Information Resources Management Journal
Volume 37 • Issue 1

Risk Prediction of the Development of the Digital Economy Industry Based on a Machine Learning Model in the Context of Rural Revitalization

Rui Luan, Shenyang Polytechnic College, China
Ping Xu, Xihua University, China*

ABSTRACT

In today's society, rural areas face challenges such as complex terrain and uneven population distribution, and infrastructure construction is exceptionally difficult. At the same time, poor information transmission and low communication efficiency have also become a major obstacle to the promotion of the digital economy in rural areas. This study aims to use gradient advancement models to identify potential risks in the growth of the digital economy sector related to rural revitalization. In this study, we used an enhanced hierarchical gradient boosting algorithm. The research results indicate that the introduction of this technology can provide us with a more comprehensive and reliable risk prediction model, thereby more scientifically assisting the development and decision-making of the digital economy in rural areas. This article provides a new perspective and solutions for development issues in rural areas, promoting sustainable development and economic growth in rural areas.

KEYWORDS

Digital Transformation, Enhanced Hierarchical Based Gradient Boost Algorithm (EHGBA), Principal Component Analysis (PCA), Risk Prediction, Rural Revitalization

In today's society, rural areas face challenges such as complex terrain and unequal population distribution, which makes infrastructure construction extremely difficult. At the same time, poor information transmission and low communication efficiency have also become a major obstacle to the promotion of the digital economy in rural areas. Therefore, this study aims to use gradient advancement models to identify potential risks in the growth of the digital economy sector related to rural revitalization. To achieve this goal, we first need to standardize the collection and preprocessing of the dataset to ensure its accuracy and comparability. Next, we will use principal component analysis (PCA) to extract features related to risk prediction from the digital economy field. This will help us better understand the risk factors that the digital economy may face in the development of rural areas. In this study, we also introduced an enhanced hierarchical gradient boosting algorithm called EHGBA technology. By improving and optimizing the algorithm, we hope to assess the potential risks of the

DOI: 10.4018/IRMJ.343095
*Corresponding Author

digital economy sector more accurately in the growth process of rural areas. The research results demonstrate that the introduction of this technology can provide us with a more comprehensive and reliable risk prediction model, thereby more scientifically assisting the development and decision-making of the digital economy in rural areas. This article provides a new perspective and solutions for the development issues in rural areas, promoting sustainable development and economic growth in rural areas.

The innovation points of this article are as follows:

(1) This article proposes an EHGBA technique that enhances the hierarchical gradient boosting algorithm to identify potential risks to the growth of the digital economy sector related to rural revitalization. This technology combines the ideas of hierarchical learning and gradient enhancement.
(2) This article extracts features through standardized collection and preprocessing of datasets, as well as principal component analysis (PCA), in order to better adapt to complex data distributions and problem structures and improve the learning ability and generalization performance of the model.
(3) The method proposed in this article has great potential for application in the field of rural revitalization and provides new ideas and methods for solving similar problems.

LITERATURE REVIEW

The percentage of digitalized agriculture will reach 22.5% by 2020, while the percentage of digitalized quality safety traceability will reach 22.1%. Livestock and poultry output, as well as the cultivation of their facilities, will rise to levels that outpace those of the whole country (Zhang, 2022). For rural economies to thrive, it is crucial to increase the degree of digitization in agricultural output (Zhao et al., 2022). Establishing a new pattern of coordinated development of e-commerce services, packaging and transportation, and other related industries through the digital sales model, centered on the sale of local characteristics, is essential if rural e-commerce is to realize its full potential in large-scale industrial development (Henrique et al., 2019). One of the most visible indicators of the growth of the digital economy in rural regions is the rise of online shopping (Petropoulos et al., 2020). As a result of the spread of digital technology into China's rural areas, analysts anticipate that e-commerce sales from rural areas will increase by a factor of roughly six by 2021 (Huateng et al., 2021). As a result of rising consumer interest in buying food and other farm goods online, the number of marketplaces that cater to this niche has grown (Obthong et al., 2020). Agricultural commodity trade has increased during the last five years. E-commerce in rural areas has helped boost sales on a massive scale, made it easier for information to flow more smoothly, and leveled the playing field between farmers in different parts of the country and a sizable market (Wei et al., 2020).

The agriculture sector is a prime example of how scientific and technological innovation may positively impact economic growth due to the many ways in which it has been used there (Vadlamudi et at., 2020). As part of a broader drive to revitalize rural areas, today's farmers have embraced the digital age, therefore advancing agricultural research and technology. Science and technology in agriculture are directly responsible for 60.7% of the development made thus far (Schwartz, 2022). Insistence on using cutting-edge innovation has fueled the demonstration effect, leading to a number of game-changing discoveries (Lee et al., 2011).

The development of new technologies has had a significant effect on farming in rural areas (Zhang et al., 2021). As an example, we discuss rural digital finance, investigating how big data and other technologies may be utilized to address the funding gaps that exist, overcome the unique challenges of rural financing, and ultimately contribute to the growth of the rural digital economy. New enterprises are the best evidence of industrialization in rural regions. New growth is being sparked in rural economies and the new economy as a result of the use of information technology and electronic commerce to address the supply-side imbalance. In 2020, several new rural companies mimicking rural tourism sprung up in response to the New Coronary Pneumonia pandemic. The rural tourist sector is expected to begin its revival in 2021. Economic growth in China stands to benefit from ecotourism, which is expected to attract over 2.1 billion visitors to the nation by 2020. In 2020, there will be an estimated 10.1 million city dwellers who have moved to the country in search of better economic and creative opportunities (Tao, 2018). There is a new type of rural business that is contributing to the digital transformation of rural areas by maintaining their own distinctive products and services and gaining product appreciation based on local peculiarities (Johnson & Noguera,2012).

People who live in rural areas tend to have lower cultural literacy, a more static way of life and mindset, and a propensity to reject new ideas as a matter of habit. It is challenging to complete the transformation of ideas on the first try in the face of new technologies, new platforms, new channels, and new business models; thus, it is important to raise people's level of digital economy knowledge (Zhang et al., 2023). Some rural inhabitants have benefited from state and federal aid programs for the poor for a long time and depend on these programs for their basic survival requirements. Consequently, they still lack a robust understanding of how digital tools might boost productivity. However, due to financial constraints, businesses and institutions are hesitant to invest in data, information, and technology. As a result, they are slow to move forward with plans to integrate the digital economy with agricultural products, and therefore, enterprise resource support in rural areas is low.

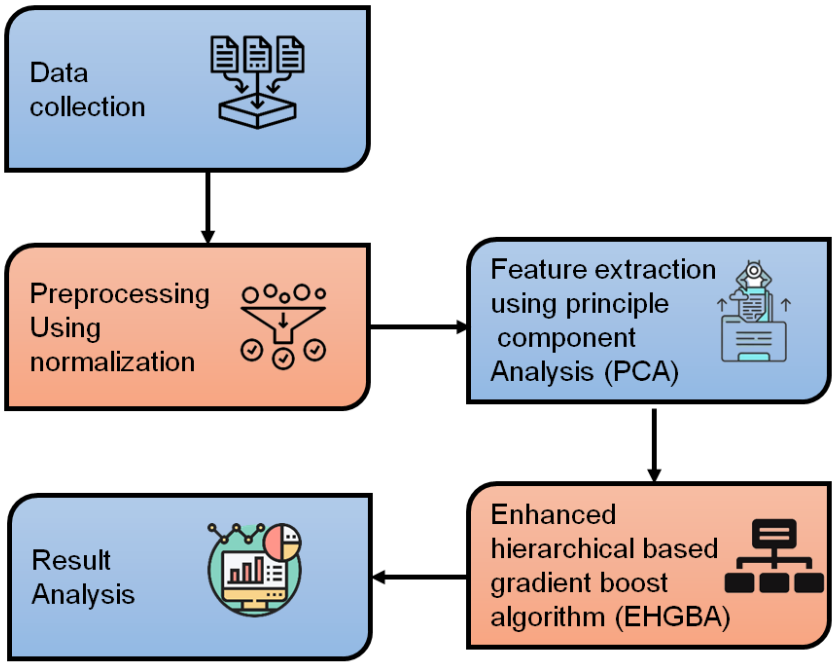
The rapid growth of the digital economy in rural regions requires individuals to have access to high-quality human resources. There is now a widening gap between the technical and professional skills of those who are fluent in digital technology and digital literacy (Heng & Xu, 2023). Young and middle-aged workers, as well as graduates from universities that are open all year, have better educations and more familiarity with technology than their rural counterparts. Numerous top-tier individuals have been lost, which is an undeniable reality. Many retirees choose to remain in the country. Their minds can't be changed easily. There are many low-income rural families with little or no access to network information technology because they lack the skills and knowledge necessary to utilize network platforms themselves. Furthermore, the number of "new farmers" has risen over time (Zhang, 2014) due in part to government incentives. Despite making up for the dearth of rural human resources to some degree, their limited scope prevents them from being employed as a primary means of bolstering talent.

The growth of the digital economy cannot occur without the establishment of stable supporting infrastructure. High-velocity, high-quality information processing is essential to the information and digital economy (Ma, 2020). The challenges of building infrastructure in rural areas are exacerbated by variables such as complex topography and unequal population distribution. Poor information transmission, low communication efficiency, and other phenomena hinder the further promotion of the digital economy in the countryside (Wang, 2020).

Manual model construction and statistics were formerly the backbone of economic modeling and forecasting, but the sheer amount of data accessible in today's digital economy has rendered these methods obsolete (Claveria et al., 2019). Moreover, conventional approaches lack the quantitative rigor of large-scale sampling and objective scientific prediction (Claveria et al., 2017), making their analysis overly intuitive. Since this is the case, several researchers have proposed economic modeling and prediction based on big data analysis and AI-based technologies. Numerous variables have a role in the economy, and some of them are more consequential than others at certain junctures (Ptak-Chmielewska, 2019). Establishing the interrelationships between economic variables by hand is a time-consuming and error-prone process due to the inherent intricacy of these linkages.

A model has been proposed after analyzing the relevant components using a graph network economics analysis. This graph network architecture allows for the simultaneous input of several pieces, which are then automatically correlated based on their respective learning weights. Additionally, a model of economic forecasting is presented using long-term memory (LSTM) (Guo & Miao, 2022). The future of the economy might be predicted with the help of this model. Optimizing and enhancing

Figure 1. The Working Strategy of the Suggested System



the rural innovation and entrepreneurship environment (Li, 2022) via the integration of advanced and innovative technology across the whole agricultural process and industrial chain has opened new opportunities for business startups and job creation in rural regions, in addition to facilitating new forms of trade (Wen & Zhang, 2022). Smart agriculture, rural e-commerce, agricultural scientific and technological innovation, and new rural enterprises are the primary entry points for the digital economy into rural agricultural production (Yao & Liu, 2022).

RELATED MATERIALS AND METHODS

The term digital economy may be used to describe a broad spectrum of businesses, from homegrown, personal companies utilizing rudimentary techniques to global conglomerates employing existing tools to mine and refine wealth. A major obstacle to the digital evolution of the economy might be an inadequate investment in technology and information systems (Wen et al., 2021). A framework of economic and interpersonal interactions based on the utilization of digital innovation is needed, the widespread adoption of which might eventually force the integration of the physical and digital worlds, with far-reaching implications for the character of the economy and the financial markets. The working strategy of the suggested system is presented in Figure 1. In this section, we describe the proposed system with its relevant features.

Sampling

This research uses two micro-entrepreneurs' datasets on a significant number of Chinese manufacturing firms in order to test the predictions. One is the Chinese Industrial Enterprise Dataset, which includes both government and private industrial firms with annual revenues more than the set threshold. This dataset contains information about 1,000 micro entrepreneurs or companies, including their names, registered locations, industry classifications, establishment dates, number of employees, etc. These pieces of information were collected and organized through government agencies and commercial databases. The other is the most trustworthy industrial environmental management sample dataset in China, the Enterprise Pollution Database from the Republic of the union of China's Ministries of Environment and Natural Resources (or State Environmental Protection Organization, as it was originally known). This dataset contains information on about 1,000 micro-entrepreneurs or companies.

Beijing Forecast Information Systems Co., Ltd., which manages the earnings per share (EPS) dataset, is responsible for the synchronization of the two different datasets. Several macro variables from the EPS database are also used in this analysis. It is constrained, on the one side, by the pace at which the information technology sector is evolving, and, on the other, by the timing of the most recent version of the industrial pollution dataset. From 2012 to 2021, all data was collected. This article investigates publicly available sources of macroeconomic data (such as the National Bureau of Statistics, Ministry of Finance, etc.) to obtain specific data on relevant macroeconomic variables, such as gross domestic product (GDP), inflation rate, unemployment rate, etc., and includes them in the analysis scope. If missing macro variable data cannot be directly obtained, data comparison analysis can be considered. Compare and analyze existing earnings per share data with other known macroeconomic indicators to infer the potential impact of macroeconomic variables on corporate earnings. In addition, this article also conducts sensitivity analysis to evaluate the potential impact of missing macro variable data on research conclusions by simulating the impact of different macro variable values on earnings per share. This helps to provide more comprehensive analysis and conclusions in research, while also reducing reliance on data integrity.

According to best practices in data coordination, Beijing Predictive Information System Co., Ltd. has taken the following measures to ensure synchronization and merging of two different datasets:

(1) Data cleaning: Firstly, the original data should be cleaned, including handling missing values, outliers, and erroneous data. This can ensure the accuracy and reliability of subsequent analysis.
(2) Data standardization: As two datasets may use different formats and units, they need to be converted to the same format and units for comparison and analysis. For example, converting all currencies to the same currency and converting all dates to the same time format.
(3) Data matching and integration: For two different enterprise datasets, it is necessary to match their common fields, such as enterprise name or registration number. After successful matching, they need to be integrated into a dataset for analysis and prediction.
(4) Data validation and verification: In the process of data coordination, data validation and verification should be carried out to ensure the accuracy and consistency of the data. For example, checking data through random sampling or using algorithms to detect outliers and duplicate data.
(5) Data confidentiality: Due to the involvement of sensitive data, appropriate measures need to be taken to protect the security and confidentiality of the data. For example, restricting access to data by using encryption techniques or access control.
(6) Data update: As data may change over time, it is necessary to regularly update the data to ensure the accuracy and timeliness of the analysis results.

Data Preprocessing

Normalization is a data preprocessing technique used to convert data of different scales and ranges into a unified standard range to eliminate dimensional differences between different variables. The role of normalization is as follows:

(1) Enhance model performance: Some machine learning algorithms are sensitive to the scale and range of input data, and if not normalized, it may lead to a decrease in model performance. By utilizing normalization, the weights of different features can be more balanced, avoiding the significant impact of certain features on the model due to their large numerical range.

(2) Assisted convergence: Normalization can help optimize algorithms to converge faster during model training. For gradient based optimization algorithms, if the features have different scales, the speed of gradient update will be affected, leading to slower convergence speed. Normalization can make the gradient of features smoother and improve the efficiency of optimization algorithms.
(3) Improving model stability: Normalization can reduce model instability caused by changes in data distribution. If the data range of different features varies greatly, the model may become more sensitive to certain features, leading to unstable performance on new data. Normalization can reduce the differences between features and improve the model's generalization ability toward new data.
(4) Data visualization: In the process of data visualization, normalization can make data with different features comparable and better display the relationships between data. For example, normalizing data to a range between 0 and 1 can provide a more intuitive comparison of the impact of different features on the target variable.

In summary, normalization plays an important role in data preprocessing, helping to improve the performance, stability, and interpretability of machine learning algorithms. This article uses the minmax scaling method, which is a common normalization method and has the following main advantages:

(1) Easy to understand and implement: The calculation process of minimum maximum normalization is simple and easy to understand and implement. It only needs to find the minimum and maximum values of the data and linearly convert the data to a given range (usually between 0 and 1).
(2) Preserves the relative relationship of data: Minmax normalization can preserve the relative relationship of the original data without changing the distribution shape of the data. This is very useful for algorithms that require preserving data structure and distribution characteristics, such as clustering and classification algorithms.
(3) Non-dependent distribution assumption: The minimum maximum normalization does not depend on the distribution assumption of the data and is applicable to various types of datasets. In some cases, data distributions may have different shapes and ranges, and minimum maximum normalization can make different features comparable.

The goals of data preprocessing are to ensure that only essential information is stored, that every data block has exactly one piece of information, and that any redundant information has been removed. Independent variables are adjusted consistently across this procedure to ensure they stay within a predetermined limit. There are a variety of normalizing techniques available, however, in this instance, we employ min-max normalization. The normalization expression then looks like this:

$$n = \left(\left(\frac{\left(a - a_{min}\right)}{\left(a_{max} - a_{min}\right)}\right) * \left(1 - 0\right) + 0\right) \tag{1}$$

The maximum and minimum values of the normalized data are denoted by a max and a min, respectively, and the range is from 0 to 1. The average and mode of a character are represented by the attribute's observed or estimated values.

Skewness is a statistical measure that describes the shape of a data distribution and measures the degree of asymmetry in the data distribution. Positive skewness indicates that the data distribution is skewed to the right, with a longer tail on the right side; negative skewness indicates that the data distribution is skewed to the left, with a longer tail on the left side; and zero skewness indicates that the data distribution is basically symmetrical. In the process of data standardization, skewness can provide important information about the shape of data distribution, thereby affecting the selection of standardization methods and the interpretation of results. Specifically, the importance of skewness in the standardization process is reflected in the following aspects:

(1) The selection of standardization methods may be affected by skewness when the data exhibits a clear skewed distribution. Simple minimum maximum normalization or Z-score normalization may be affected by skewness, resulting in the standardized data still retaining the original skewed characteristics. In this case, it is possible to consider using non-linear normalization methods that are more suitable for skewed data, such as Box Box transformation or Yeo Johnson transformation, to better eliminate skewness in the data distribution.
(2) Interpreting standardized data: Understanding the skewness of the original data can help us better understand the characteristics of standardized data when interpreting it. If the original data shows a significant skewed distribution, the standardized data may also retain this skewed feature, so it is necessary to carefully explain the distribution shape and corresponding statistical characteristics of the standardized data.
(3) Impact on modeling results: When conducting modeling analysis, data skewness may also affect the performance and stability of the model. For some machine learning algorithms, such as linear regression or logistic regression, skewed data may lead to an increase in model bias, thereby affecting the predictive ability of the model. Therefore, when conducting data standardization and modeling analysis, it is necessary to fully consider the impact of skewed distribution of data on model results.

In summary, understanding the skewness of data is crucial for the process of data standardization and modeling analysis. It can help us choose appropriate standardization methods and more accurately interpret the distribution characteristics of data, thereby improving the reliability and interpretability of data analysis.

Minimum skewness indicates proper data sources, which was formerly defined as the discrepancy of a dataset's spectrum. Therefore, Equations (2), (3), and (4) represent the average, midpoint, and deviation, respectively.

$$M = \frac{\sum_{i-1}^{n} y_1}{n} \tag{2}$$

$$M_d = \begin{cases} y\dfrac{n}{2}, ifniseven \\ \dfrac{\left(y\left[\dfrac{n-1}{2}\right]+y\left[\dfrac{n+1}{2}\right]\right)}{2}, \wedge ifnisodd \end{cases} \tag{3}$$

$$S_k = n\sum_{i=1}^{n} \tag{4}$$

If 'n' represents the total quantity of values contained within the dataset, then 'y' represents those values. In addition, $\overline{y}$ stands for the average, while σ denotes the standard deviation. In order to normalize security, it is necessary to find and collect all pertinent information. Attributes in normalization play an important role in several methods for detecting anomalies.

Extracting Features

Principal component analysis (PCA) is a commonly used method for data dimensionality reduction and feature extraction. It achieves the goal of dimensionality reduction and feature extraction by projecting the original data onto a new coordinate system, resulting in the maximum variance of the projected data. The main advantages of PCA are as follows:

(1) Data visualization and exploratory analysis: In the early stages of data analysis and research, we hope to visualize the data to understand its distribution, structure, and correlation. PCA can help us better understand patterns and trends in data by reducing high-dimensional data to two-dimensional or three-dimensional space, visualizing the data as scatter plots or surface graphs.
(2) Feature selection and extraction: For high-dimensional data, there are a large number of redundant or irrelevant features that may interfere with model training, increase computational complexity, and may lead to overfitting. PCA can reduce the dimensionality of data by retaining the most important principal components, namely features with high variance, and extract the most representative features for subsequent modeling and analysis.
(3) Data preprocessing and removal of linear correlation: In some studies, data may have linear correlation, which may lead to model instability and multicollinearity issues. By applying PCA, we can transform data into a new coordinate system where the principal components are independent of each other, thus solving the problem of linear correlation.
(4) Noise filtering and signal enhancement: In some studies, data may contain noise or interference, which may have an impact on subsequent analysis and model building. Through PCA, we can identify and remove noise components from the data, thereby improving the quality of the data and the clarity of the signal.

In summary, the PCA method can help us understand the structure and correlation of data, select the most important features, solve data correlation and multicollinearity problems, and improve data quality and signal clarity, laying a foundation for subsequent research and analysis.

Principal component analysis (PCA) is a well-known method for evaluating large datasets with several parameters and characteristics for every interpretation. The goal is to save as much of the original information as possible while making the information easier to understand, and this is done quantitatively. Quantization, facial recognition, neurology, the digital economy, and graphic design are just a few of the many applications that make use of PCA. The PCA approach relies on the feature set to generate the eigenvalues of the correlation matrix. Only a few of the features, some with high eigenvalues, will be separated from each other allowing for more investigation. The remaining characteristics may be ignored. The vast depth of the subset of features is greatly simplified as a consequence. The correlation matrix can be depicted as follows.

$$B = \frac{1}{x}\sum_{n=1}^{x}\left\{G_n - \alpha\right\}\left\{G_n - \alpha\right\}^T \tag{5}$$

where G_n is the design (n = 1 to x), x is the set of entries, and x is the attribute vector. Where T represents transpose of matrix.

$$CV_j = \tau_j u_j \tag{6}$$

(j = 1, 2, 3…m), m = number of attributes, τ_j is the eigenvalue, and u_j represents the eigenvector.

Fisher's criterion refers to an expansion of principal component analysis known as the Fisher discriminant analysis (FDA). Discriminant separation $q_n^{x,y}$ between subclasses x and y may be computed given the standard deviations and mean values of the desired characteristics. To compute the characteristic value using PCA, use the solution given in Equation (7).

$$q_n^{x,y} = \frac{\left|\left(meanp_n^x\right) - mean\left(p_n^y\right)\right|^2}{\left[Sd\left(p_n^x\right)\right]^2 + \left[Sd\left(p_n^y\right)\right]^2} \tag{7}$$

Here, p_n^x and p_n^y denote the nth attribute under conditions x and y, respectively, for transporting objects. For the nth capability, the separation between the two classes of directions x and y is given by $q_n^{x,y}$, where mean () and standard deviations () represent the mean and standard deviation, respectively.

The dispersion within a category usually refers to the differences or degree of dispersion between data points within the same category. By measuring the degree of dispersion within a category, we can better understand and describe the internal structure of the data. The dispersion across subcategories usually refers to the difference or degree of dispersion between data points from different subcategories. By measuring the degree of dispersion across subcategories, we can better understand and describe the overall structure of the data and provide a foundation for subsequent data analysis and model building. The following equation makes clear that the dispersal within a category equals the total of its independent sector, whereas the dispersal across subcategories equals the squares of the range between such individual means.

It is a two-dimensional representation of a one-dimensional difference across two data sets. For the numerous (category both x and y) test, the data may be found in Equation (8). The gap between both the overall mean and the test variances is accurately shown.

$$T_n^{x,y} = \frac{\left|\left(Meanp_n^x\right) - Mean\left(p_n^y\right)\right|^2}{\left[Std\left(p_n^x\right)\right]^2 + \left[Std\left(p_n^y\right)\right]^2} \tag{8}$$

where x and y are weight limits specific to the nth aspect and component, and n is the index by which those features were isolated. Generate an A-shaped matrix out of the separation characteristics data for each section.

$$Y = \left[Y_1, Y_2, \ldots\ldots, Y_n\right] \tag{9}$$

It contains r sets of data, whereas n represents the total number of observations. Next, we use Equation 10 to determine the correlation matrix S.

$$T = \frac{1}{m}\sum_{k=1}^{m}\left(Y_k - \bar{Y}\right)\left(Y_K - \bar{Y}\right)^S \tag{10}$$

In this case, Y is used as an illustrative number. Since S is a r r matrix and its eigenvalues are [δ_1, δ_2,...., δ_n] ($\delta_1 \geq \delta_2 \geq ... \delta_2 \geq$), the eigenvector S = [u1, u2,......, un] must be generated. This eigenvector provides an orthogonal basis for business information. Features with a higher value may be more helpful. Using normalized methods, one may obtain the percentage Qk as given in Equation 11.

$$QK = \lambda k \left(\sum_{i=1}^{m} \delta j \right)^{-1} \tag{11}$$

Eliminate any eigenvalues that contribute very little to the selected attribute. Using model Q as a reference, the restoring vector Y is constructed using the first d vectors. The physical dataset used in the automated processes places more emphasis on the aspects with the highest reliability, as determined by the correlation matrix than on the elements with the lowest correlation. This provides a strategy for reducing the number of dimensions, which speeds up the search for optimal features.

$$\overline{X = \sum_{i=1}^{d} u_i^T X u_i} \tag{12}$$

Enhanced Hierarchical Based Gradient Boost Algorithm (EHGBA)

Enhanced hierarchical gradient enhancement algorithm (EHGBA) is an improved machine learning algorithm aimed at improving the performance and efficiency of gradient enhancement algorithms. This algorithm combines the ideas of hierarchical learning and gradient enhancement, reducing the problem space through hierarchical learning and improving the accuracy and generalization ability of the model through gradient enhancement. The main steps of the EHGBA algorithm include:

(1) Hierarchical learning: Divide the original dataset into multiple subsets and organize them in a hierarchical manner according to certain rules. This can reduce the problem space and improve learning efficiency.
(2) Basic model training: Use the basic model for training on each subset to obtain preliminary prediction results.
(3) Residual calculation: Calculate the residual for each sample, which is the difference between the observed value and the predicted value.
(4) Gradient enhancement: Use gradient enhancement techniques to further optimize the model. Based on the difference calculated by the residuals, update the weights of the base model to better fit the residuals.
(5) Repeated iteration: By repeating steps 2 to 4, continuously optimize the model until the preset stop condition is reached.

The EHGBA algorithm, by combining the advantages of hierarchical learning and gradient enhancement, can better adapt to complex data distributions and problem structures and improve the learning ability and generalization performance of the model. It has wide applications in many fields, such as image processing, natural language processing, and recommendation systems. Compared to other machine learning algorithms, the EHGBA algorithm has the following advantages:

(1) Efficiency: The EHGBA algorithm reduces problem space and improves learning efficiency through hierarchical learning. At the same time, the gradient enhancement technique was used to

further optimize the model, significantly improving the training speed and generalization ability of the algorithm.

(2) Robustness: The EHGBA algorithm has good robustness and can better adapt to complex data distributions and problem structures. It exhibits strong anti-interference ability when facing situations such as noise, missing values, and outliers.
(3) Generalization performance: The EHGBA algorithm further optimizes the model through gradient enhancement techniques, significantly improving the accuracy and generalization ability of the algorithm. It shows good performance in handling high-dimensional data, large-scale datasets, and multi-classification problems.
(4) Flexibility: The EHGBA algorithm can be integrated with other machine learning algorithms to form a hybrid model. It can also adapt to different application scenarios and problem types by adjusting parameters and optimization strategies.

Suppose we have a training set consisting of pairs $\left\{\left(x_1,y_1\right),\left(x_2,y_2\right),\ldots,\left(x_N,y_N\right)\right\}$, where xi is the characteristic of the ith instance and $y_i\epsilon\left\{0,1\right\}$ is its attribute. Credit scoring is implemented by ML techniques by creating a variable $F\left(x_i\right)$ that minimizes the prediction error $L\left(y_i,F\left(x_i\right)\right)$.

$$F^*=arg\,\underset{F}{min}\sum\nolimits_{i=1}^{N}L\left(y_i,F\left(x_i\right)\right) \tag{13}$$

Mathematically, Equation (13) is implemented using gradient boosting methods through an additional optimal system:

$$F\left(x\right)=\sum\nolimits_{t=1}^{T}f_t\left(x\right) \tag{14}$$

Here, T is the total number of iterations. Using Equation (14), we observe that the integration of $F\left(x_i\right)$ is performed cumulatively at each successive iteration. By iteration t, the complete loss of the aggregation generated so far, $\left\{f_j\right\}_{j=1}^{t-1}$, is even further optimized by ft. Each function f in EHGBA is realized using an ML model that may be thought of as a learner permit starting point.

The loss function is denoted as $L\left(y_i,F_{t-1}\left(x_i\right)+f_t\left(x_i\right)\right)$, and it is optimized instead at the t-th phase. Approximating the prediction error using Taylor series expansion yields the following:

$$L\left(y_i,F_{t-1}\left(x_i\right)\right)+g_if_t\left(x_i\right)+\frac{1}{2}f_t\left(x_i\right)^2 \tag{15}$$

The first derivative of the loss function, denoted by g_i, may be found by solving:

$$g_i=\left[\frac{\partial L\left(y_i,F_{t-1}\left(x_i\right)\right)}{\partial F\left(x_i\right)}\right]_{F\left(x_i\right)=F_{t-1}\left(x_i\right)} \tag{16}$$

As a result, the solution to Equation (16) may be recast as an optimal solution:

$$f_t^* = arg\,\underset{f_t}{min} \sum_{i=1}^{N} \left(f_i, \left(x_i \right) - g_i \right)^2 \tag{17}$$

Equation (17) demonstrates that the downward gradients of the error function seem to be the adaptation objective of ft. We change the learning goals of each node individually while training them in EHGBA in this way:

$$\left\{ y_i \right\}_{i=1}^{N} = -\left[\left[\frac{\partial L\left(y_i, F\left(x_i \right) \right)}{\partial F\left(x_i \right)} \right]_{F\left(x_i\right) = F_{t-1}\left(x_i\right)} \right]$$

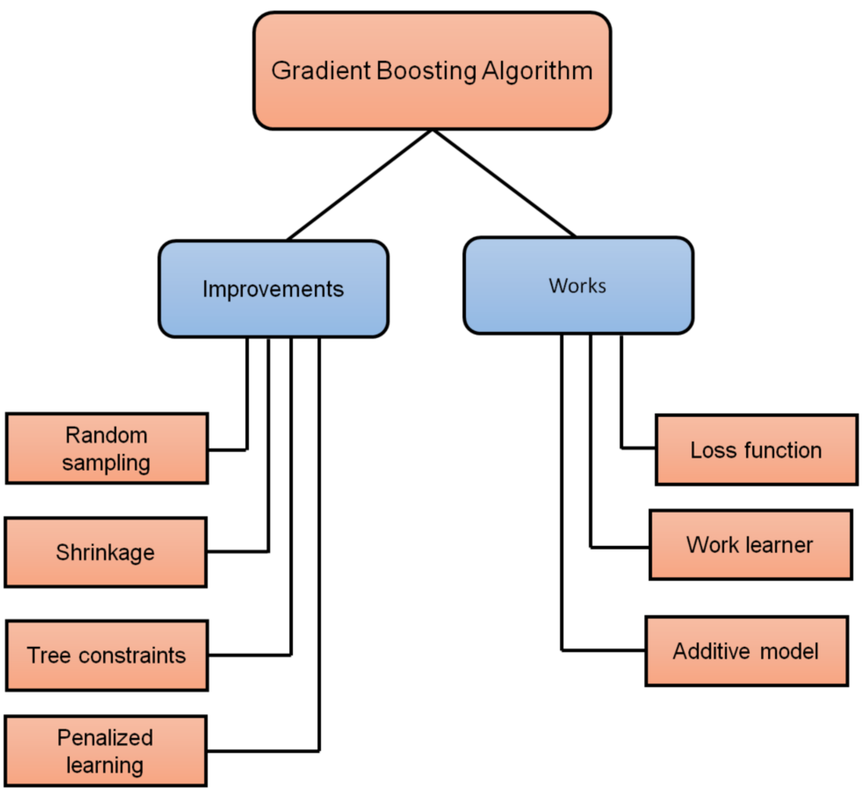
Figure 2 explains the working model of the proposed algorithm which can be used for risk prediction.

RESULTS AND ANALYSIS

Analysis of Experimental Results

The information and communications technology that allows the global network of economic operations, business dealings, and professional relationships is known as the “digital economy.” A suggested technique is the improved hierarchical based gradient boost algorithm (EHGBA). Analysis

Figure 2. EHGBA for Risk Prediction



of the system's performance has been done. The findings were compared to those obtained from more conventional methods, such as the bi-directional long short-term memory-convolutional neural network (BLSTM-CNN) (Hou et al., 2021), the digital financial inclusion, utilizing information and communications technology (DFI-ICT) (Chen et al., 2021), and the backpropagation neural network (BPNN) (Liu, 2022). The measuring measures utilized to compare the current and suggested approaches were accuracy, precision, recall, and F-measure.

The accuracy of the suggested method is seen in Figure 3. The measure to which the assessments of a quantity are closer to the value that really corresponds to that number is the accuracy of the device. The predictions that the recommended approach makes on consumer consumption are found to be more accurate when compared to the method that is currently being used. The level of accuracy is reported as a percentage of the total. The risk of inaccurate prediction is indicated both in the systems that already exist and in the system that is being proposed. In comparison, the suggested method achieves 97% accuracy, while BLSTM-CNN only reaches 66%, DFI-ICT only reaches 72%, and BPNN only reaches 87%. It demonstrates that the recommended method is more effective than the one that is currently being used. Table 1 shows the suggested method's accuracy.

The suggested method has substantially higher precision than the current approaches (refer to Figure 4). In contrast, the recommended technique achieves 93% precision, whereas BLSTM-CNN only achieves 67% accuracy, DFI-ICT only gets 83% accuracy, and BPNN only achieves 75%. This is because the risk prediction of precision in current systems has the following level. As a result, the suggested system has the highest degree of performance. The suggested method's precision is shown in Table 2.

Figure 5 shows a recall of both the proposed and existing techniques. A recall is the number of relevant incidents that were located and retrieved. Recall is another term for the sensitivity or true positive rate. The suggested strategy has the greatest degree of recall compared to the current methods. The recommended technique achieves 95% accuracy, whereas BLSTM-CNN only reaches 68%, DFI-ICT only reaches 89%, and BPNN only reaches 79%. The risk prediction of recall in current systems has the following degree of recall. Thus, a comparison of the provided way achieves 98%

Figure 3. Accuracy of Proposed and Existing Methods

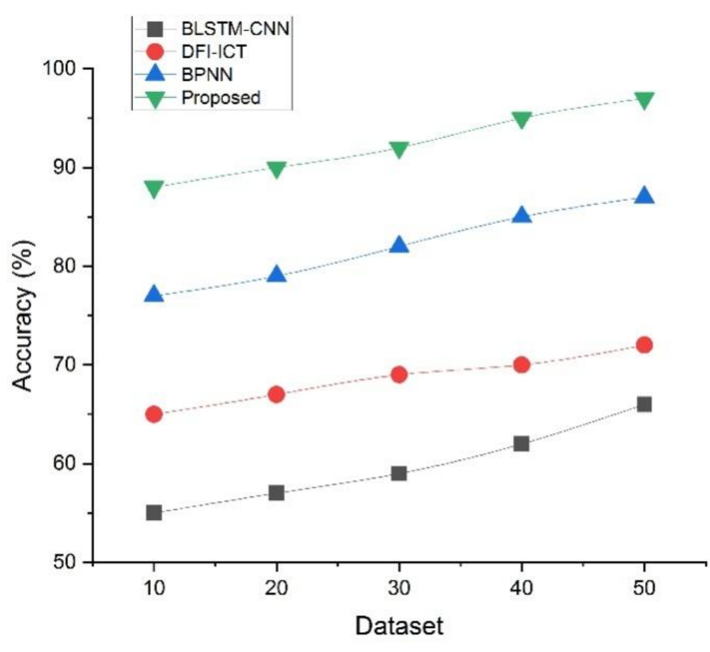


Table 1. Comparison of Accuracy

Dataset	Accuracy (%)			
	BLSTM-CNN	DFI-ICT	BPNN	Proposed
10	55	65	77	88
20	57	67	79	90
30	59	69	82	92
40	62	70	85	95
50	66	72	87	97

Figure 4. Precision of Proposed and Existing Methods

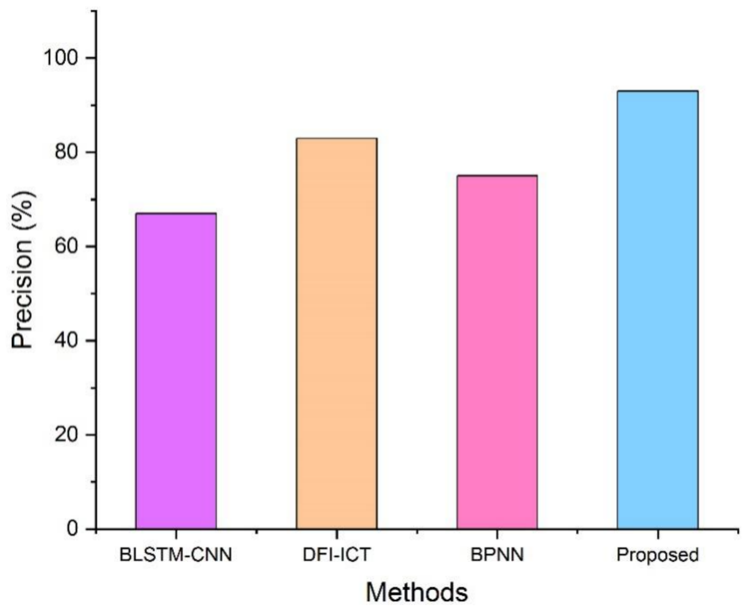


Table 2. Comparison of Precision

Methods	Precision (%)
BLSTM-CNN	67
DFI-ICT	83
BPNN	75
Proposed	93

recall. This indicates that the efficiency of the suggested task is appropriate. The suggested procedure recall is shown in Table 3.

Figure 6 shows the F1-measure for both old and new strategies. The harmonic means of a system's clarity and memory are determined and merged into a single statistic known as the F1-score. It mostly serves as a comparison of the two methods' efficacy. Higher F1 scores are seen as representing superior

Figure 5. Recall of Proposed and Existing Methods

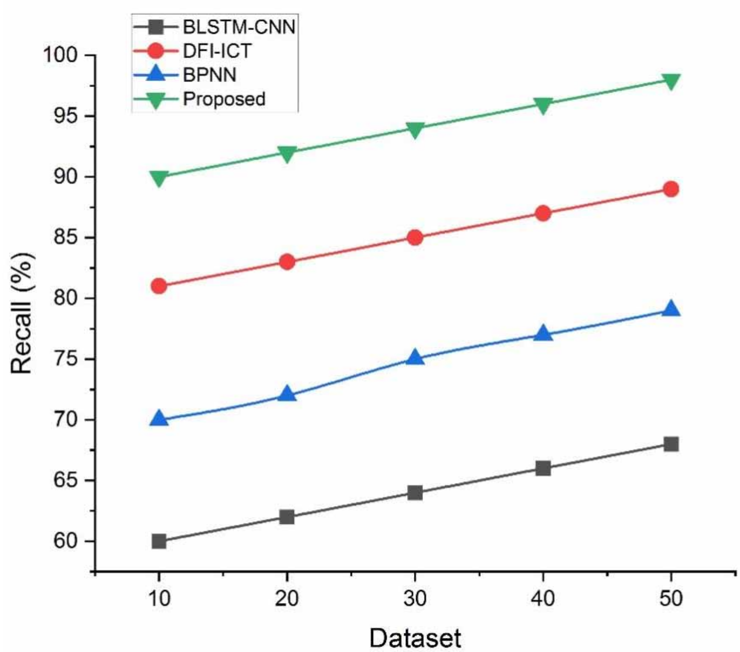


Table 3. Comparison of Proposed Methods

Dataset	Recall (%)			
	BLSTM-CNN	DFI-ICT	BPNN	Proposed
10	60	81	70	90
20	62	83	72	92
30	64	85	75	94
40	66	87	77	96
50	68	89	79	98

system performance. The suggested method has an F-measure rate of 90%, whereas BLSTM-CNN, DFI-ICT, and BPNN only manage 66%, 73%, and 81%, respectively. It implies that the suggested system performs better. The F-measure for the suggested technique is shown in Table 4.

Analysis of Practical Applications

The development of rural areas has always been an important issue, especially in the era of booming digital economy. However, due to complex terrain and uneven population distribution, rural areas face a series of challenges in infrastructure construction and digital economy promotion. In order to better understand and address these challenges, researchers have begun to explore the possibility of digital economy risk prediction in rural areas. So, how should we identify and evaluate the potential risks faced by the digital economy in rural areas? In this study, we realized that there are some limitations that need further in-depth research and exploration. The limitations of this article include the following aspects:

Figure 6. F1-Measure of Proposed and Existing Methods

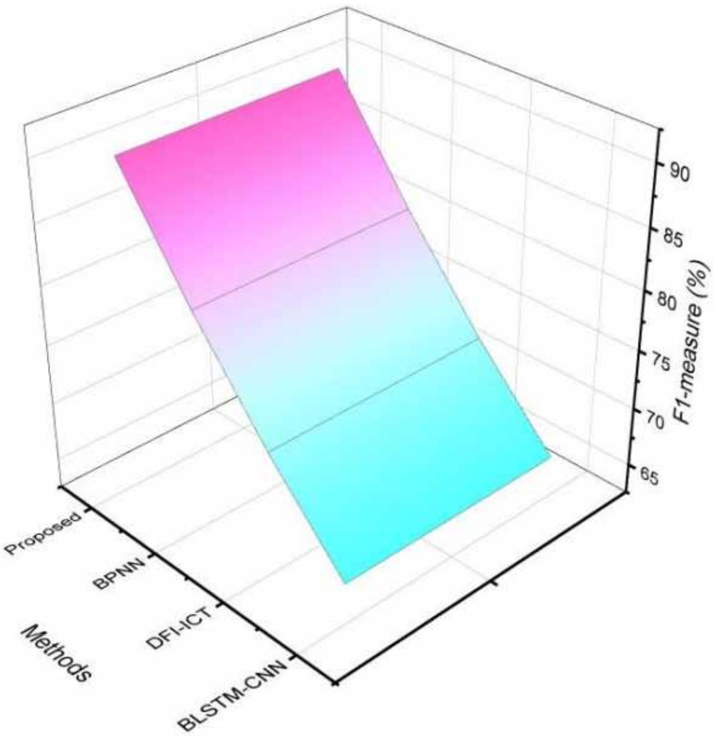


Table 4. Comparison of Proposed Methods

Methods	F1-measure (%)
BLSTM-CNN	66
DFI-ICT	73
BPNN	81
Proposed	90

(1) Limitations of method selection: This study only used gradient advancement model and principal component analysis (PCA) method to identify risks and extract features. Although these methods have certain applications in the field of digital economy, there are other algorithms and methods that can be further explored and applied. Future research can consider using more types of models and technologies, such as deep learning algorithms, support vector machines, etc., to obtain more comprehensive and accurate risk prediction results.

(2) Limitations of data collection and preprocessing: There may be some challenges in the process of data collection and preprocessing. For example, data acquisition in rural areas may be more difficult than in urban areas, and there may be uncertainty in data quality. In addition, due to the diversity and complexity of rural areas, it is necessary to consider and solve the problem of data collection more carefully to ensure the accuracy and reliability of the data. In the process of data collection and preprocessing, it is possible to consider combining multiple data sources, such as satellite remote sensing, local government data, etc., to improve data quality and coverage.

In addition, establishing sound data standards and quality control mechanisms will help ensure the accuracy and reliability of data.

(3) Limitations of regional specificity: This study analyzed the risks in rural areas, but the situation in rural areas has significant regional specificity. The level of rural development, economic structure, and social and cultural factors in different regions will have an impact on risks. Therefore, there may be differences in the applicability of research results in different rural areas, and a comprehensive analysis is needed based on specific circumstances.
(4) Limitations of research scope: This study focuses on the risk prediction of the digital economy in rural areas, but the problems faced by rural development go far beyond the digital economy. Other factors such as education, health, and environment also have a significant impact on the development of rural areas. Future research can further expand the scope of research, taking into account factors from multiple fields to obtain more comprehensive research results.

In today's digital age, the digital economy has become an important force in promoting social development and improving people's lives. However, we cannot ignore the risks and challenges faced by the digital economy in the development of rural areas. In order to better achieve sustainable development of the digital economy in rural areas, we need to comprehensively understand and evaluate its risks. The practical application of this article is mainly reflected in the following aspects:

(1) Policy formulation and planning: The development of rural areas requires targeted policy formulation and planning. By identifying and evaluating the risks of the digital economy in rural areas, the government can better understand the development potential and challenges of rural areas and accordingly formulate policies and plans to promote the healthy development of the digital economy in rural areas.
(2) Investment decision-making and resource allocation: Predicting the risks of the digital economy in rural areas can provide reference for investors and help them make wise investment decisions. At the same time, the government and relevant departments can also allocate resources reasonably based on risk prediction results to support the development of digital economy in rural areas.
(3) Risk management and emergency response: By analyzing the risks of the digital economy in rural areas, relevant departments and organizations can develop effective risk management strategies to reduce the potential impact of risks on the digital economy in rural areas. At the same time, in case of emergencies, timely emergency response and resource allocation can be carried out based on the risk prediction results.
(4) Development guidance and capacity building: Risk prediction results can provide guidance and decision-making support for the development of rural areas. By comparing and analyzing the risks in different regions, specific development paths and capacity building plans can be provided for rural areas, promoting the sustainable development of the digital economy in rural areas.

CONCLUSION

In summary, variables such as complex terrain and unequal population distribution have indeed exacerbated the challenges of building infrastructure in rural areas. Meanwhile, issues such as poor information transmission and low communication efficiency also hinder the further promotion of the digital economy in rural areas. In this study, we employed a gradient advancement model to identify potential risks to the growth of the digital economy sector related to rural revitalization. By standardizing the collection and preprocessing of the dataset, we can ensure the accuracy and comparability of the data. Meanwhile, the application of principal component analysis (PCA) has helped us extract key features for risk prediction in the digital world. These methods provide important support for us to better understand the risk factors that the digital economy may face in the development of rural areas.

In addition, this study also introduced the enhanced hierarchical gradient boosting algorithm (EHGBA) to improve the accuracy of risk assessment. The optimization of this technology will help establish more comprehensive and reliable risk prediction models, providing guidance for the development and decision-making of digital economy in rural areas.

Through the implementation of this study, we have provided new perspectives and solutions for infrastructure construction and digital economy promotion in rural areas. Our research findings will make a positive contribution to the sustainable development and economic growth of rural areas. However, we also need to be aware of the limitations of this study. For example, our research is limited to using gradient advancement models to identify risks, and there are other algorithms and methods that can be further explored and applied. In addition, we may face some challenges in data collection and preprocessing, which require more careful consideration and resolution.

In order to promote sustainable development in rural areas and the flourishing development of the digital economy, we encourage future researchers to conduct more in-depth research in this field. Through joint efforts, we believe that we can find more effective methods and strategies to promote modernization and prosperity in rural areas.

The development of the digital economy in rural areas faces many risks and challenges. In order to better achieve the sustainable development of the digital economy in rural areas, it is necessary to deeply explore its risk identification and assessment issues and explore future research directions. Future research directions can include the following aspects:

(1) Optimization of risk prediction models: Further improve and optimize the risk prediction models of the digital economy in rural areas, enhancing their accuracy and reliability. We can explore using more data sources and introducing machine learning algorithms and artificial intelligence technologies to improve the accuracy and timeliness of risk prediction.
(2) Risk management and prevention strategies: Conduct in-depth research on the risk management and prevention strategies of the digital economy in rural areas, including formulating corresponding policies and regulations, establishing effective regulatory mechanisms, providing training and support services, etc. At the same time, it is also possible to explore the construction of emergency response systems for digital economy in rural areas and strengthen the ability to handle unexpected risk events.
(3) Development path and policy support: Further research can be done on the development path and policy support of the digital economy in rural areas, including providing comprehensive guidance and support from aspects such as industrial structure adjustment, technological innovation, and talent cultivation. Focus on the development characteristics and needs of the digital economy in rural areas, formulate differentiated policies and measures, and promote the sustainable development of the digital economy in rural areas.
(4) Social impact and sustainability assessment: Study the social impact and sustainability of the digital economy in rural areas, including assessments of rural residents' living standards, employment opportunities, environmental protection, and other aspects. Comprehensively evaluating the impact of the digital economy on rural areas provides reference for the government and relevant departments to formulate more comprehensive and sustainable development strategies.

Future research directions may focus on optimizing risk prediction models, risk management and prevention strategies, development paths and policy support, as well as social impact and sustainability assessments, thereby providing more scientific and effective support for the development of the digital economy in rural areas.

DATA AVAILABILITY

The figures and tables used to support the findings of this study are included in the article.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

FUNDING STATEMENT

This work was supported by Research and Practice on the Training Path of Management Talents in Supply Chain Enterprises in Liaoning Province [JYTMS20231864]; General project of basic scientific research project of Liaoning Provincial Department of Education in 2023 (unveiling the list and serving local projects).

ACKNOWLEDGMENT

The authors would like to show sincere thanks to those techniques that have contributed to this research.

PROCESS DATES

Received: 12/24/2023, Revision: 1/29/2024, Accepted 2/4/2024

REFERENCES

Chen, Y., Kumara, E. K., & Sivakumar, V. (2021). Investigation of finance industry on risk awareness model and digital economic growth. *Annals of Operations Research, volume*(issue), 1-22.

Claveria, O., Monte, E., & Torra, S. (2017). Using survey data to forecast real activity with evolutionary algorithms. A cross-country analysis. *Journal of Applied Econometrics*, *20*(2), 329–349.

Claveria, O., Monte, E., & Torra, S. (2019). Evolutionary computation for macroeconomic forecasting. *Computational Economics*, *53*(2), 833–849. doi:10.1007/s10614-017-9767-4

Guo, C. X., & Miao, Y. F. (2022). The mechanism and path of digital economy promoting rural industry revitalization. *Journal of Beijing University of Technology, volume*(issue), page numbers.

Heng, S. & Xu, Y. (2023), Research on rural e-commerce poverty alleviation under the background of rural revitalization strategy, *Electronic Commerce, 2*(2020).

Henrique, B. M., Sobreiro, V. A., & Kimura, H. (2019). Literature review: Machine learning techniques applied to financial market prediction. *Expert Systems with Applications*, *124*, 226–251. doi:10.1016/j.eswa.2019.01.012

Hou, H., Tang, K., Liu, X., & Zhou, Y. (2021). Application of artificial intelligence technology optimized by deep learning to rural financial development and rural governance. [JGIM]. *Journal of Global Information Management*, *30*(7), 1–23. doi:10.4018/JGIM.289220

Huateng, M., Zhaoli, M., Deli, Y., & Hualei, W. (2021). *The Chinese digital economy*. Springer Singapore. doi:10.1007/978-981-33-6005-1

Johnson, R. C., & Noguera, G. (2012). Accounting for intermediates: Production sharing and trade in value added. *Journal of International Economics*, *86*(2), 224–236. doi:10.1016/j.jinteco.2011.10.003

Lee, P. S., Leung, L., Lo, V., Xiong, C., & Wu, T. (2011). Internet communication versus face-to-face interaction in quality of life. *Social Indicators Research*, *100*(3), 375–389. doi:10.1007/s11205-010-9618-3

Li, M. Z. (2022). Research on innovation and development of rural digital supply chain financial model under the background of rural revitalization. *Journal of Agricultural Economics*, *2022*(08), 120–122.

Liu, C. (2022). Risk prediction of digital transformation of manufacturing supply chain based on principal component analysis and backpropagation artificial neural network. *Alexandria Engineering Journal*, *61*(1), 775–784. doi:10.1016/j.aej.2021.06.010

Ma, D. (2020). Research on the countermeasures of rural e-commerce in helping poverty alleviation. *Rural Economy and Technology, 31*(2020).

Obthong, M., Tantisantiwong, N., Jeamwatthanachai, W., & Wills, G. (2020). A survey on machine learning for stock price prediction. *Algorithms and Techniques, volume*(issue), page numbers.

Petropoulos, A., Siakoulis, V., Stavroulakis, E., & Vlachogiannakis, N. E. (2020). Predicting bank insolvencies using machine learning techniques. *International Journal of Forecasting*, *36*(3), 1092–1113. doi:10.1016/j.ijforecast.2019.11.005

Ptak-Chmielewska, A. (2019). Predicting micro-enterprise failures using data mining techniques. *Journal of Risk and Financial Management*, *12*(1), 30. doi:10.3390/jrfm12010030

Schwartz, Z. P. (2022). Fascism: Transnational frameworks. *The Webster Review of International History, 2*(1).

Tao, H. (2018). No. 1 central document and Rural E-commerce Development. *China National Conditions and Strength*, *4*, 24–26.

Vadlamudi, S. (2020). The impacts of machine learning in financial crisis prediction. *Asian Business Review*, *10*(3), 171–176. doi:10.18034/abr.v10i3.528

Wang, J. (2020). The historical evolution and future prospects of rural payment methods in the new era. *Journal of the Party School of the Taiyuan Municipal Committee of the Communist Party of China, 4*(2020).

Wei, J., Cheng, H., Fan, B., Tan, Z., Tao, L., & Ma, L. (2020). Research and practice of "one opening-one closing" productivity testing technology for deep water high permeability gas wells in South China Sea. *Fresenius Environmental Bulletin*, *29*(10), 9438–9445.

Wen, H., Lee, C. C., & Song, Z. (2021). Digitalization and environment: How does ICT affect enterprise environmental performance? *Environmental Science and Pollution Research International*, *28*(39), 54826–54841. doi:10.1007/s11356-021-14474-5 PMID:34014481

Wen, L., & Zhang, R. C. (2022). Research on the influence mechanism of digital economy enabling rural revitalization. *Journal of Henan University of Science and Technology*, *40*(04), 25–30.

Yao, J. W., & Liu, X. C. (2022). The influence of digital economy on county economic growth under the background of rural revitalization strategy. *Journal of Commercial Economy*, *2022*(11), 137–139.

Zhang, C., Zhang, Y., Li, Y., & Li, S. (2023). Coupling coordination between fintech and digital villages: Mechanism, spatiotemporal evolution and driving factors—An empirical study based on China. *Sustainability (Basel)*, *15*(10), 8265. doi:10.3390/su15108265

Zhang, J. (2014). Research on the promotion path of non- cash payment tools in rural areas. *Chinese Market, 43*(2014).

Zhang, J. P. (2022). Empowering rural areas in ethnic minority areas with digital economy. *Journal of Contemporary Guangxi*, *2022*(16), 25.

Zhao, B., Wang, W., & Li, H. D. (2022). Research on the mechanism and path of new e-commerce helping rural revitalization under digital economy. *Journal of Shijiazhuang University*, *24*(05), 65–68.

Rui Luan, female, Han nationality, research direction: economy, trade and logistics management.

Ping Xu was born in Yueyang, Hunan Province, China in 1970, he received a master's degree from the University of Science and Technology of China. Now he is working in the College of Science, Xihua University, Chengdu, Sichuan. His main research field is digital economy.