

Research Article

Stock Price Prediction Methods based on FCM and DNN Algorithms

Wennan Wang,^{1,2,3} Wenjian Liu,² Linkai Zhu ,^{4,5} Ruijie Luo,⁵ Guang Li,² and Shugeng Dai³

¹Alibaba Cloud Big Data Application College, Zhuhai College of Science and Technology, Zhuhai, China

²Institute of Data Science, City University of Macau, Macau, China

³Department of Finance, School of Economics, Xiamen University, Xiamen, China

⁴Institute of Software Chinese Academy of Sciences, Beijing, China

⁵Faculty of Data Science, City University of Macau, Macau, China

Correspondence should be addressed to Linkai Zhu; linkai@iscas.ac.cn

Received 28 October 2021; Revised 16 November 2021; Accepted 24 November 2021; Published 17 December 2021

Academic Editor: Sang-Bing Tsai

Copyright © 2021 Wennan Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the rapid economic development and the continuous expansion of investment scale, the stock market has produced increasing amounts of transaction data and market public opinion information, making it further difficult for investors to distinguish effective investment information. With the continuous enrichment of artificial intelligence achievements, the status and influence of artificial intelligence researchers in academia and society have been greatly improved. Expert system, as an important part of artificial intelligence, has made breakthrough progress at this stage. Expert system is based on a large amount of professional knowledge and experience for a specific field. Computers of this system can be used to simulate the decision-making process of experts to provide a decision-making basis for solving some complex problems. This research mainly discusses stock price prediction methods on the basis of artificial intelligence (AI) algorithms. Fuzzy clustering is a data mining tool that has been developed in recent years and is widely used. Using this method to process super large-scale databases with various data attributes has the characteristics of high efficiency and small amount of information loss. Theoretically speaking, the use of fuzzy clustering technology and related index method can effectively reduce the massive financial fundamentals of listed companies. By analyzing the influencing factors of stock value investment, we specifically select from the financial statements of listed companies the five aspects that can reflect their profitability, development ability, shareholder profitability, solvency, and operating ability. The full text runs through a variety of AI methods that is the characteristic of the research method used in this article, which pays special attention to verifying the theoretical method model. Doing so ensures its effectiveness in practical applications. In stock value portfolio research, a portfolio optimization model, which integrates the dual objectives of portfolio risk and returns into the risk-adjusted return of capital single objective constraints and solves the portfolio, is established. The accuracy and recall of the FCM model are relatively stable, with accuracies of 0.884 and 0.001, respectively. This research can help improve the number and quality of listed companies.

1. Introduction

Traditional value investment theory is inapplicable to stock market. Securities market has been flooded with all kinds of speculative activities for a long time, which has affected the stability of the market. Therefore, the formation of the value investment concept is conducive to that of a rational investment market atmosphere. Although the trend of value return has gradually appeared in recent years, disputes exist

about the applicability of the value investment concept in state-owned stock markets. Start with the artificial intelligence theme fund, introduce the funds currently on the market that invest in artificial intelligence theme stocks, and then analyze the distribution characteristics of fund returns through descriptive statistics; then carry out statistical tests, establish relevant models, and calculate the VAR value of the fund to be studied; and finally, empirically analyze the risk characteristics of artificial intelligence theme funds.

The shortcoming of the valuation method is the excessive dependence on the model. The complexity of the stock price phenomenon cannot be described by the valuation model alone. The testing method of the model is also a difficult problem. Not many applications of artificial intelligence (AI) methods exist in stock value investment. Nevertheless, the superior performance of this type of method has attracted the attention of domestic academic circles, and the application prospect is optimistic.

The complex task is related to improving investment decisions. Zhang X believed to estimate SOH due to the different discharge/charge characteristics and operating conditions of batteries. His proposed method was validated [1]. Etmianiesfahani A introduced a new meta-heuristic algorithm. He also demonstrated the algorithm performance through two engineering design problems [2]. Shroff N found that managers of firms with financial covenants invested more (and less) when changes in GAAP increased (decreased) contract slack. He also recorded the specific mechanism embodied in the relationship [3]. Gupta R believed that no monthly frequency of economic activity data is observed; thus, he used the NBER recession dummy for his measurements, which, in turn, can easily be included [4]. Das S P predicted a financial derivative [5]. Financial data are different from other types of data, and data about extreme risks in the stock market are more special. Traditional machine learning methods for processing unbalanced data sets cannot guarantee that the generated data can also show the relevant characteristics of financial data. Adjusting the processing method of unbalanced data can be considered to adapt to unbalanced data.

Through an in-depth analysis of the virtual nature of stocks, this research explains that the virtual nature of stocks is the deep theoretical basis of stock value investment theory and the scientific theoretical basis for the success of value investment. It is based on the AI method to focus on the stock selection process during the stock investment process. Two key links of investment portfolio are noted, namely, design specific stock value investment strategy and method process. In this article, principal component analysis (PCA) is conducted on each index selected. The indexes are then divided into two groups with high collinearity according to internal and external factors. Subsequently, SPSS software is used to perform PCA. Using fuzzy clustering technology and related index method can realize the effective reduction of the massive financial fundamental information of listed companies.

2. Research Methods

2.1. Feature Index Extraction. In this study, PCA is conducted to extract feature indicators. The PCA method has the following two advantages. First, multiple indicators are comprehensively reflected through several principal components. Second, multiple indicators are converted into several principal components that are irrelevant and have most of the information of the original data, thus solving the problem of collinearity of the indicators. Dealing with the risk characteristic indicators necessary to improve the convergence and accuracy of the DNN and FCM models is

therefore important. PCA is performed on each index selected. The indexes are then divided into two groups with high collinearity according to internal and external factors. SPSS software is used to conduct PCA.

The best characteristic of the risk-adjusted capital gains of the sample stocks is that the model does not have high requirements on the data structure. It can quickly find the law from the data model itself to complete an improved data fitting. The d -order phase-by-period difference is determined through the nonstationary sequence to gradually stabilize the modeling process. As a result, the newly obtained stationary sequence can use the RAROC model to present the sequence. Then, the original sequence is obtained through inverse transformation [6].

$$\nabla^d y = \beta + \sum_{i=2}^p \kappa y + \mu + \sum_{i=2}^p \kappa t, \quad (1)$$

where $\nabla^d y$ represents a new sequence obtained by d successive difference and β is white noise random sequence μ , which is used to represent the random error at time t .

The RAROC of the sample stocks is as follows [7]:

$$\text{RAROC} = \frac{E_R}{\text{VAR}}. \quad (2)$$

The improved logistic growth model is used, which contains large numbers of assumptions and smoothness adjustment-related parameters, which are used to fit the trend [8].

$$g(t) = \frac{C(t)}{1 + \exp\left(\left(k + \partial(t)^T\right)\beta + \left(k + \partial(t)^T\right)\phi\right)}, \quad (3)$$

where C is a function of carrying capacity with time as the independent variable and $(k + \partial(t)^T)$ is a function of growth rate related to time.

2.2. Fuzzy Clustering Algorithm of Stock Value Feature Selection. Stock value information is mainly reflected in the profitability, development ability, shareholder profitability, solvency, and operating ability of listed companies. The financial indicators comprising each of them are also rich. If qualitative analysis methods are used to directly select indicators from them, then it will undoubtedly lead to a large amount of missing information, which will directly reduce the success rate of investment decision-making. Therefore, adopting a scientific method to process information is necessary. On the one hand, mass information should be reduced to an operable range. On the other hand, the integrity of the information should be ensured as much as possible. Fuzzy clustering is a data mining tool that has been developed in recent years and is widely used. Using this method to process super large-scale databases with various data attributes has the characteristics of high efficiency and small amount of information loss. Theoretically speaking, the use of fuzzy clustering technology and related index method can achieve effective reduction of the massive financial fundamentals of listed companies.

In this study, fuzzy clustering technology is used to cluster these five types of indicators. The correlation index

method is also employed to screen the indicators, compressing the stock value investment and finally selecting the stock value. An investment closely related to indicators constitutes the stock value feature set.

2.2.1. Attribute Reduction. The first step of pattern recognition is to analyze the effectiveness of various features and select the most representative features. Machine language is quite powerful in extracting mathematical features. Attribute reduction has two methods: one is feature selection, and the other is feature extraction.

Artificial intelligence is one of the hottest topics at the moment, and the daily fluctuations and changes of the stock market are also topics of concern to the public. Under the trend of artificial intelligence, listed company stocks related to artificial intelligence topics attract the attention and love of ordinary investors. There are two main ways for investors to invest in artificial intelligence-themed stocks in the stock secondary market: one is to directly buy and sell related stocks; the other is to buy and sell open funds that focus on investing in artificial intelligence-themed stocks. The purpose of feature evaluation is to analyze whether the selected features are reasonable, whether some important information is lost, and whether the distance among different categories can be widened. The purpose of feature extraction is to describe the sample with few features to achieve the purpose of reducing the dimensionality of the feature space. FCM stock value feature selection is shown in Figure 1.

2.2.2. Steps of Fuzzy Clustering and Its Key Algorithms. According to the definition of cluster analysis, one is clustering based on sample similarity, called sample clustering; the other is clustering based on indicator similarity, called index clustering. The cluster analysis method can be subdivided into many methods according to the different mathematical tools used. Among them, the cluster analysis method using fuzzy mathematical tools is called fuzzy clustering. In the pattern recognition system, the initial features of the samples are usually high dimensional, redundant, and noisy. In the initial features of the sample, the functions and meanings of the same type of features for the sample are the same or similar. Multiple correlation directly affects the efficiency and performance of the classifier. Therefore, the fuzzy clustering method can be used to cluster the multiple features of the sample to first realize their classification and then select the most representative feature composition among similar features. Through the selection of similar indicators, the purpose of feature screening can be achieved.

(1) Data standardization: the specific algorithm for data standardization includes two steps [9]:

① Translation and standard deviation transformation [10]:

$$x = \frac{s - x_k}{s}, \quad k = 1, 2, \dots, n. \quad (4)$$

Among them, $x_k = (1/m) \sum_{i=1}^m x$.

It realizes the dimensionlessness of the data but does not guarantee that they are all on interval x_k .
② Translation and range transformation [11]:

$$x'' = \frac{x' - \min\{x_k\}}{\max\{x_k\} - \min\{x_k\}}. \quad (5)$$

Obviously, all x'' are in the interval $[0, 1]$.

(2) Fuzzy similarity matrix establishment: suppose X and Y are two nonempty sets; then, the direct product is as follows [12]:

$$X \times Y = \{(x, y) | x \in X, y \in Y\}. \quad (6)$$

In this study, the correlation coefficient method is used to calculate the fuzzy similarity coefficient [13].

$$R_{ij} = \frac{\sum_{k=1}^m (x - x_i)(x_k - \bar{x})}{\sqrt{\sum_{k=1}^m (x_k - \bar{x})^2} \sqrt{\sum_{k=1}^m (x - x_i)^2}}, \quad (7)$$

where x_i and x_j are any two variables in the variable space.

The fuzzy similarity matrix established according to the correlation coefficient method is as follows [14]:

$$R = \begin{bmatrix} R(x_1, x_1) & R(x_1, x_2) & \cdots & R(x_1, x_n) \\ \cdots & \cdots & \cdots & R(x_2, x_n) \\ \cdots & \cdots & \cdots & \cdots \\ R(x_n, x_1) & R(x_n, x_2) & \cdots & R(x_n, x_n) \end{bmatrix}. \quad (8)$$

Fuzzy cluster analysis on the influencing factors of stock value investment belongs to index clustering. The fuzzy similarity matrix of the profitability index is shown in Table 1.

(3) Clustering: in this article, the transitive closure method is used for clustering. First, the fuzzy equivalence matrix is obtained by seeking the transfer closure using the square method. Second, the fuzzy equivalence matrix of the five aspects of profitability, development capability, shareholder profitability, solvency, and operating capability is calculated.

$$R \longrightarrow R^2 \longrightarrow R^4 \cdots \longrightarrow R^{2^i} \longrightarrow \cdots \quad (9)$$

The self-multiplication operation of the square matrix replaces the product and sum operation in the conventional matrix multiplication with the intersection and union in the fuzzy set operation.

$$A = R \circ R \Leftrightarrow A_{ij} = \bigvee (R_{ik} \wedge R_{ki}). \quad (10)$$

(4) Feature screening: the relevant index method is used to screen three similar indicators in each aspect. For the classification that has been performed, a representative indicator can be selected from each category as a typical indicator. The feature selection here adopts the correlation index method. For the indicators in the same category, selecting the index with

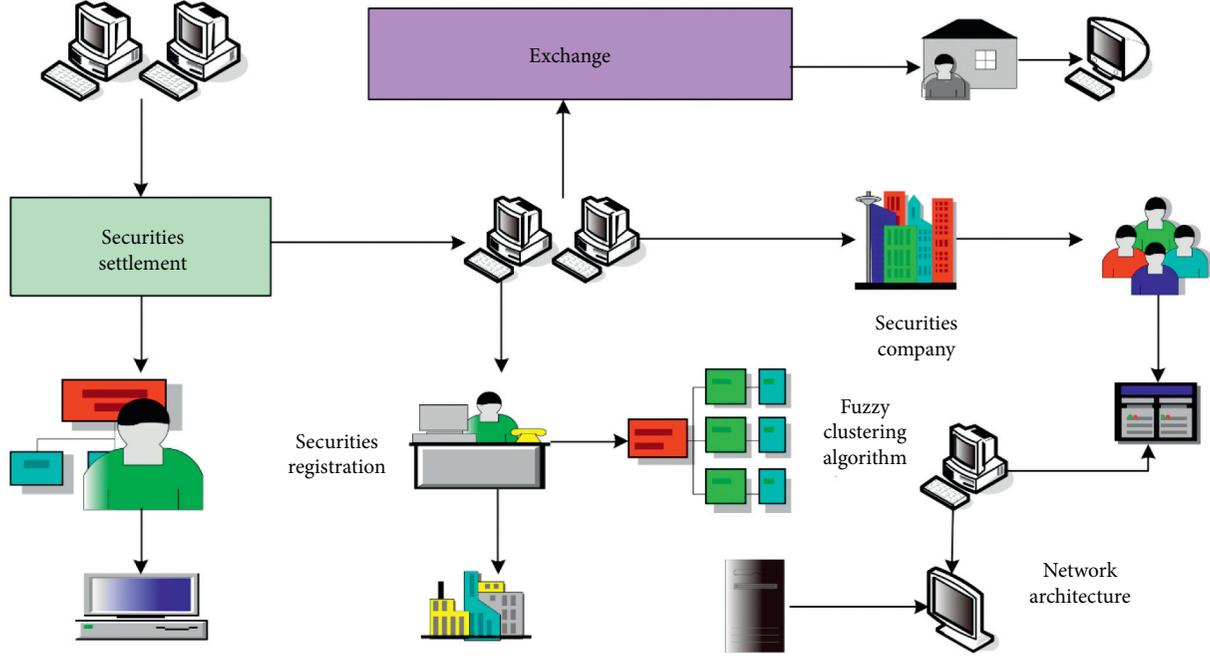


FIGURE 1: FCM stock value characteristic selection.

TABLE 1: Fuzzy similarity matrix of profitability metrics.

| A | A_1 | A_2 | A_3 | A_4 | A_5 | A_6 | A_7 | A_8 |
|-------|--------|--------|--------|--------|--------|--------|--------|--------|
| A_1 | 0.3437 | 1.0000 | 0.5437 | 0.5773 | 0.4307 | 0.3101 | 0.1973 | 0.0139 |
| A_2 | 0.3574 | 0.5437 | 1.0000 | 0.9735 | 0.7303 | 0.3535 | 0.3457 | 0.0457 |
| A_3 | 0.3477 | 0.5773 | 0.9735 | 1.0000 | 0.7355 | 0.3535 | 0.3547 | 0.0374 |
| A_4 | 0.4147 | 0.4307 | 0.7303 | 0.7355 | 1.0000 | 0.3375 | 0.3747 | 0.0091 |
| A_5 | 0.1570 | 0.3101 | 0.3535 | 0.3535 | 0.3375 | 1.0000 | 0.1450 | 0.0509 |
| A_6 | 0.1057 | 0.9903 | 0.5033 | 0.5415 | 0.3711 | 0.1777 | 0.1751 | 0.0107 |
| A_7 | 0.1115 | 0.1973 | 0.3457 | 0.3547 | 0.3747 | 0.1450 | 1.0000 | 0.0070 |
| A_8 | 0.0113 | 0.0139 | 0.0457 | 0.0374 | 0.0091 | 0.0509 | 0.0070 | 1.0000 |

the greatest degree of correlation with the same type of indicators is necessary to ensure that the selected features can cover more comprehensive information. The specific method is as follows: calculate first the correlation coefficient among the indicators in each category, and then calculate the mean value of the square of the correlation coefficient between each indicator and other indicators, that is, the correlation index; subsequently, select the one with the largest correlation index as the typical indicator. If only one indicator exists in the classification, then it can be directly included in the indicator set. If two indicators are found in the classification, then either one of them should be chosen.

Financial risk refers to the uncertainty of the results of a financial behavior or event to the parties; the risk of the stock market refers to buying and selling stocks in the stock market, and investors face the uncertainty of income due to the fluctuation of stock prices. According to the relationship between risk and return, and whether the risk can be diversified, the risk in the stock market can be divided into nonsystematic risk and systemic risk. According to the

process of establishing the stock selection model, the structure of the design stock selection model is shown in Figure 2.

2.3. Stock Price Prediction. Traditional machine learning methods are generally based on the principle of empirical risk minimization. Their learning effect is limited by the number of learning samples. In practical applications, often few known samples exist. The training effect is unsatisfactory. The function set is as follows [15]:

$$M_1 \subset M_2 \subset M_3 \subset \dots \subset M_K \dots \subset M. \quad (11)$$

It is decomposed into a nested sequence of function subsets. Each subset can be arranged according to the size of Φ , namely, VC arrangement:

$$h_1 \leq h_2 \leq \dots \leq h_k \leq \dots. \quad (12)$$

In this way, the confidence range is the same in the same subset. The best evaluation criterion is predicted expected risk R [16].

$$R(\alpha) = \int L(y, L(x, \alpha)) dF(x, y). \quad (13)$$

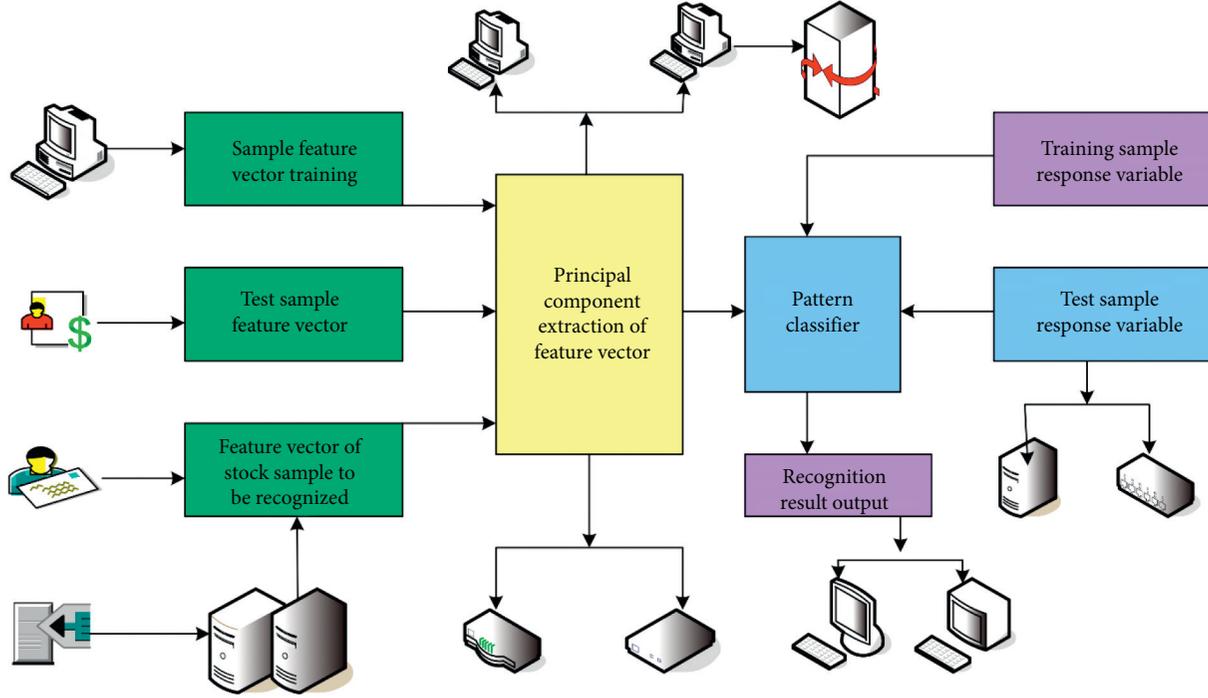


FIGURE 2: Structure of the stock selection model.

Value at Risk (VAR) is expressed in mathematical language as

$$\text{Prob}(\Delta P > \text{VAR}) = 1 - c. \quad (14)$$

VAR is the value at risk under the confidence level c [17]. The expected return rate of the sample stocks is

$$E_Q = \frac{1}{n-1} \sum_{t=1}^{n-1} \left(\frac{p_{t+1} - p_t}{p_t} \right). \quad (15)$$

The financial market presents obvious multifractal characteristics, and the volatility state of the financial market can be characterized by these characteristics.

First, for each trading day, the probability that the price series is located in a certain area is calculated [18]:

$$P(r) = \frac{l_i}{\sum_{i=1}^n l_i}. \quad (16)$$

Probability measure $P(r)$ and r^α have a power law relationship as follows:

$$P(r) \sim r^\alpha. \quad (17)$$

The same length interval with the same mouth is denoted as follows [19]:

$$N(r) \sim r^{-f(\alpha)}. \quad (18)$$

The same $f(\alpha)$ also obeys the power law distribution division function, namely,

$$S(r) = \sum_{i=1}^m P(r) \sim r^q. \quad (19)$$

The value range of q is $[-100, 100]$ and is finally obtained through Legendre transformation [20]:

$$F(\alpha) = \alpha q - \tau(q), \quad (20)$$

where α is the parity index. The exponential weight function is

$$\varphi = \frac{ke^{-k(1-p)}}{1 - e^{-k}}, \quad (21)$$

where $k > 0$ is the Arrow-Pratt absolute risk aversion coefficient.

According to the value range of the coefficient, φ is divided into two sections [21]:

$$\varphi = \begin{cases} \gamma p^{\gamma-1}, & \gamma > 1, \\ \gamma(1-p)^{\gamma-1}, & 0 < \gamma < 1, \end{cases} \quad (22)$$

where γ is the risk aversion coefficient. Expected shortfall (ES) is

$$\text{ES} = \frac{1}{1-\varphi} \int_{\varphi}^1 \text{VAR}(x) dx. \quad (23)$$

For a known training sample, accurate prediction is aimed for by determining the corresponding relationship between the input and the output. The relationship between variables x and y can be used as a joint probability $F(x, y)$; then, n samples are independent of each other and equally distributed. The optimal function solution of the prediction function set $\{f(x, w)\}$ can be obtained, and the expected risk obtained by $F(x, y)$ can be estimated with this function [22].

$$\varphi = \begin{cases} \gamma p^{\gamma-1}, & \gamma > 1, \\ \gamma(1-p)^{\gamma-1}, & 0 < \gamma < 1. \end{cases} \quad (24)$$

$L(\gamma, f(x, w))$ represents the loss incurred during the forecast. Assuming that the expected risk has a uniform probability distribution, the following formula is used as an estimate of the expected risk definition [23]:

$$R(w) = \frac{1}{n} \sum_{i=1}^n L(\gamma, f(x, w)). \quad (25)$$

3. Results

This article will conduct an in-depth exploration of the application of value investment in the stock market from both theory and method. First of all, this article starts from the virtual nature of stocks, draws on the traditional value investment theory, and combines the development stage of the stock market to study the value investment problem of stock market, and then, proposes a new value investment framework to improve the value investment theory from a theoretical perspective. Compared with R_1 and R_2 , the proposed stock selection model has achieved good results in average yield and accumulated net worth. On the one hand, the model defeated the average return of R_1 in seven of the nine test periods and beat the average return of R_2 in eight periods. Except for the third quarter of 2017 and the first quarter of 2018, the model achieved positive returns in all test periods. The failure to achieve positive returns in these two periods may be due to the impact of the stock market crash. These results show that the proposed stock selection model is an effective investment strategy. The results of the average rate of return and cumulative net worth are displayed in Figure 3.

Value investment is a kind of investment theory with rich connotation. Using value investment to guide investment practice requires a comprehensive, objective, and dynamic understanding of value investment theory. It is meaningless to study value investment theory out of the development stage of the stock market. In the face of a new round of investment boom, it is necessary for us to carry out in-depth research and reinterpretation of value investment theory based on the phased characteristics of the stock market development. On the one hand, this is the enrichment and development of value investment theory, and on the other hand, this enables securities investors to re-examine the theory of value investment based on national conditions and form a comprehensive and objective understanding of it. To discuss the importance of predictive indicators, the weights of predictive and financial factors are described, as presented in Figure 4. The average predictive index (PR) weight is 0.515. Among all the indicators, the predictive index weight ranks second, second only to the net income year-on-year growth rate. Compared with other indicators, the predictive factor has made a huge contribution to the model.

As shown in Module A in Table 2, the stock selection model using FCM performs more prominently in AR,

SharpeRatio, Prob.(R_1), Prob.(R_2), and HitRatio. At the same time, the stock selection model of SVR Max and Min have achieved a few advantages, whereas the stock selection model of back propagation neural network (BPNN) is at a disadvantage in any evaluation index. Therefore, the predictive factors constructed by FCM can well assist stock selection decisions. To statistically prove that the stock selection model using FCM is significantly better than other benchmark models, t -tests are performed on each benchmark model. The process is as follows: Normality test, which is the prerequisite for t -test, is first performed on each model. Then, a null hypothesis is constructed, H_0 : the AR of the stock selection model using FCM is significantly lower than that of other benchmark models. According to Module B, all models have passed the normality test, and their p -values are less than 5%. At the 95% confidence level, the FCM-based stock selection model and the predictive factors constructed by FCM can well assist stock selection decisions. The decision-making comparison in Module A is shown in Table 2.

Deepening the research on the concept of value investment is conducive to promoting the maturity and perfection of value investment theory, cultivating value investment concepts suitable for China's stock market, and objectively helping to curb excessive speculation, making value investment deeply rooted in the hearts of the people, and indirectly regulating stock market order to promote the healthy development of the stock market and give full play to the function of the stock market. In the securities market, where risks and opportunities coexist, we must understand the nature of the market and adapt to the changes in the market environment to better grasp the fleeting market opportunities. To statistically prove that the stock selection model using FCM is significantly better than other benchmark models, t -tests are performed on each benchmark model. The process is as follows: Normality test, which is the prerequisite for the t -test, is performed on each model. Then, a null hypothesis is constructed, H_0 : the AR of the stock selection model using FCM is significantly lower than that of other benchmark models. According to Module B, all models have passed the normality test, and their p -values are less than 5%. At the 95% confidence level, the predictive factors constructed by FCM can well assist stock selection decision-making. The decision-making comparison in Module B is presented in Table 3.

The evaluation indicators of all prediction methods in Module C are presented in Table 4. Obviously, the method of using FCM to predict stock prices crushes all benchmark models in computing time, directional accuracy D-STAT, prediction accuracy MAPE, and RMSE. In this study, FCM has a relatively prominent predictive ability. To statistically prove that the prediction result based on the extreme learning machine FCM is significantly better than those based on other benchmark models; the Diebold–Mariano test (DM test) is performed on each benchmark model. The process is as follows: construct a null hypothesis, H_0 : the prediction result based on FCM is significantly lower than that based on the benchmark model.

The DM test results are shown in Module D in Table 5. All p -values are less than 5%, indicating that the FCM-based

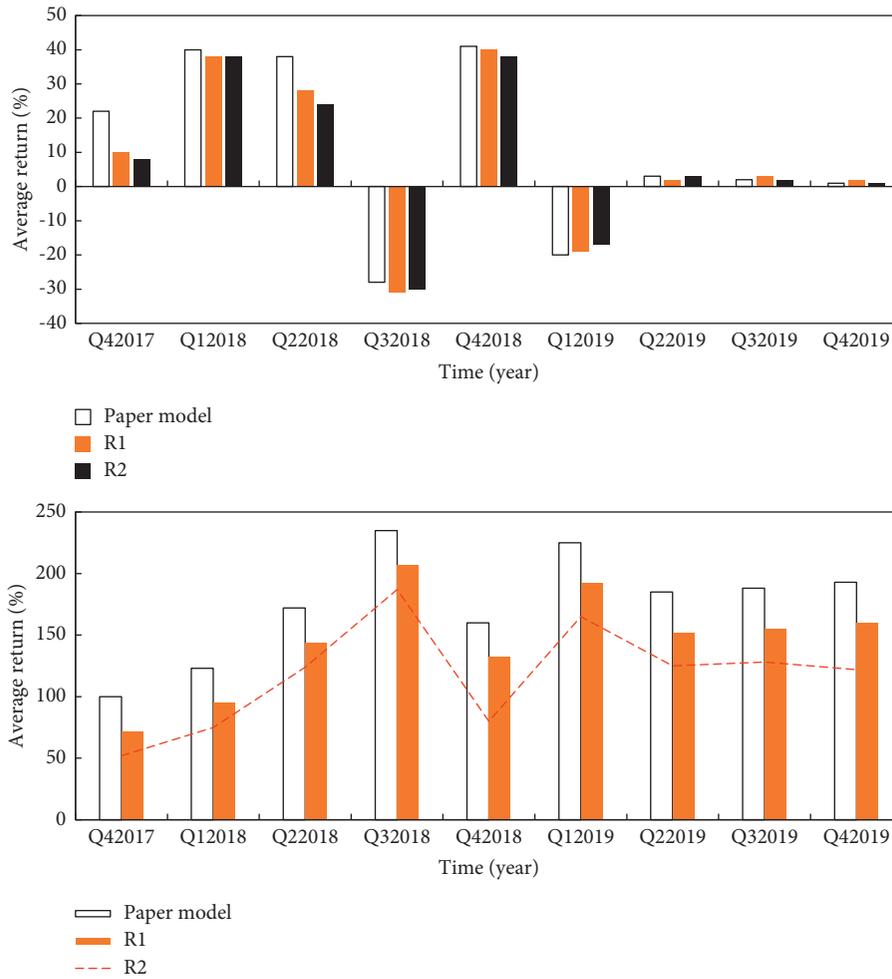


FIGURE 3: Results of the average yield and cumulative net value.

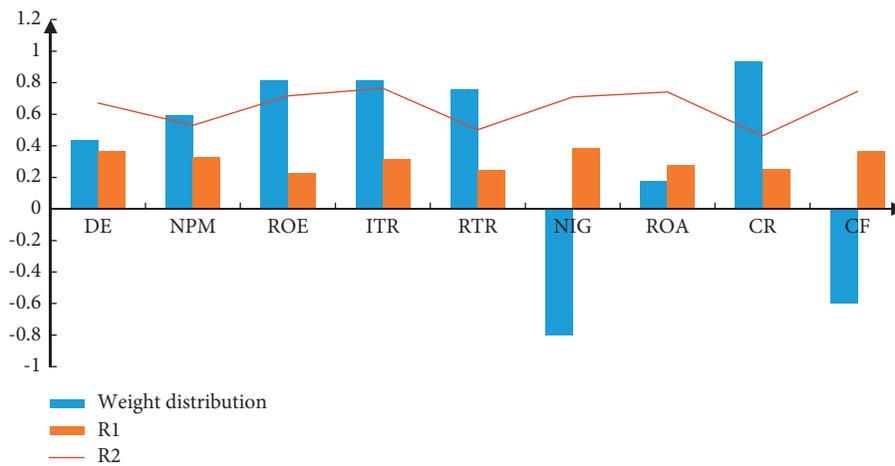


FIGURE 4: Weights of the predictors and financial factors.

prediction results are better than the benchmark model at the 95% confidence level. It proves that FCM has stronger predictive ability. The predictor based on FCM structure can better assist the stock selection decision-making.

From an algorithm perspective, the pure genetic algorithm (GA) converges quickly in the initial stage but is extremely unstable; meanwhile, the fuzzy clustering algorithm achieves the convergence of the optimal solution

TABLE 2: Decision contrast in Module A.

| Module A: inspection | | | |
|------------------------|---------|---------|---------|
| Model | FCM | BPNN | SVR |
| AR | 0.2070 | 0.0855 | 0.0872 |
| SharpeRatio | 0.4874 | 0.4085 | 0.4227 |
| Max | 0.4277 | 0.4040 | 0.4424 |
| Min | -0.4068 | -0.4272 | -0.2727 |
| Prob.(R ₁) | 0.6444 | 0.4047 | 0.5285 |
| Prob.(R ₂) | 0.7742 | 0.7444 | 0.7527 |
| HitRatio | 0.7574 | 0.7556 | 0.5742 |

TABLE 3: Decision contrast in Module B.

| Module B: inspection (H ₀ : AR _{FCM} ≤ SAR _{Benchmark}) | | | |
|---|--------|---------|---------|
| Model | FCM | BPNN | SVR |
| Positive testability (<i>p</i> value) | 0.3066 | 0.1886 | 0.4361 |
| <i>t</i> -stat | N.A. | 16.6068 | 13.3034 |
| <i>p</i> value | N.A. | <0.0001 | <0.0001 |

TABLE 4: Evaluation metrics for all prediction methods in Module C.

| Model | FCM | BPNN | SVR |
|----------------|---------|------------|------------|
| Computing time | 78.7280 | 17088.2270 | 22217.8770 |
| D-stat | 0.7272 | 0.7070 | 0.7027 |
| MAPE | 0.2281 | 0.2280 | 0.2777 |
| RMSE | 18.2818 | 22.2171 | 27.8878 |

TABLE 5: DM test results.

| Module D: model checking (H ₀ : the prediction result based on the benchmark model is better than that based on ELM) | | | |
|---|------|---------|---------|
| Model | ELM | BPNN | SVR |
| DM-stat | N.A. | -17.15 | -3.57 |
| <i>p</i> value | N.A. | <0.0001 | <0.0001 |

robustly. The performance comparison of the algorithms is shown in Figure 5.

Batch_size is the size of each training and gradient update block. With the increase of the epoch value, the overall accuracy of the model showed a trend that first decreased significantly and then gradually increased. The volatility was not obvious, and the value of the recall rate first increased significantly and then gradually stabilized. Comprehensively comparing the above results, under the condition of comprehensively weighing the model accuracy and the sensitivity to crisis samples, this article takes the epoch value as the optimal parameter. At this time, the accuracy rate is 0.702, and the recall rate is 0.731. It performs well and can effectively predict the crisis state under the condition of ensuring an accurate prediction of the state of the stock market. In summary, this article has completed the construction of the FCM extreme risk early warning model. The evaluation result of the overall iteration number is illustrated in Figure 6.

The five-fold cross-validation of the DNN model is generally stable. The average accuracy rate is 0.723, the

standard deviation is 0.010, the average recall rate is 0.705, and the standard deviation is 0.014. The fluctuations in the values of accuracy and recall are very small, and the model is excellent. The stability and the fitting effect are good. From the results, the accuracy and recall of the FCM model are relatively stable. The average accuracy is 0.884, and the standard deviation is 0.0015. Therefore, the FCM model also has excellent stability. The overall inspection and comparison of the DNN model is shown in Figure 7.

Figure 8 shows the data randomly selected within 135 days. The daily return rate sequence of stock indexes fluctuated sharply in 2016, and the extreme risks of upward and downward were obvious. After the introduction of the position-limiting policy, the volatility of the sequence has eased, but the downward risk still appears to be an obvious volatility agglomeration.

This article now builds the DNN and FCM models for the early warning of extreme stock market risks. Both models are applied to the real Shanghai and Shenzhen 300 markets. The goal is to provide an early warning of extreme stock risks. The market state and whether extreme risks will occur are predicted.

During the period from May 2017 to October 2017, extreme risks appeared in the market. The crisis warnings of FCM and DNN covered this period intensively. A few days before the occurrence of the “stock market crash,” the model also carried out extreme risk warnings. FCM and DNN are still sensitive to extreme risks and can play certain roles in early warning. For the crisis state, the accuracy rate of DNN is 0.862, and that of FCM model is 0.872. On the whole, the accuracy rate of DNN is 0.682, and that of FCM is 0.891. The result is the same as the conclusion obtained from theoretical analysis. The DNN model has achieved a good predictive effect on the crisis state samples, but it has produced more “false alarms.” That is, the normal state is predicted as a crisis state, and the FCM model performs well as a whole, but not sensitive enough to the crisis state. Comprehensive comparison, DNN, and FCM extreme risk early warning models have achieved good prediction results. Each model has advantages and disadvantages and to a certain extent has complementary effects. However, the FCM extreme risk early warning model performs better than the DNN model. The performance comparison of extreme risk early warning models is shown in Figure 9.

The upper-tail risk measurement values output by the three stock prediction models at the 95% confidence level are illustrated in Figure 10, which are VAR, ES, and TDRM (tail distortion when the risk aversion coefficient $\gamma = 0.5$ Risk). From Figure 10, TDRM > ES > VAR shows that when the tail risk is given a high weight, the degree of risk aversion is enhanced, and the risk management or investment portfolio strategy based on the TDRM measurement value is cautious and conservative.

4. Discussion

After a comprehensive review of the quantitative stock selection model, this article proposes a new quantitative stock selection model, which optimizes the predictive factors and the weights of financial factors in the previous step through

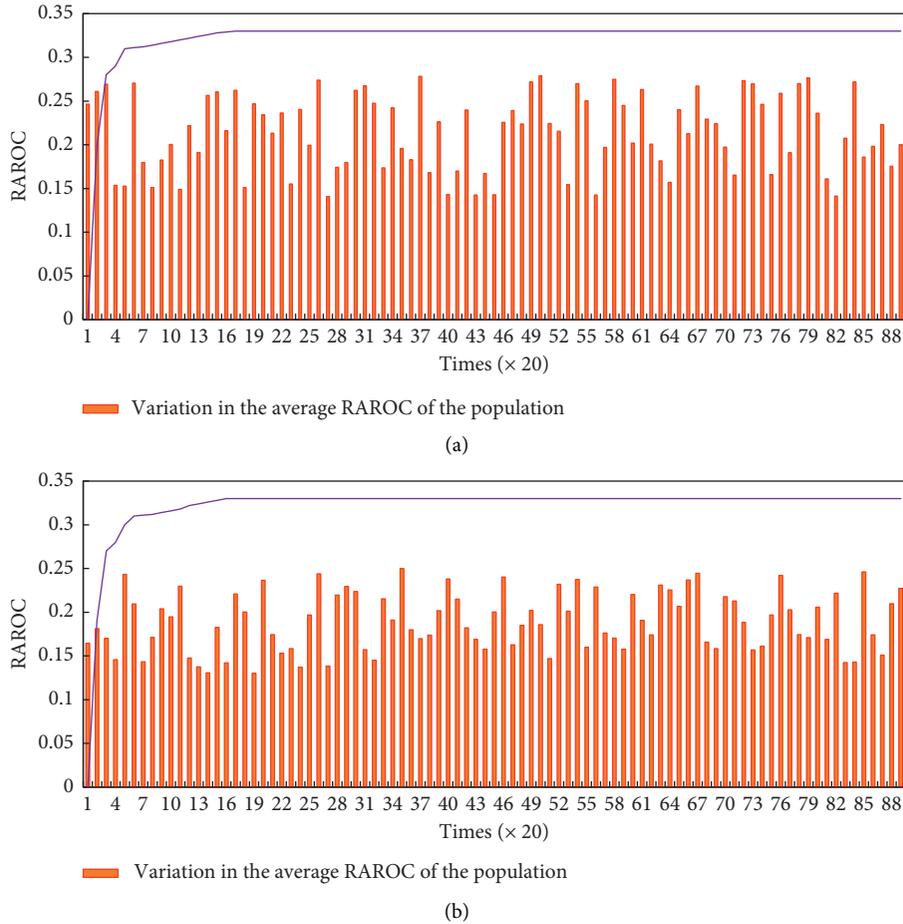


FIGURE 5: Performance comparison of the algorithm. (a) Traditional genetic algorithm. (b) Fuzzy clustering algorithm.

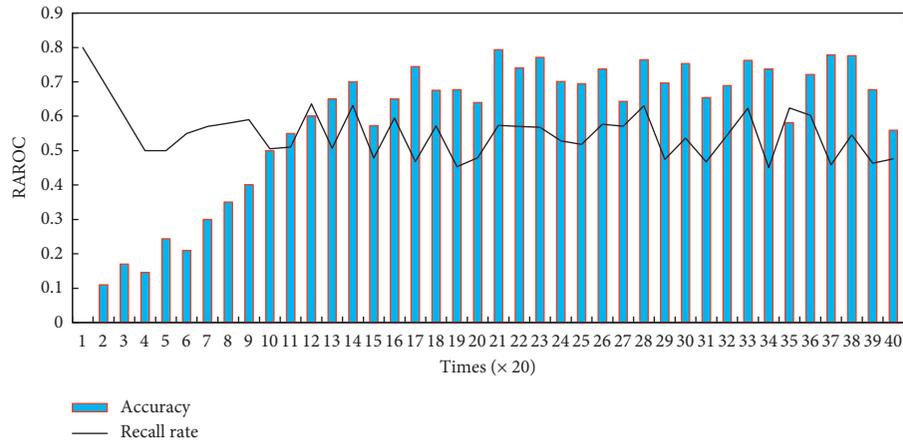


FIGURE 6: Evaluation results of the overall number of iterations.

the FCM. Doing so allows the optimization of weights and various factors accordingly. The linear combination of values evaluates each stock. High-scoring stocks are selected, and equal-weight investments are made. Our analysis reveals that the proposed model can be used as an effective tool to invest in stocks and obtain substantial market returns. Its income is much higher than the average market performance (average

return of all stocks) and the A-share index. At the same time, the weight coefficient of the constructed predictor is higher than most financial factors. It shows that the predictive factors have a great influence on the multifactor stock selection model and play a positive role [24].

Through the research of AI methods applied in stock investment, we find that AI methods that have emerged in

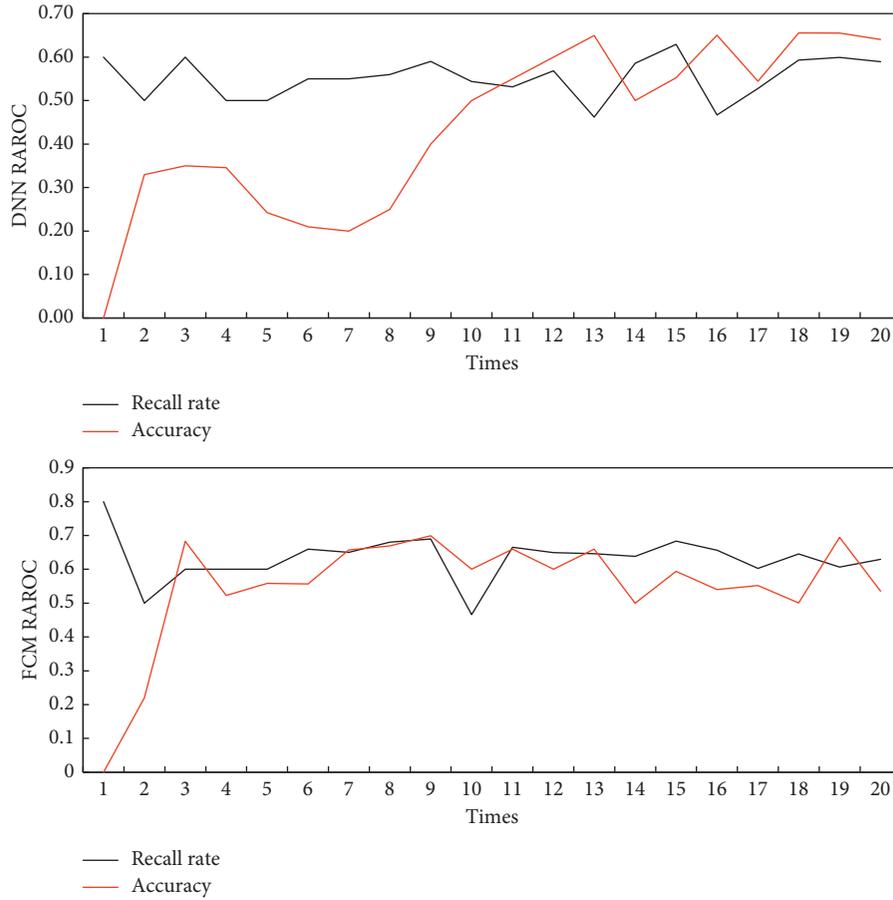


FIGURE 7: DNN model overall test comparison.

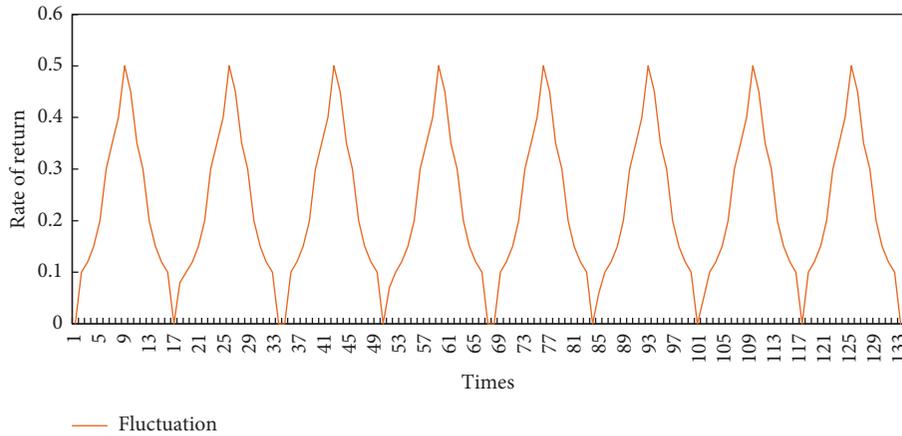


FIGURE 8: Randomly selected data within 135 days.

recent years have incomparable advantages over traditional stock investment methods. The prospects for application in the field of stock investment are also optimistic. Securities models form the basis of modern securities investment theory research, but the three models imply efficient market assumptions, which cause their application value in inefficient markets to be greatly reduced. Nevertheless, their theoretical methods and portfolio investment strategies in analyzing the factors affecting stock returns provide a solid

foundation for stock value investment. The stock transaction price information identifies stocks that may have rates of return that exceed the market rate and uses GA to select stocks to construct a portfolio; the stock price information also evaluates the actual effect of the algorithm in selecting stocks through the evaluation of the combination [25].

As far as the financial fundamental information of listed companies is concerned, the amount of information contained in is large, which are mostly contained relationships

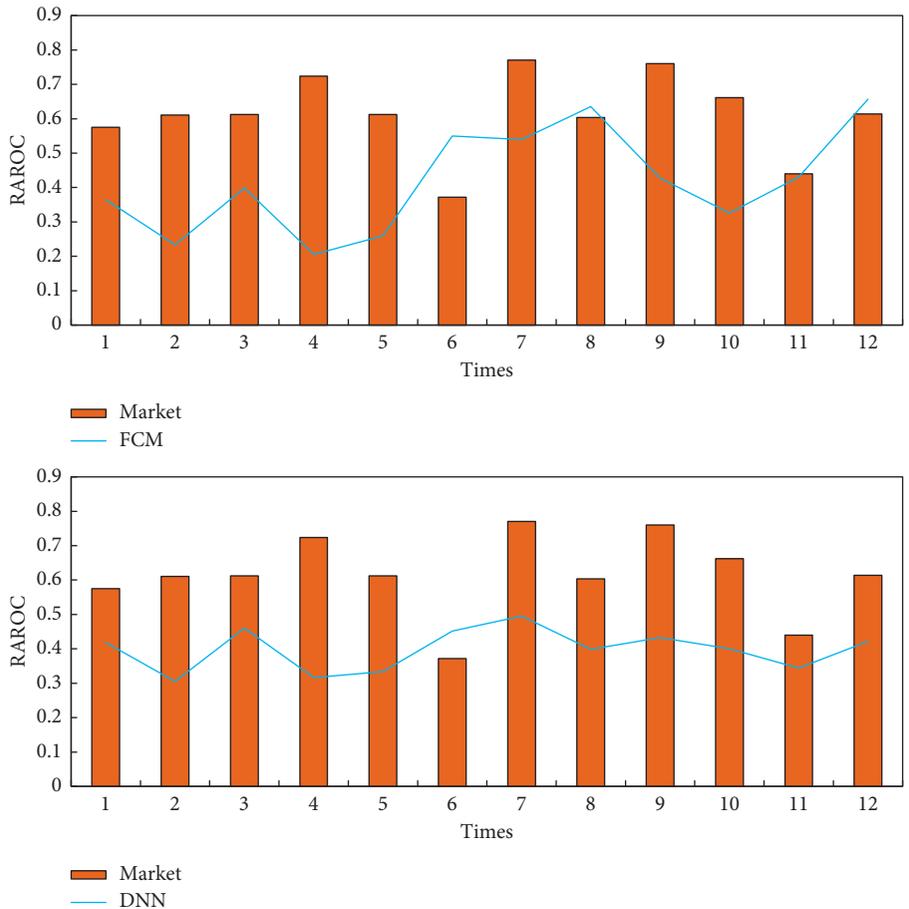


FIGURE 9: Performance comparison of the extreme risk early warning model.

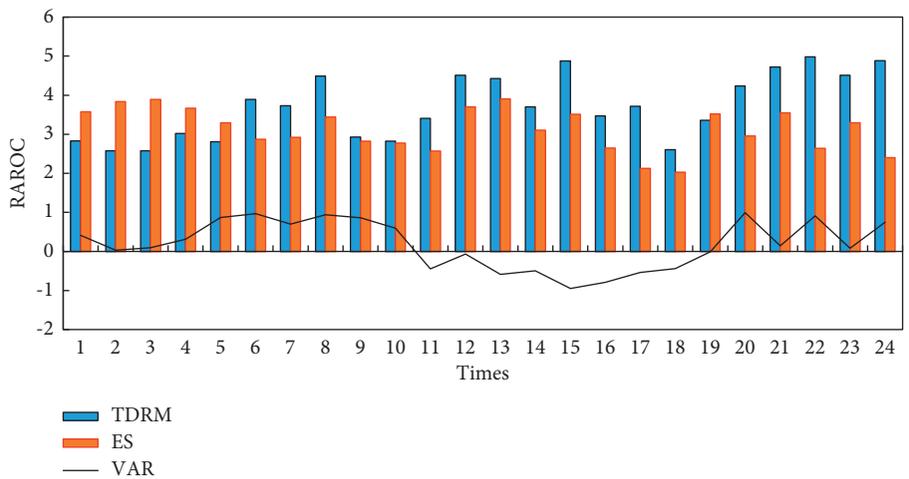


FIGURE 10: Using upper-tail risk measures with outputs from the three stock prediction models at the 95% confidence level.

or strong correlations between the information. To effectively use this information in the practice of value investment, using scientific methods to process massive amounts of information and then digging out the key information needed for stock value investment decisions are necessary [26].

The application of value investment theory in the stock market is not a question of absolute can or absolutely not, but a question of the degree of applicability, that is, the level of investment performance and how to use it. Therefore, whether traditional value investment theory can be enriched and developed in combination with the stage characteristics

of the stock market to solve the applicability of value investment to the stock market is the key to the success or failure of value investment theory in the stock market.

Value investment is a broad concept, but the stock investment guided by the value investment concept is specific. It needs to dig deep into value information, that is, financial information. The intrinsic value of an enterprise should be the present value of all dividends that investors can obtain. Fuzzy model methods are mainly based on expert experience or statistical methods to establish fuzzy models for prediction. Many companies have established an evaluation index system for the investment value of listed companies, and the evaluation angles are relatively consistent. The appraisal result test or its credibility is low. Therefore, the applicability of this method in the stock investment value appraisal is still debatable. Stock investment usually includes two contents: stock selection and investment portfolio construction. After the selection of stocks with investment value through scientific methods, portfolio investment is used to diversify investment risks and achieve relatively high investments at low-risk levels. Portfolio investment is also an important research content of stock investment.

We suggest using AI methods to design stock value investment methods suitable for stock market. The value investment method of stocks is to select and extract the value characteristics of the stock selection model according to the stock investment process, construct the stock selection model, optimize the stock investment portfolio, and develop portfolio solution methods and other aspects. The profitability, development ability, shareholder profitability, solvency, and operating ability of listed companies are an organic whole that is independent and complementary. Stock investors are most directly concerned about the profitability, development ability, and shareholder profitability of listed companies. The solvency and operational capabilities of listed companies are the basis for guaranteeing company profits, shareholder profits, and investment safety. Therefore, a comprehensive grasp of these five aspects is the prerequisite and guarantee for the profitability of stock value investors [27].

5. Conclusion

This research analyzes predictive factors and finds that when FCM is used to predict stock returns, the prediction results have advantages in directional accuracy, prediction accuracy, and running time. Therefore, when using the predictive factors obtained by FCM to select stocks, an improved return on investment can be obtained. The return of stock selection models, which only consider financial factors, is much smaller than that of stock selection models that consider financial and predictive factors, suggesting the importance of predictive factors in multifactor stock selection models. FCM plays a more prominent role in the optimization of the multifactor stock selection model than GA DNN. When the optimization goal is set to the long-short portfolio income, the stock selection model can achieve improved results during the test period. Finally, machine learning is used for time series forecasting, so that all indicators are predictive.

Intelligent optimization calculations are also used for stock selection research. This study will investigate these interesting questions in the future.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This study was supported by Macau Foundation (No. MF2012) and the National Key Research and Development Program of China (No. 2020YFB806504).

References

- [1] X. Zhang, Y. Wang, C. Liu, and Z. Chen, "A novel approach of battery pack state of health estimation using artificial intelligence optimization algorithm," *Journal of Power Sources*, vol. 376, pp. 191–199, 2018.
- [2] A. Etminaniesfahani, A. Ghanbarzadeh, and Z. Marashi, "Fibonacci indicator algorithm: a novel tool for complex optimization problems," *Engineering Applications of Artificial Intelligence*, vol. 74, pp. 1–9, 2018.
- [3] N. Shroff, "Corporate investment and changes in GAAP," *Review of Accounting Studies*, vol. 22, no. 1, pp. 1–63, 2017.
- [4] R. Gupta and M. Wohar, "Forecasting oil and stock returns with a Qual VAR using over 150 years off data," *Energy Economics*, vol. 62, pp. 181–186, 2017.
- [5] S. P. Das and S. Padhy, "A novel hybrid model using teaching-learning-based optimization and a support vector machine for commodity futures index forecasting," *International Journal of Machine Learning and Cybernetics*, vol. 9, no. 1, pp. 97–111, 2018.
- [6] A. Plaat, J. Schaeffer, W. Pijls, and A. Bruin, "A minimax algorithm better than alpha-beta? no and yes," *Journal of Telemedicine and Telecare*, vol. 6, no. 13, pp. 47–49, 2017.
- [7] B. K. Patle, D. R. Parhi, A. Jagadeesh, and S. K. Kashyap, "On firefly algorithm: optimization and application in mobile robot navigation," *World Journal of Engineering*, vol. 14, no. 1, pp. 65–76, 2017.
- [8] L. Kothhoff, B. Hurley, and B. O'Sullivan, "The ICON challenge on algorithm selection," *AI Magazine*, vol. 38, no. 2, pp. 91–93, 2017.
- [9] S. Yao, F. Xu, P. Zhao, J. Wang, and J. Chen, "Feature selection algorithm based on neighborhood valued tolerance relation rough set model," *Moshi Shibie yu Rengong Zhineng/Pattern Recognition and Artificial Intelligence*, vol. 30, no. 5, pp. 416–428, 2017.
- [10] W. A. Hussein, S. Sahran, and S. N. H. Sheikh Abdullah, "The variants of the Bees Algorithm (BA): a survey," *Artificial Intelligence Review*, vol. 47, no. 1, pp. 67–121, 2017.
- [11] F. Zhong, H. Li, and S. Zhong, "An improved artificial bee colony algorithm with modified-neighborhood-based update operator and independent-inheriting-search strategy for global optimization," *Engineering Applications of Artificial Intelligence*, vol. 58, pp. 134–156, 2017.

- [12] M. A. Remli, S. Deris, M. S. Mohamad, S. Omatu, and J. M. Corchado, "An enhanced scatter search with combined opposition-based learning for parameter estimation in large-scale kinetic models of biochemical systems," *Engineering Applications of Artificial Intelligence*, vol. 62, pp. 164–180, 2017.
- [13] H. S. Le and T. M. Tuan, "Dental segmentation from X-ray images using semi-supervised fuzzy clustering with spatial constraints," *Engineering Applications of Artificial Intelligence*, vol. 59, pp. 186–195, 2017.
- [14] M. Ehteram, M. F. Allawi, H. Karami et al., "Optimization of chain-reservoirs' operation with a new approach in artificial intelligence," *Water Resources Management*, vol. 31, no. 7, pp. 2085–2104, 2017.
- [15] Z. Zhang, X. Wang, and J. Lu, "Multi-objective immune genetic algorithm solving nonlinear interval-valued programming," *Engineering Applications of Artificial Intelligence*, vol. 67, pp. 235–245, 2018.
- [16] N. R. Sabar, L. M. Kieu, E. Chung, T. Tsubota, and P. E. Maciel de Almeida, "A memetic algorithm for real world multi-intersection traffic signal optimisation problems," *Engineering Applications of Artificial Intelligence*, vol. 63, pp. 45–53, 2017.
- [17] J. Renić, "Feature selection methods," *Springer International*, vol. 6, no. 4, pp. 1–4, 2017.
- [18] H. Y. Kim and C. H. Won, "Forecasting the volatility of stock price index: a hybrid model integrating LSTM with multiple GARCH-type models," *Expert Systems with Applications*, vol. 103, pp. 25–37, 2018.
- [19] J. Sen and T. D. Chaudhuri, "A predictive analysis of the Indian FMCG sector using time series decomposition-based approach," *SSRN Electronic Journal*, vol. 4, no. 2, pp. 206–226, 2017.
- [20] B. Roni, G. Abbas, and S. Wang, "Return and volatility spillovers effects: study of asian emerging stock markets," *Journal of Systems Science & Information*, vol. 6, no. 2, pp. 97–119, 2018.
- [21] A. A. Salisu and K. O. Isah, "Revisiting the oil price and stock market nexus: a nonlinear Panel ARDL approach," *Economic Modelling*, vol. 66, pp. 258–271, 2017.
- [22] P. Maio and D. Philip, "Economic activity and momentum profits: further evidence," *Journal of Banking & Finance*, vol. 88, pp. 466–482, 2018.
- [23] F. Wang, Y. Zhang, Q. Rao, K. Li, and H. Zhang, "Exploring mutual information-based sentimental analysis with kernel-based extreme learning machine for stock prediction," *Soft Computing*, vol. 21, no. 12, pp. 3193–3205, 2017.
- [24] H. Herwartz, "Stock return prediction under GARCH-an empirical assessment," *International Journal of Forecasting*, vol. 33, no. 3, pp. 569–580, 2017.
- [25] C. Chiyachantana, P. K. Jain, C. Jiang, and V. Sharma, "Permanent price impact asymmetry of trades with institutional constraints," *Journal of Financial Markets*, vol. 36, pp. 1–16, 2017.
- [26] G. Zhang, L. Xu, and Y. Xue, "Model and forecast stock market behavior integrating investor sentiment analysis and transaction data," *Cluster Computing*, vol. 20, no. 1, pp. 1–15, 2017.
- [27] M. Gerges, H. Eng, H. Chhina, and A. Cooper, "Modernization of bone age assessment: comparing the accuracy and reliability of an artificial intelligence algorithm and shorthand bone age to Greulich and Pyle," *Skeletal Radiology*, vol. 49, no. 9, pp. 1449–1457, 2020.