ying, "Research on the dilemma and countermeasures of the industrialization development of the new type of agricultural management," *Agricultural Economy (China)*, vol. 3, pp. 62–64, 2019.

[5] H. Huang, "Thoughts and suggestions on financial support for the development of new agricultural business entities to help alleviate poverty," *Wuhan Finance (China)*, vol. 11, pp. 63–65, 2018.

[6] J. Min, "Analysis on the innovation of the new business entity business model——based on the case analysis of Heilongjiang Renfa Cooperative," *Issues in Agricultural Economy (China)*, vol. 10, pp. 50–59, 2018.

[7] L. J. Mester, "What's the point of credit scoring?" *Business Review*, vol. 3, pp. 3–16, 1997.

[8] E. I. Altman, "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy," *The Journal of Finance*, vol. 23, no. 4, pp. 589–609, 1968.

[9] W. Dobbie, G.-P. Paul, N. Mahoney, and J. Song, "Bad credit, no problem? Credit and labor market consequences of bad credit reports," *Asia-Pacific Journal of Financial Studies*, vol. 49, no. 3, pp. 368–409, 2020.

[10] C. H. Sossou, F. Noma, and J. A. Yabi, "Rural credit and farms efficiency: modelling farmers credit allocation decisions, evidences from Benin," *Economics Research International*, vol. 2014, Article ID 309352, 2014.

[11] K. K. Bolarinwa and E. O. Fakoya, "Impact of farm credit on farmers socio-economic status in ogun state Nigeria," *Journal of Social Sciences*, vol. 26, no. 1, pp. 67–71, 2011.

[12] J. Du and D. Lu, "Operation mode, existing problems and countermeasures of peasant household biological asset mortgage," *Journal Of South China Agricultural University (Social Science Edition) (China)*, vol. 12, no. 2, 2013.

[13] W. Li, "Exploring and research on the financing model of live animal mortgage——taking hainan prefecture, qinghai province as an example," *Finance Economy (China)*, vol. 20, pp. 161-162, 2014.

[14] F. Heywood, "Secured transactions: the power of collateral," *Finance and Development*, vol. 6, pp. 44–46, 1996.

[15] J. Du, D. Lu, and F. Chunyan, "Influencing factors and hierarchical analysis of farmers' biological asset mortgage financing willingness," *Journal of China Agricultural University (China)*, vol. 19, no. 2, pp. 238–244, 2014.

[16] P. Luo, "Analysis of influencing factors of farmers' biological asset mortgage financing willingness—based on logistic model estimation," *Forestry Economics (China)*, vol. 37, no. 12, pp. 24–29, 2015.

[17] W. Fan, X. Zhang, and H. Hu, "Study on the selection scheme of biological asset value evaluation method," *Friends of Accounting (China)*, vol. 20, pp. 17–19, 2013.

[18] E. Baltensperger, "Credit rationing issues and questions," *Journal of Money, Credit and Banking*, vol. 10, no. 2, pp. 170–183, 1978.

[19] C. John, "A note on the comparison of logit and discriminant models of consumer credit behavior," *Journal of Financial and Quantitative Analysis*, vol. 15, no. 3, 1980.

[20] W. Liang and S. Wen, "Research on loan default risk assessment of agricultural small and medium-sized enterprises-Based on the data of agriculture, forestry, animal husbandry and fishery enterprises in the "new third board"," *Rural Economy (China)*, vol. 11, pp. 93–100, 2019.

[21] A. Samad, "Credit risk determinants of bank failure: evidence from US bank failure," *International Business Research*, vol. 5, no. 9, pp. 10–21, 2012.

[22] M.-F. Hsieh and C.-C. Lee, "Bank liquidity creation, regulations, and credit risk," *The Journal of Finance*, vol. 75, no. 5, pp. 2377–2419, 2020.

[23] E. Kamalloo and M. S. Abadeh, "Credit risk prediction using fuzzy immune learning," *Advances in Fuzzy Systems*, vol. 2014, Article ID 651324, 2014.

[24] Y. Yan and W. Fan, "Research on the risk of mortgage financing of live animals," *Rural Economy and Science-Technology (China)*, vol. 21, pp. 163–165, 2016.

[25] G. Teles, J. Joel, P. C. Rodrigues, R. A. L. Rabêlo, and S. A. Kozlov, "Comparative study of support vector machines and random forests machine learning algorithms on credit operation," *Software: Practice and Experience, Early View*, vol. 21, 2020.

[26] J. C. Westland, "Predicting credit card fraud with Sarbanes-Oxley assessments and Fama-French risk factors," *Intelligent Systems in Accounting, Finance and Management*, vol. 27, no. 3, pp. 95–107, 2020.

[27] J. Futoma, J. Morris, and J. Lucas, "A comparison of models for predicting early hospital readmissions," *Journal of Biomedical Informatics*, vol. 56, pp. 229–238, 2015.

[28] Y. Wu, L. Feng, Z. Jiang, and W. Kong, "Research on risk early warning of P2P online loan platform based on machine learning model," *Journal of Finance and Economics (China)*, vol. 9, pp. 18–25, 2019.

[29] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[30] M. Z. Joharestani, C. Cao, X. Ni et al., "PM2.5 prediction based on random forest, XGBoost, and deep learning using multisource remote sensing data," *Atmosphere*, vol. 10, no. 7, p. 373, 2019.

[31] Y. Zhou, T. Li, J. Shi, and Z. Qian, "A CEEMDAN and XGBOOST-based approach to forecast crude oil prices," *Complexity*, vol. 2019, pp. 1–15, 2019.

[32] L. Xia, Y. Zhang, Q. Lu, and G. Tang, "Credit rating method combining XGBoost algorithm and Logistic regression," *Credit Reference (China)*, vol. 37, no. 11, pp. 56–59, 2019.

[33] C. Li, X. Zheng, Z. Yang, and Li Kuang, "Predicting short-term electricity demand by combining the advantages of ARMA and XGBoost in fog computing environment," *Wireless Communications and Mobile Computing*, vol. 2018, pp. 1–18, 2018.