

RESEARCH

Open Access



The analysis of financial market risk based on machine learning and particle swarm optimization algorithm

Tao Liu¹ and Zhongyang Yu^{2*}

*Correspondence:

zhongyang.yu@vonechain.com

² Shanghai Vonechain

Information Technology Co.,
Ltd, Shanghai 200443, China

Full list of author information
is available at the end of the
article

Abstract

The financial industry is a key to promoting the development of the national economy, and the risk it takes is also the largest hidden risk in the financial market. Therefore, the risk existing in the current financial market should be deeply explored under blockchain technology (BT) to ensure the functions of financial markets. The risk of financial markets is analyzed using machine learning (ML) and random forest (RF). First, the clustering method is introduced, and an example is given to illustrate the RF classification model. The collected data sets are divided into test sets and training sets, the corresponding rules are formulated and generated, and the branches of the decision tree (DT) are constructed according to the optimization principle. Finally, the steps of constructing the branches of DT are repeated until they are not continued. The results show that the three major industries of the regional economy account for 3.5%, 51.8%, 3.2%, 3.4%, and 3.8% of the regional GDP, respectively, the secondary industry makes up 44.5%, 43%, 45.1%, 44.8%, and 43.6%, respectively, and the tertiary industry occupies 20%, 3.7%, 52.3%, 52.9%, 54%, and 54.6%, respectively. This shows that with the development of the industrial structure under BT, the economic subject gradually shifts from the primary industry to the tertiary industry; BT can improve the efficiency of the financial industry and reduce operating costs and dependence on media. Meanwhile, the financial features of BT can provide a good platform for business expansion. The application of BT to the supply chain gives a theoretical reference for promoting the synergy between companies.

Keywords: Machine learning, Random forest, Clustering method, Financial market, Blockchain

1 Introduction

With the popularity of the e-commerce platform [1–3] in the international market, almost all software applications involve capital transactions. This brings great convenience for people, but also huge risks. In the new era, the financial market [4–6] has to transform its traditional operational mode and adapt itself to the new trend, and the risk of the financial industry [7–9] is also increasing. Chen et al. (2020) [10] argued that deep learning (DL) could detect possible risks in finance, and blockchain technology (BT) [11–13] can settle the disputes between e-commerce platforms and

users. Because it can store the data dispersedly and complete point-to-point transactions, BT is widely used in many experiments, such as securities, banking, and insurance [9, 14, 15]. It is a distributed shared ledger and database and has the characteristics of decentralization, non-tampering, and traceability. Therefore, it can significantly curb the hidden dangers and risks brought by the explosive growth and expansion of the financial market, and solve the problems in the financial field such as high credit risk, low capital utilization efficiency, and high payment processing cost.

Scholars all over the world have done a lot of research on the risk of financial markets. Financial risks [16–18] are the loss of an entity's activities in the financial field. Relevant scholars have confirmed that a portfolio can reduce the probability of risks. And foreign scholar Markowitz once proposed the mean–variance model, but the shortcomings of this model are also obvious. Its setting conditions are too strict about adapting to the changing financial market. Subsequently, British economists, Fishburn, and Stone improved and optimized the mean semi-variance model, but the model is only suitable for ordinary scenarios. Japanese scholars have studied and introduced the first-order absolute deviation method of portfolio theory and used this method as the standard to measure risks. Some domestic scholars use generalized autoregressive conditional heteroskedasticity (GARCH) to study financial risks, and their research provides a solution for the development of financial markets. Zhao et al. (2021) [19] studied the relationship between financial risks and global climate change and evaluated the relationship between global financial risks and carbon dioxide emissions. The results show that technological innovation and financial risk significantly affect global carbon emissions in the 10 quantiles. Kim et al. (2020) [20] conducted a case study on predicting financial risk behavior, extracting features from structured data for DL. The results show that DL can classify traders' financial risks more accurately. Yang et al. (2019) [21] applied and studied the financial risk management model of the Internet supply chain based on data science. The empirical results show that the model has high accuracy in data evaluation. In short, the research on financial risk assessment is carried out based on hedging combinations. And relevant scholars have studied and constructed a portfolio model with transaction costs and put forward specific data solution methods and required conditions.

As for the risk in international financial markets, most relevant researchers also introduced some methods to reduce the risk. However, there is little research on preventing financial market risks based on BT, and there is no relevant research under machine learning (ML) and particle swarm optimization (PSO). Therefore, the research on preventing financial market risks based on BT, ML, and PSO will be carried out. The innovation is that artificial intelligence (AI) and DL are used to find possible risks, and PSO and BT analyze the risk of financial markets. The research content provides a theoretical basis for subsequent research and has great significance. Its structure is as follows: the first section introduces the research background and relevant literature; the second displays PSO and maximum likelihood clustering algorithm; the third analyzes the experimental results using the maximum likelihood algorithm, BT and PSO; the fourth summarizes the empirical conclusion.

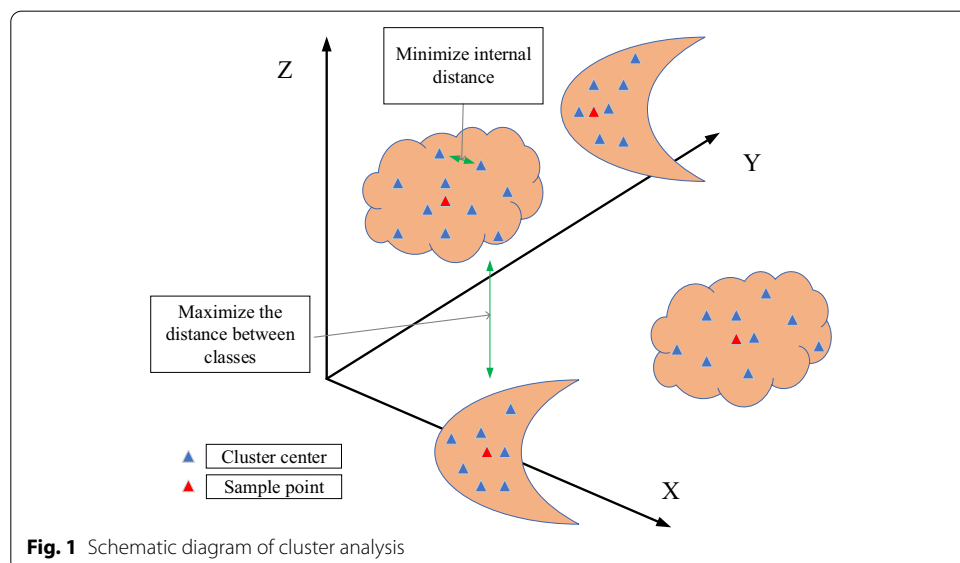
2 Method

2.1 The clustering analysis algorithm based on ML and DL

ML algorithm is an interdisciplinary subject in many fields, and there are many related subjects, such as statistics, algorithm complexity theory, and probability theory. ML is a learning algorithm based on simulating human behavior to further obtain new knowledge and information. Moreover, it can continuously improve its performance in the process of operation. From the academic perspective, DL belongs to ML. In other words, DL is actually a special case of ML. The power and flexibility of the DL function are to learn and then express the real world in the way of nested concepts. A more specific expression is that, in the DL, any definition can find its simplified definition, and DL can also concretize and calculate the abstract things.

The ML algorithm [22–24] used here consists of two parts. The first part is the clustering method, and the second part is the RF [25–27] classification model. The clustering method learns from the unlabeled data and then further analyzes and interprets the internal relationship of the data. Specifically, the concept of set in a mathematical model is adopted. Initially, classification is conducted on the samples and multiple disjoint subsets. Then, the samples in each subset are basically similar, and the samples between different sets are different. The mathematical model of clustering method can be interpreted in this way. The set of samples is set as D , which contains many unlabeled samples. The method of multi-dimensional eigenvector is used to express a single sample. On this basis, the next step of clustering method is to find K subset C in the set of samples, in which each of K subsets is called class cluster. Figure 1 illustrates the scheme of clustering method analysis.

Based on the above description and the principle analysis diagram of clustering [28–30], the collected samples will be selected and matched according to the fitness of each class cluster, which will avoid that each sample does not have the same class cluster. Therefore, this sample set can be clustered. Equation (1) expresses the specific calculation of the objective function.



$$\min D = \sum_j \sum_{x_i \in C_j} d(x_i, y_j) \tag{1}$$

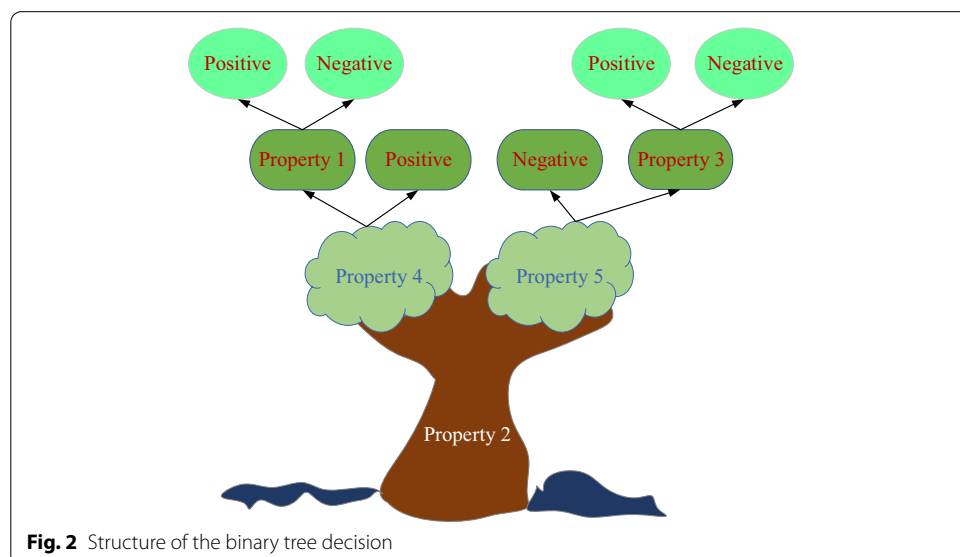
In Eq. (1), y_j means the centroid of the cluster C_j . The function $d(x, y)$ refers to a similarity measurement function, which is used to measure the similarity of sample x and centroid y . To obtain a similar relationship between a single sample set with another set, and summarize the characteristics of the data in the set, the clustering method is indeed a common method with extraordinary function.

The generation of RF algorithm is based on random subspace. Researchers carried out repeated sampling experiments in the random subspace many times, and finally proposed the RF algorithm. RF has a unique method for the completion of learning tasks. For the centralized learning model of DT, a large number of classifiers can be constructed and combined according to a certain method, which is the reason why the RF can complete the learning task well. In the experiment, a new sample set is generated according to the self-help sampling method and the randomly selected samples. Figure 2 depicts the final structure of the binary tree decision.

The reason for the great difference between learners may be the randomness of the collected samples, which will have a great impact on the whole learning result. Candidate feature sets are randomly selected in the current feature set when a single DT performs optimal feature segmentation. Randomness can be controlled by controlling the size of the feature subset. Equation (2) is used to calculate the size of the frequently used subset.

$$S = \lfloor \log_2 M \rfloor + 1 \tag{2}$$

In Eq. (2), M means the total number of current node characteristics. Equation (3) denotes the expression of the eigenvector of the random training sample. In Eq. (3), r represents the characteristic dimension.



$$A = \{t_1, t_2, \dots, t_r, y\} \tag{3}$$

Equation (4) reveals the expression of multiple classifiers. In Eq. (4), index l means the number of classifiers. According to the principle of ensemble learning, the voting result can be obtained as shown in Eqs. (5) and (6), where $H(x)$ represents the classification effect evaluation function of the classifier.

$$\{h(x, \theta_k), k = 1, 2, \dots, L\} \tag{4}$$

In Eq. (4), L represents the number of base classifiers. According to ensemble learning, the final voting result can be obtained by:

$$H(x) = \operatorname{argmax} \sum_{i=1}^L I(h(x, \theta_i) = y) \tag{5}$$

$$I(x = y) = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{if } x \neq y \end{cases} \tag{6}$$

In Eqs. (5) and (6), x represents the sample and y represents the centroid. Construction of RF model consists several steps: Firstly, the collected data sets should be classified into two parts: test set and training set; secondly, corresponding rules are formulated and produced according to certain methods; thirdly, the branch of DT is constructed according to the optimal principle; fourthly, the corresponding branch is repeatedly branched until the branching cannot be continued. Figure 3 demonstrates the detailed process of RF generation.

According to the principle of DT, the prediction results of DT are independent from each other. Besides, due to the guidance of ensemble learning, the performance of DT is obviously worse than that of RF. From many important perspectives, the performance of RF is better.

Because there are many characteristic attributes in the data set, some of them are useful and the others are useless. Therefore, they are classified into two categories. The first category is related features, and the other category is irrelevant features. Based on this situation, the Gini coefficient is used to select the characteristics of the data. The score of importance features can be marked as VIM (very importance score), and the evaluation

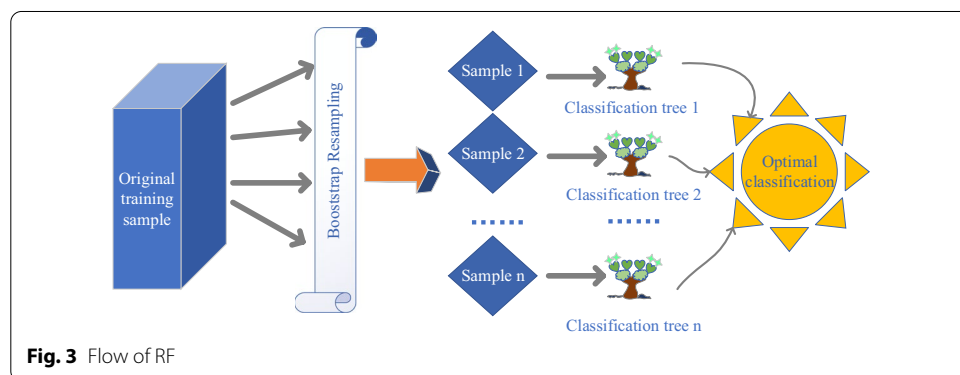


Fig. 3 Flow of RF

score of each feature can be marked as VIM_i , which means the mean value of the i th feature with respect to the change in the purity of node splitting in the DT. Equation (7) displays the calculation process.

$$GI_m = \sum_{k=1}^{|K|} p_{mk}(1 - p_{mk}) \tag{7}$$

In Eq. (7), K means the number of sample categories, p_{mk} is the specific gravity of node m on the k_{th} sample. On this basis, Eq. (8) demonstrates the importance of feature t_i with respect to node m .

$$VIM_{im}^{Gini} = GI_m - GI_l - GI_r \tag{8}$$

In Eq. (8), GI_l and GI_r are the Gini coefficients of the new node after branching based on the node m . On this basis, Eq. (9) reveals the importance of characteristic t_i with respect to the number of the j_{th} tree.

$$VIM_{ji}^{Gini} = \sum_{m=1}^N VIM_{im}^{Gini} \tag{9}$$

To sum up, Eq. (10) demonstrates the importance of feature t_i on the overall RF.

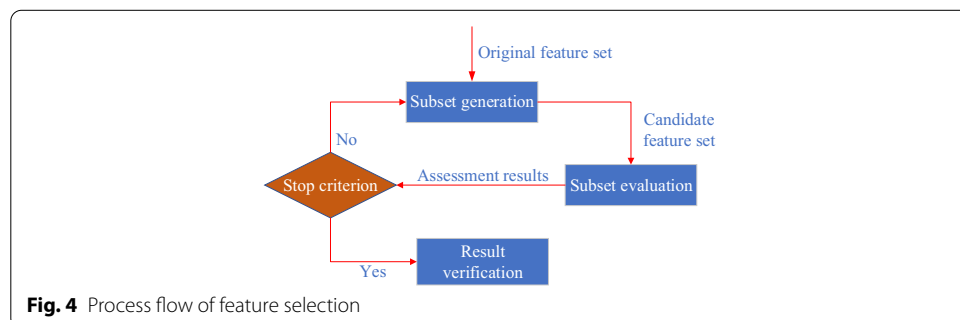
$$VIM_i^{Gini} = \frac{1}{n} \sum_{j=1}^n VIM_{ji}^{Gini} \tag{10}$$

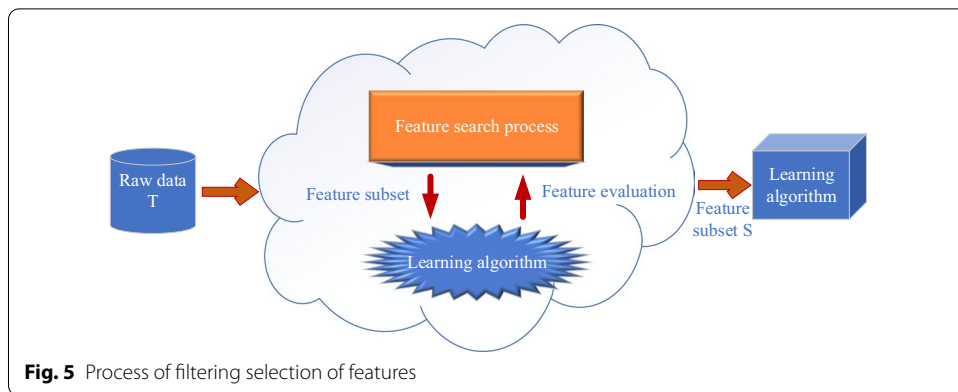
In Eq. (10), n means the number of base classifiers in the RF.

Feature selection mainly includes four aspects. The first aspect is subset generation, the second aspect is subset evaluation, the third aspect is stop criterion, and the fourth aspect is result verification. The generation of each subset is equivalent to the search process, in which the appropriate optimal subset is selected step by step. Figure 4 refers to the process flow of feature selection.

At first, the filtering method selects the features of the data set, and then trains the learner. The advantage of filtering selection is that it can quickly remove useless noise. Figure 5 presents the process of filtering selection of features.

The wrapping method first improves the learner, and then carries out multiple trainings. The selecting effect of this method is obviously better than that of filtering



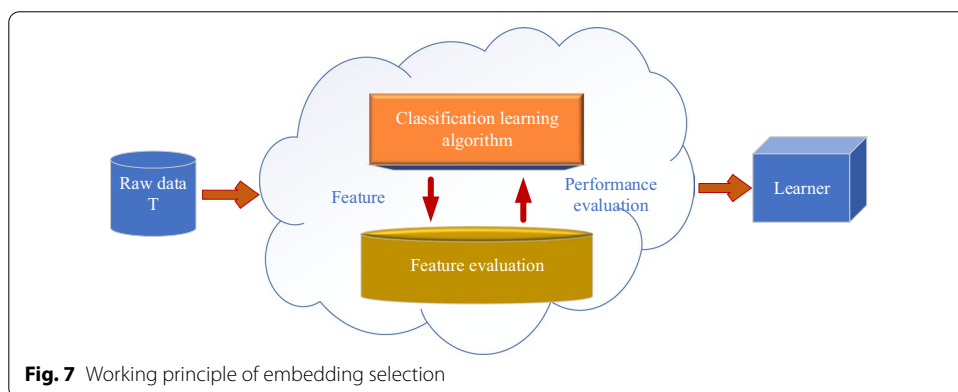
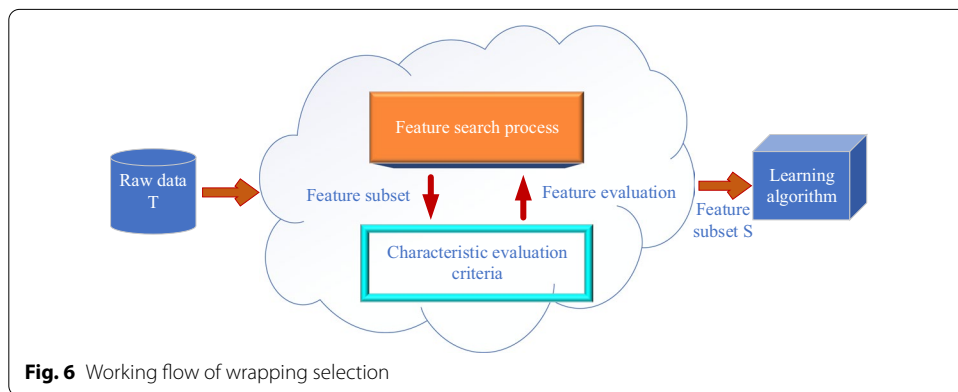


selection, but its calculation process is complex and needs a lot of time. Figure 6 depicts the working flow of the wrapping selection.

Embedding selection is obviously different from filtering selection and wrapping selection. Processes of embedding and learner training are mixed together, so they can improve the effect simultaneously. Figure 7 illustrates the working principle of embedding selection.

2.2 Particle swarm optimization algorithm

PSO is a heuristic swarm intelligence algorithm proposed by Dr. Eberhart and Dr. Kennedy in 1995 to simulate bird predation behavior. It is assumed there are different food



sources in a region, and the task of birds is to find the largest food source (global optimal solution). Throughout the search, birds communicate information about each other's location so that other birds know where the food comes from. Finally, the entire bird swarm can gather around the largest food source, where the optimal solution is found, and the problem converges. The specific strategies are as follows:

Firstly, each bird sets out randomly in one direction to find a food source.

Secondly, each bird shares the optimal food source and food stock found by itself to the bird group after flying for one minute, and then calculates the optimal position of the group to find the optimal food source.

Thirdly, each bird looks back on its path, considering its optimal location and group optimal location to determine the next direction.

Fourthly, if every bird is near the same food source, they should stop looking, otherwise continue to repeat the second and third steps.

In the scale of particle swarm [31–33], a single particle is normally a multi-dimensional vector, and a means the maximum velocity of the artificially set particle. The adjustment of the position and velocity of each particle in the particle population depends on Eqs. (11) and (12).

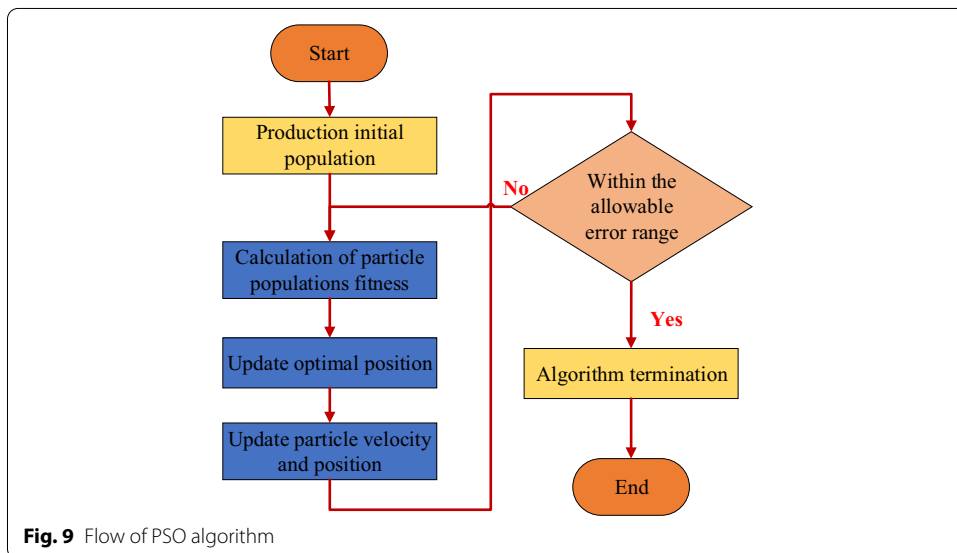
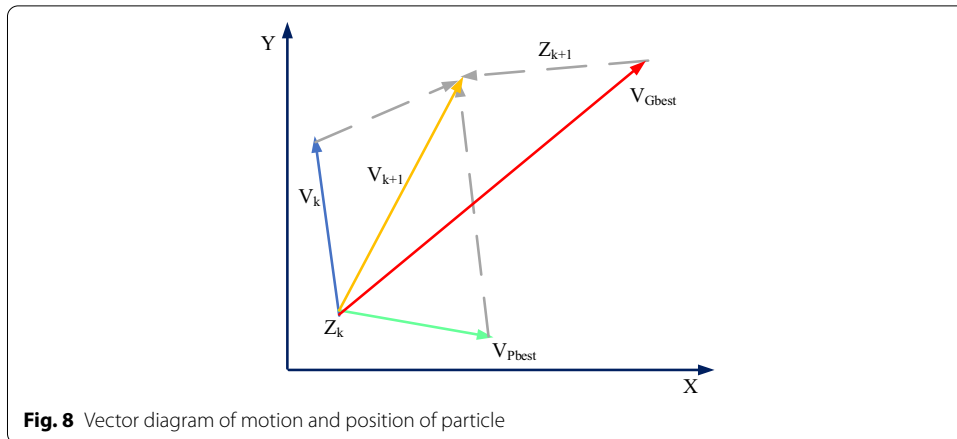
$$\begin{aligned} v_{id}^{t+1} &= \omega v_{id}^t + c_1 r_1 (p_{id}^t - x_{id}^t) \\ &\quad + c_2 r_2 (p_{gd}^t - x_{id}^t) \text{Prob}(\Delta p \leq -VaR) \\ &= 1 - \alpha \end{aligned} \quad (11)$$

$$x_{id}^{t+1} = x_{id}^t + v_{id}^{t+1} \quad (12)$$

In Eqs. (11) and (12), v_{id}^{t+1} represents the velocity of the particle. ωv_{id}^t represents the current state of the particle; $c_1 r_1 (p_{id}^t - x_{id}^t)$ is a part often called “cognition,” which mainly reflects the thinking and continuous cognition of particles on their own flight process when they fly; $c_2 r_2 (p_{gd}^t - x_{id}^t) \text{Prob}(\Delta p \leq -VaR)$ is the social part, and reflects the learning ability and particle following when flying. As for $d=1,2, \dots, D$, $i=1,2, \dots, NP$, $t=1,2, \dots$ represents the times of iterations of the algorithm model. If t is the current generation, $t+1$ represents the next generation. ω (0,1) represents inertia weight, and it is a random number. c_1 and c_2 are acceleration constants and they usually random numbers between (0, 2). c_1 is used as self-learning factor and c_2 is social learning factor. r_1 and r_2 are pseudo-random numbers between [0, 1].

According to the description of the Equation, Fig. 8 depicts the particles' position after each time of the particles' motion. Figure 8 is set as a two-dimensional space. Particles start out from the point Z_k . Z_{k+1} indicates the position of the particles after motion. K means the current velocity of the particle, and $K+1$ means the velocity of the particle after motion. In the process of moving back and forth of particle swarm, the speed and position of particles are also changing, and this process is the program to obtain the optimal solution of PSO algorithm [34–36]. Unless the optimal solution is obtained, the particle will move continuously. Figure 8 is the vector diagram of motion and position of particle.

The larger the number of particles moving back and forth in the particle swarm, the longer the reaction time, which indicates that the particle swarm is large. The specific



flow of PSO algorithm is as follows: primarily, a subset of PSO is generated in the original feature set; then, the set of candidate features is selected in these subsets; afterwards, these subsets are evaluated and the evaluation results are obtained to judge whether the evaluation result meets the stop criterion; finally, decision is made on whether to verify the evaluation results according to the judgment, or continue to enter the step of subset generation and circulate operations until an appropriate subset is selected. Figure 9 denotes the flow of subgroup algorithm.

Equation (13) expresses the model of PSO algorithm. In Eq. (13), $f(x)$ is a continuous function in space, which represents there is no need to make constrained optimization.

$$\min_{x \in R^n} f(x), \quad R = \left\{ x \mid a_i^T x \leq b, \quad i = 1, 2, \dots, m \right\} \tag{13}$$

If, $f(x) \in C1$, any x is full rank in $f(x)$. $H = H(x)$ refers to a continuous symmetric matrix function, and it is bounded and uniformly positive definite on R .

$$\eta_1 \|y\|^2 \leq y^T H(x) y \leq \eta_2 \|y\|^2 \tag{14}$$

In Eq. (14), $\|\cdot\|$ stands for the 2-norm of the vector. Equations (15–19) refer to the specified marks at point x .

$$g(x) = -\nabla f(x) \tag{15}$$

In Eq. (15), $g(x)$ represents the gradient function, and $-\nabla f(x)$ represents the negative gradient direction. For particle swarm optimization matrix, x denotes each element in the matrix, and ∇ indicates the matrix diagonalization.

$$Q(x) = \text{diag}(b_1 - a_1^T x, b_2 - a_2^T x, \dots, b_m - a_m^T x) \tag{16}$$

For the continuous variables in space, Eq. (17) signifies the unitized data.

$$B(x) = (A^T H A + Q(x))^{-1} A^T H \tag{17}$$

In Eq. (17), $A^T H A$ denotes the search matrix, and $B(x)$ represents the weight value of the search matrix. Equation (18) indicates the calculation of the weight vector.

$$u(x) = B(x)g(x) = (u_1, u_2, \dots, u_m)^T \tag{18}$$

Concurrently, Eq. (19) expresses the relationship between the weight function with the objective function and the search matrix.

$$P(x) = H(E - AB(x)) = H \left| E - A(A^T H A + Q(x))^{-1} A^T H \right| \tag{19}$$

Equation (20) illustrates the calculation of the function to solve the gradient.

$$v_j = \begin{cases} -\Phi_1(\lambda - u_j), & u_j \leq 0 \\ -\Phi_2(u_j)(a_j^T x - b_j), & u_j > 0 \end{cases} \tag{20}$$

In Eq. (20), there are two parameters, which are $u_j \leq 0, u_j > 0$.

Equations (21) and (22) describe the relationship between gradient descent variables and scalar values.

$$d^k = P_k g^k + B_k^T v^k \tag{21}$$

$$(g^k)^T d^k = 0 \tag{22}$$

In Eq. (21), d^k stands for the value of the direction where gradient values fall fastest. Equations (23) and (24) demonstrate the relational expression of gradient descent.

$$x^{k+1} = x^k + \lambda_k d^k \tag{23}$$

$$\lambda_k = \text{argminf} \{ x^k + \lambda d^k \mid x^k + \lambda d^k \in R \} \tag{24}$$

In Eqs. (23) and (24), x^k denotes the input value and λ_k represents the coefficient of differential operator. Equation (25) illustrates the limitations on the object function.

$$\begin{cases} x^k + \lambda d^k \in R \\ f(x^k) - f(x^k + \lambda d^k) \geq \tau \lambda (g^k)^T d^k \end{cases} \quad (25)$$

$$H = H(x) = \nabla^2 f(x) \quad (26)$$

$$H(x) = \nabla^2 f(x) + \mu E (\mu > 0) \quad (27)$$

In Eqs. (26) and (27), $f(x)$ is a second-order continuous and strictly convex function; then, $H(x)$ is taken as the Hessian matrix of the $f(x)$.

3 Results and discussion

3.1 Financial risk analysis based on maximum likelihood algorithm and BT

Figure 10 shows the regional gross domestic product (GDP) and year-on-year growth rate in 2020. The national annual economic growth data are collected and sorted on the website of the National Bureau. According to the statistical results of economic development in a certain region of China, the GDP of the primary industry is 5.5 billion yuan, an increase of 3.5% over the previous year; GDP of the secondary industry region is 6 billion yuan, an increase of 7.2% over the previous year; the total output value of the tertiary industry is 8.8 billion yuan, an increase of 8.9% over the previous year.

GDP of the three major industries increases in turn in 2020, of which GDP of the tertiary industry is the highest and that of the primary industry is the lowest. In addition, the year-on-year growth rate of the three industries also increases. The year-on-year growth rate of the tertiary industry is the highest and that of the primary industry is the lowest. This shows that the economic center begins to transfer from the primary to the tertiary industry gradually.

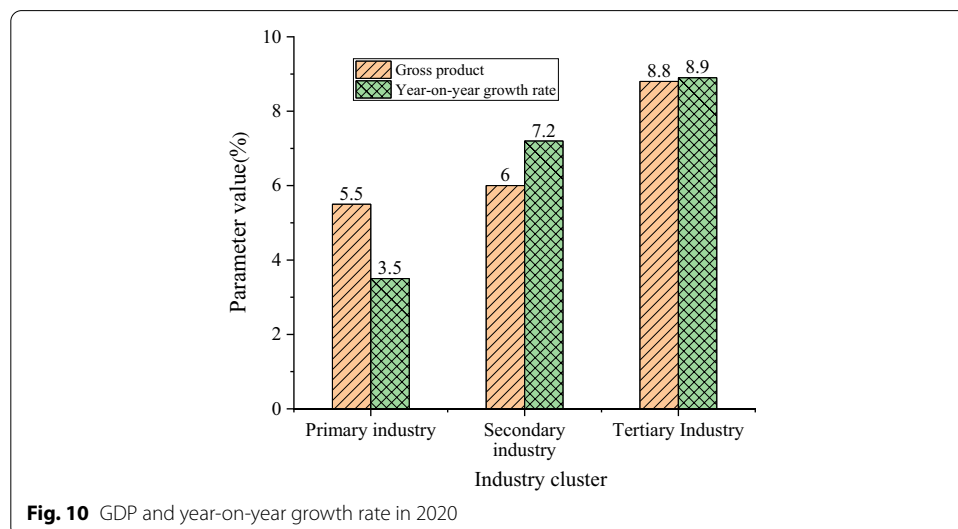


Figure 11 shows the growth rate of regional financial investment. The specific situation of regional financial investment from 2016 to 2020 is as follows: the growth rate of fixed assets and real estate investment in 2016 are 1.7% and 2.1%, respectively; in 2017, they are 5.7% and 6%, respectively; in 2018, they are 14% and 13%, respectively; in 2019, they are 12% and 13%, respectively; in 2020, they are 10% and 9%, respectively.

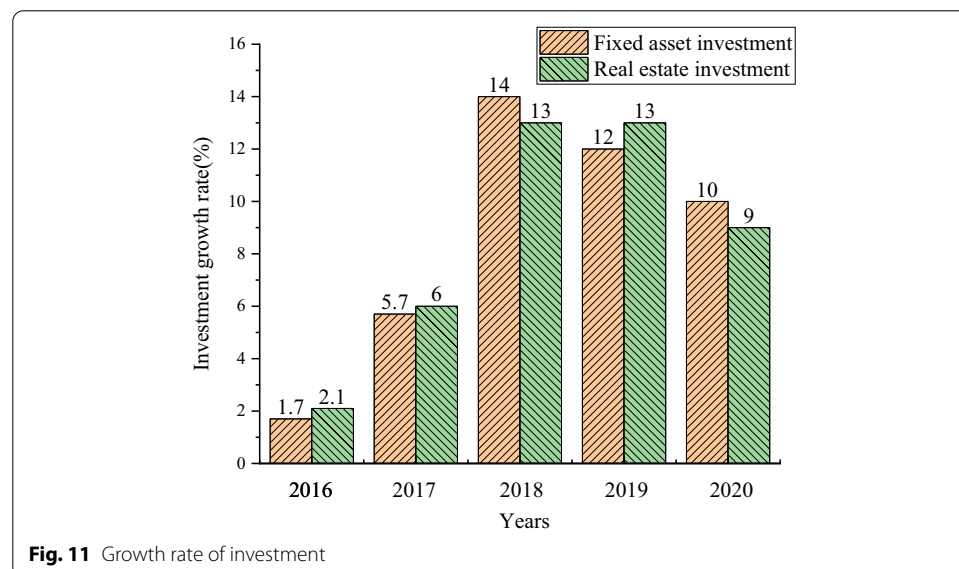
This proves that the investment in fixed assets and real estate increases from 2016 to 2018, but decreases from 2019 and 2020.

Figure 12 shows the proportion of the three major industries in GDP. The details are as follows: in 2016, the primary industry accounts for 3.5%, the secondary industry for 44.5%, and the tertiary industry for 51.8%; in 2017, the primary industry accounts for 3.7%, the secondary industry for 43%, and the tertiary industry for 52.3%; in 2018, the primary industry accounts for 3.2%, the secondary industry for 45.1%, and the tertiary industry for 52.9%; in 2019, the primary industry accounts for 3.4%, the secondary industry for 44.8%, and the tertiary industry for 54%; in 2020, the primary industry accounts for 3.8%, the secondary industry for 43.6%, and the tertiary industry for 54.6%.

This shows that the primary industry develops steadily, the development of the secondary industry is not stable, and the position of the tertiary industry improves gradually.

3.2 Financial risk analysis based on BT and PSO

Figure 13 depicts the factors affecting the regional economy. According to the regional economic development, various indicators are detected, and the results are analyzed. Figure 13a shows the indicators of regional economic development level. In 2018, the growth rate of GDP is 7.6%, the growth rate of fixed asset investment is 13%, the growth rate of real estate investment is -3.2% , and the growth rate of consumer price indicator (CPI) is 2%; in 2019, they are 8.1%, 12.3%, -6.1% , and 2.1%; in 2020, they are 8.3%, 11%, -8.9% , and 1.5%, respectively. Figure 13b shows the regional government regulation indicators. In 2018, the growth rate of fiscal revenue is 8.5%, the growth rate of fiscal expenditure is 9%, and the proportion of fiscal revenue in GDP is 27.2%. In 2019, they



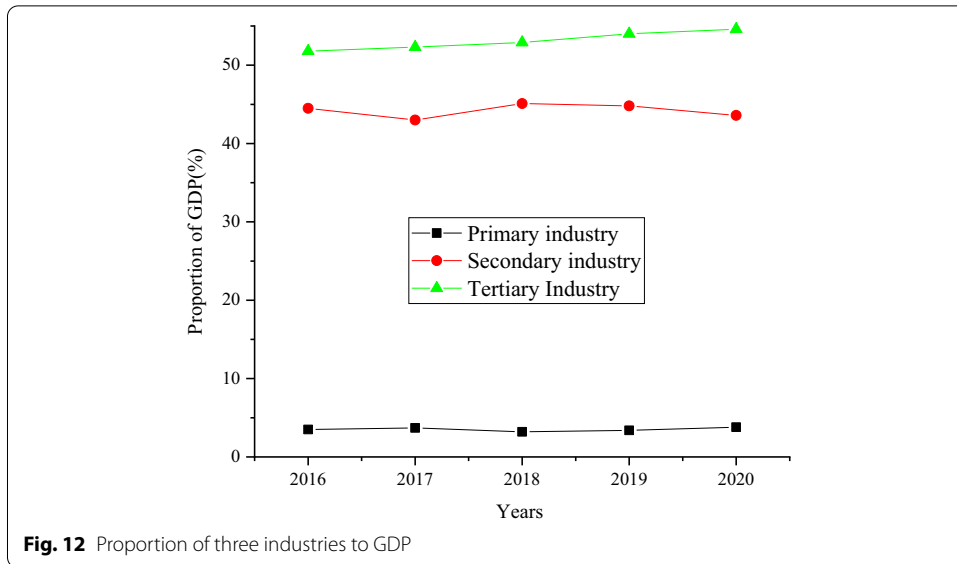


Fig. 12 Proportion of three industries to GDP

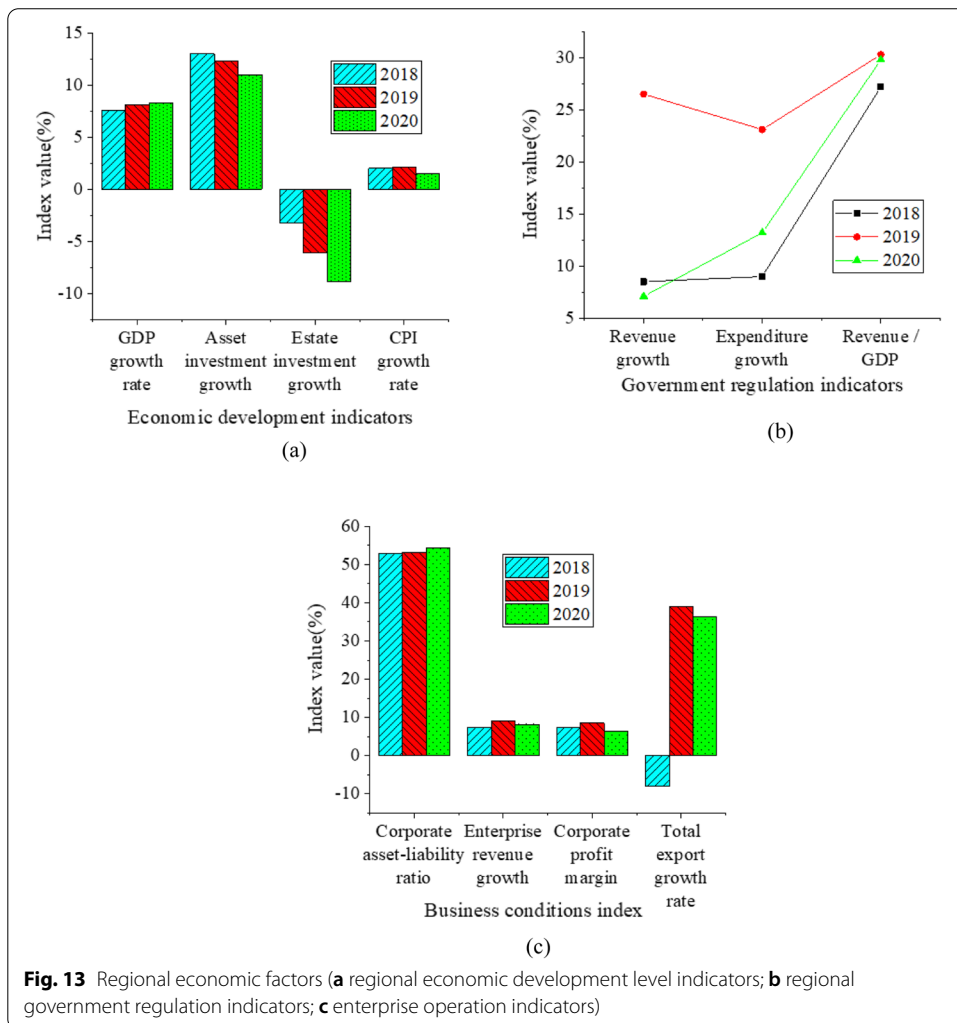
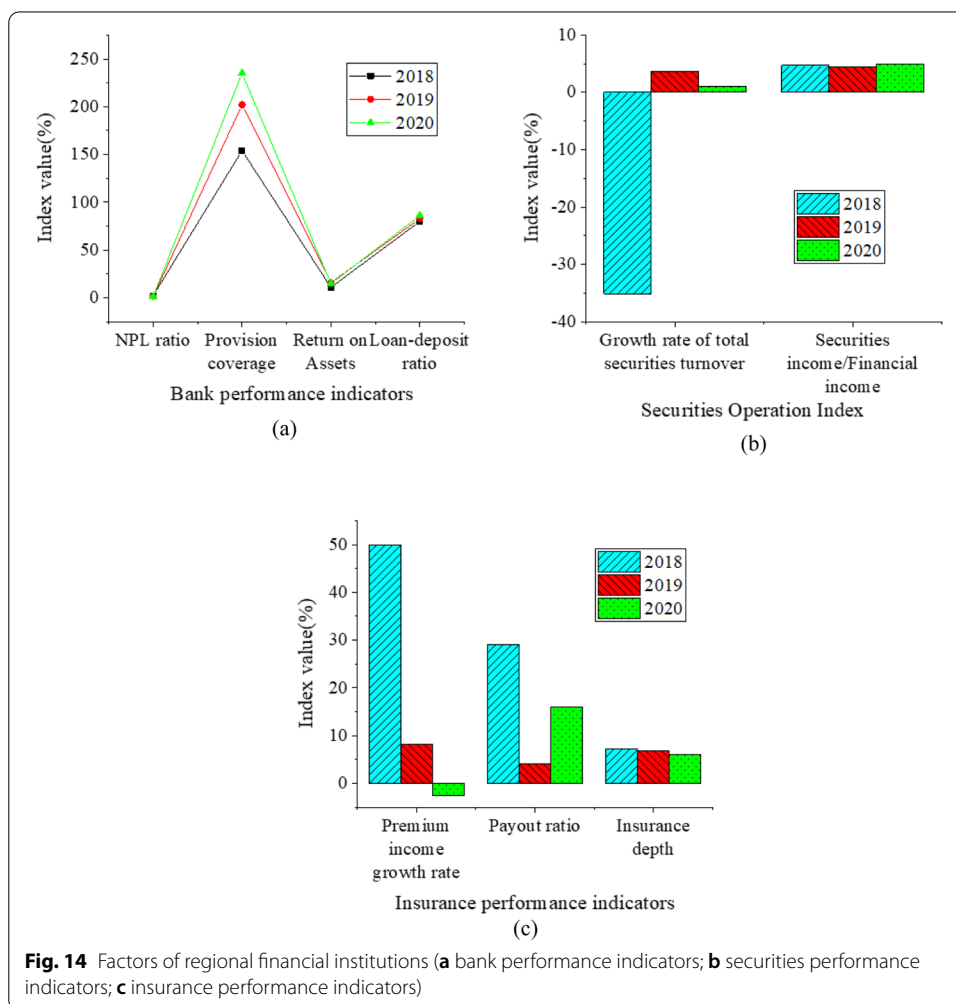


Fig. 13 Regional economic factors (**a** regional economic development level indicators; **b** regional government regulation indicators; **c** enterprise operation indicators)

are 26.5%, 23.1%, and 30.3%. In 2020, they are 7.1%, 13.2%, and 29.8%, respectively. Figure 13c shows the indicators of enterprise operation, in which the asset/liability ratio of enterprises in 2018 is 52.9%, the growth rate of enterprise income is 7.3%, the profit rate of enterprises is 7.45%, and the growth rate of total exports is - 8.1%. In 2019, they are 53.25%, 9%, 8.56%, and 39%. In 2020, they are 54.3%, 8.2%, 6.4%, and 36.2%, respectively.

The above shows that the growth rate of GDP in 2018 and 2020 continues to rise and the proportion of real estate investment gradually increases; the growth rate of fiscal revenue decreases and the proportion of fiscal revenue in GDP gradually increases; the asset/liability ratio gradually increases, while the growth rate of total exports continues to decline.

Figure 14 shows the risk factors of regional financial institutions. Figure 14a shows the performance indicators of the industry. In 2018, the non-performing loan ratio is 2.3%, the provision coverage ratio is 154%, the return on assets is 11%, and the deposit loan ratio is 80%; in 2019, they are 1.7%, 202%, 16%, and 83%; in 2020, they are 1.6%, 235%, 15%, and 86%, respectively. Figure 14b shows the performance indicators of securities. In 2018, the growth rate of total securities transactions is - 35.1%, and the proportion of securities industry revenue in financial industry revenue is 4.7%;



in 2019, they are 3.6% and 4.5%; in 2020, they are 1% and 5%, respectively. Figure 14c shows the insurance performance indicators. In 2018, the growth rate of premium income is 50%, the loss rate is 29.1%, and the insurance coverage rate is 7.3%; in 2019, they are 8.2%, 4.1%, and 6.9%; in 2020, they are – 2.5%, 16%, and 6.1%, respectively.

4 Conclusion

With the rapid development of the world economy, the risk of the financial industry also increases. On this basis, BT is introduced, and it can effectively solve the risk problem in financial markets. It can disperse stored data and complete point-to-point transactions. Therefore, it is combined with DL and AI to prevent and control the risk in the financial market. The cluster analysis algorithm and PSO based on ML and DL are used to analyze and evaluate the financial risk, and the results are summarized as follows: (1) BT has the advantages of decentralization and distrust, which enables the financial industry to carry out new innovation, improve the working efficiency, reduce operating costs, and weaken the financial industry's dependence on media, and it can connect the information between enterprises, promote the cooperation between enterprises, and improve the service quality; (2) the financial risk monitoring system plays a good indicator detection role for China's regional economy, and has obvious advantages in controlling economic situations; (3) BT can provide a good platform for the business expansion, effectively enhances the synergy between companies, and helps companies obtain effective information quickly and conveniently. Based on the above, the study can provide a reference for the future relevant research. However, there are also some deficiencies. For example, the size of the samples is small, which makes the conclusion one-sided. In the future, the size of samples will be expanded, and more attention needs to be paid to interpreting financial risks in subsequent research.

Abbreviations

ML: Machine learning; RF: Random forest; RFML: Random forest machine learning; DT: Decision tree; GDP: Gross domestic product; AI: Artificial intelligence; DL: Deep learning; PSO: Particle swarm optimization; CPI: Consumer price index.

Acknowledgements

The authors acknowledge the help from the university colleagues.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work, and approved it for publication. All authors read and approved the final manuscript.

Funding

This research received no external funding.

Declarations

Ethics approval and consent to participate

This article does not contain any studies with human participants or animals performed by any of the authors. Informed consent was obtained from all individual participants included in the study.

Competing interests

All authors declare that they have no conflict of interest.

Author details

¹School of Business Administration, Faculty of Economics and Management, East China Normal University, Shanghai 200241, China. ²Shanghai Vonechain Information Technology Co., Ltd, Shanghai 200443, China.

Received: 2 December 2021 Accepted: 23 March 2022

Published online: 02 April 2022

References

1. C. Wang, X. Fan, Z. Yin, Financing online retailers: bank vs. electronic business platform, equilibrium, and coordinating strategy. *Eur. J. Oper. Res.* **276**(1), 343–356 (2019)
2. S. Popa, P. Soto-Acosta, D. Perez-Gonzalez, An investigation of the effect of electronic business on financial performance of Spanish manufacturing SMEs. *Technol. Forecast. Soc. Chang.* **136**, 355–362 (2018)
3. K. Täuscher, S.M. Laudien, Understanding platform business models: A mixed methods study of marketplaces. *Eur. Manag. J.* **36**(3), 319–329 (2018)
4. M. Raddant, D.Y. Kenett, Interconnectedness in the global financial market. *J. Int. Money Finance* **110**, 102280 (2021)
5. B.M. Henrique, V.A. Sobreiro, H. Kimura, Literature review: Machine learning techniques applied to financial market prediction. *Expert Syst. Appl.* **124**, 226–251 (2019)
6. T. Fischer, C. Krauss, Deep learning with long short-term memory networks for financial market predictions. *Eur. J. Oper. Res.* **270**(2), 654–669 (2018)
7. M. Azam, S.A. Raza, Financial sector development and income inequality in ASEAN-5 countries: does financial Kuznets curve exists? *Global Bus. Econ. Rev.* **20**(1), 88–114 (2018)
8. C. Alexiou, S. Vogiazas, J.G. Nellis, Reassessing the relationship between the financial sector and economic growth: dynamic panel evidence. *Int. J. Financ. Econ.* **23**(2), 155–173 (2018)
9. H. Pollitt, J.F. Mercure, The role of money and the financial sector in energy-economy models used for assessing climate and energy policy. *Clim. Policy* **18**(2), 184–197 (2018)
10. Y. Chen, S. Hu, H. Mao et al., Application of the best evacuation model of deep learning in the design of public structures. *Image Vis. Comput.* **102**, 103975 (2020)
11. M. Andoni, V. Robu, D. Flynn et al., Blockchain technology in the energy sector: a systematic review of challenges and opportunities. *Renew. Sustain. Energy Rev.* **100**, 143–174 (2019)
12. G. Chen, B. Xu, M. Lu et al., Exploring blockchain technology and its potential applications for education. *Smart Learn. Environ.* **5**(1), 1–10 (2018)
13. Q. Wang, M. Su, Integrating blockchain technology into the energy sector—from theory of blockchain to research and application of energy blockchain. *Comput. Sci. Rev.* **37**, 100275 (2020)
14. A.J. Asaleye, J.I. Adama, J.O. Ogunjobi, Financial sector and manufacturing sector performance: evidence from Nigeria. *Invest. Manag. Financ. Innov.* **15**(3), 35–48 (2018)
15. A.M. Acquah, M. Ibrahim, Foreign direct investment, economic growth and financial sector development in Africa. *J. Sustain. Finance Invest.* **10**(4), 315–334 (2020)
16. S. Prinja, P. Bahuguna, I. Gupta et al., Role of insurance in determining utilization of healthcare and financial risk protection in India. *PLoS One* **14**(2), e0211793 (2019)
17. I. Korol, A. Poltorak, Financial risk management as a strategic direction for improving the level of economic security of the state. *Baltic J. Econ. Stud.* **4**(1), 235–241 (2018)
18. T. Pinjisakikool, The influence of personality traits on households' financial risk tolerance and financial behaviour. *J. Interdiscip. Econ.* **30**(1), 32–54 (2018)
19. J. Zhao, M. Shahbaz, X. Dong et al., How does financial risk affect global CO2 emissions? The role of technological innovation. *Technol. Forecast. Soc. Change* **168**, 120751 (2021)
20. A. Kim, Y. Yang, S. Lessmann et al., Can deep learning predict risky retail investors? A case study in financial risk behavior forecasting. *Eur. J. Oper. Res.* **283**(1), 217–234 (2020)
21. Q. Yang, Y. Wang, Y. Ren, Research on financial risk management model of internet supply chain based on data science. *Cogn. Syst. Res.* **56**, 50–55 (2019)
22. K.T. Chui, D.C.L. Fung, M.D. Lytras et al., Predicting at-risk university students in a virtual learning environment via a machine learning algorithm. *Comput. Hum. Behav.* **107**, 105584 (2020)
23. T.S. Kumar, Data mining based marketing decision support system using hybrid machine learning algorithm. *J. Artif. Intell.* **2**(03), 185–193 (2020)
24. S.M. Othman, F.M. Ba-Alwi, N.T. Alsohybe et al., Intrusion detection model using machine learning algorithm on Big Data environment. *J. Big Data* **5**(1), 1–12 (2018)
25. J.L. Speiser, M.E. Miller, J. Tooze et al., A comparison of random forest variable selection methods for classification prediction modeling. *Expert Syst. Appl.* **134**, 93–101 (2019)
26. T. Hengl, M. Nussbaum, M.N. Wright et al., Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ* **6**, e5518 (2018)
27. P.A.A. Resende, A.C. Drummond, A survey of random forest based methods for intrusion detection systems. *ACM Comput. Surv. (CSUR)* **51**(3), 1–36 (2018)
28. R. Janani, S. Vijayarani, Text document clustering using spectral clustering algorithm with particle swarm optimization. *Expert Syst. Appl.* **134**, 192–200 (2019)
29. T. Lei, X. Jia, Y. Zhang et al., Significantly fast and robust fuzzy c-means clustering algorithm based on morphological reconstruction and membership filtering. *IEEE Trans. Fuzzy Syst.* **26**(5), 3027–3041 (2018)
30. A. Azad, G.A. Pavlopoulos, C.A. Ouzounis et al., HipMCL: a high-performance parallel implementation of the Markov clustering algorithm for large-scale networks. *Nucleic Acids Res.* **46**(6), e33–e33 (2018)
31. G. Li, W. Wang, W. Zhang et al., Grid search based multi-population particle swarm optimization algorithm for multimodal multi-objective optimization. *Swarm Evolut. Comput.* **62**, 100843 (2021)
32. A.A. Nagra, F. Han, Q.H. Ling et al., An improved hybrid method combining gravitational search algorithm with dynamic multi swarm particle swarm optimization. *IEEE Access* **7**, 50388–50399 (2019)

33. Z. Xin-gang, L. Ji, M. Jin et al., An improved quantum particle swarm optimization algorithm for environmental economic dispatch. *Expert Syst. Appl.* **152**, 113370 (2020)
34. X. Xu, P. Lin, Parameter identification of sound absorption model of porous materials based on modified particle swarm optimization algorithm. *PLoS ONE* **16**(5), e0250950 (2021)
35. X. Zhang, R. Zhang, J. Wang et al., An adaptive particle swarm optimization algorithm based on aggregation degree. *Recent Adv. Electr. Electron. Eng.* **11**(4), 443–448 (2018)
36. M. Issa, A.E. Hassanien, D. Oliva et al., ASCA-PSO: Adaptive sine cosine optimization algorithm integrated with particle swarm for pairwise local sequence alignment. *Expert Syst. Appl.* **99**, 56–70 (2018)

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
