

Retraction

Retracted: Predicting the Link between Stock Prices and Indices with Machine Learning in R Programming Language

Journal of Mathematics

Received 19 December 2023; Accepted 19 December 2023; Published 20 December 2023

Copyright © 2023 Journal of Mathematics. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1].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

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

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

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

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

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

References

- [1] M. Cao, "Predicting the Link between Stock Prices and Indices with Machine Learning in R Programming Language," *Journal of Mathematics*, vol. 2021, Article ID 1275637, 10 pages, 2021.

Research Article

Predicting the Link between Stock Prices and Indices with Machine Learning in R Programming Language

Mengya Cao 

Olin Business School, Washington University in St. Louis, St. Louis 63130, MO, USA

Correspondence should be addressed to Mengya Cao; c.mengya@wustl.edu

Received 9 November 2021; Revised 20 November 2021; Accepted 26 November 2021; Published 10 December 2021

Academic Editor: Naeem Jan

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

This paper provides an in-depth analysis machine study of the relationship between stock prices and indices through machine learning algorithms. Stock prices are difficult to predict by a single financial formula because there are too many factors that can affect stock prices. With the development of computer science, the author now uses many computer science techniques to make more accurate predictions of stock prices. In this project, the author uses machine learning in R Studio to predict the prices of 35 stocks traded on the New York Stock Exchange and to study the interaction between the prices of four indices in different countries. Further, it is proposed to find the link between stocks and indices in different countries and then use the predictions to optimize the portfolio of these stocks. To complete this project, the author used Linear Regression, LASSO, Regression Trees, Bagging, Random Forest, and Boosted Trees to perform the analysis. The experimental results show that the MRDL deep multiple regression model proposed in this paper predicts the closing price trend of stocks with a mean square error interval [0.0043, 0.0821]. Additionally, 80% of the proposed DMISV, KDJSV, MACDV, and DKB stock buying and selling strategies have a return greater than 10%. The experimental results validate the effectiveness of the proposed buying and selling strategies and stock price trend prediction methods in this paper. Compared with other algorithms, the accuracy of the algorithm in this study is increased by 15%, and the efficiency of prediction is increased by 25%.

1. Introduction

A stock is a certificate of ownership issued by a joint-stock company to raise funds, which allows the shareholder to receive dividends and bonuses. With the development of China's economy, the stock market has become a bigger and bigger part of our economic market, even becoming a "barometer" of our economic development [1]. The stock market is a very complex and sizeable financial system, so various economic and political factors affect the changes in the stock market at every moment. Changes in stock price trends are of the utmost concern to stockholders in the stock market. In our stock market, stock prices are influenced by numerous factors, such as policy adjustments, economic environment, and international situation. Therefore, making reasonable forecasts of stock price trends has been a critical difficulty for economists to study [2]. Suppose one can make a good prediction of the stock

price trend. In that case, he or she can reduce the investment risk and combine the predicted stock price trend with the stock buying and selling strategy to help investors make a reasonable adjustment to their investment structure and maximize the return.

At present, there are numerous indicators for judging stock quotes in the stock market, such as MACD (Exponential Smoothed Moving Average), KDJ (Stochastic Indicator), and DMI (Movement Indicator or Tendency Indicator). It is impossible to take all of them into account when making a judgment on the stock market. The selection of one or more of these indicators as a reference, combined with the market environment to judge the stock market, is what we call a buying and selling strategy. Choosing the right buying and selling strategy can help us choose the most desirable stocks from among many and determine when to buy and when to sell the stock so that the shareholder's investment risk is reduced [3].

Machine learning techniques include many different methods to do analysis. A deeper understanding of how the different methods work can help make predictions more robust and more accurate. Therefore, machine learning is an essential tool that can be used in the financial world. In this project, the author used Linear Regression, LASSO, Regression Trees, Bagging, Random Forest, and Boosted Trees to do the analysis. The author will give a brief description of each method in the following section.

The linear regression model is a linear approach to do the analysis, and the model can be written specifically as

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon_1. \quad (1)$$

As it can be seen from the model formulation, the linear regression model is easy to interpret. β_0 is the intercept. β_2 is the slope of the variable x_2 . The linear regression model uses the least-squares method to estimate the parameters [4].

The LASSO model is a more modern alternative analysis. Traditionally, models like linear regression models and ridge regression models would include all variables in the results. However, the LASSO model can force some coefficients to be zero, which makes it easier to interpret. LASSO estimation $\hat{\beta}_j$ is done by minimizing the following equation:

$$\sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2. \quad (2)$$

The regression tree model has many advantages over the first two approaches. First, regression tree models are easy to use, and the resulting rules are easy to interpret and implement. Second, the selection and reduction of variables in regression tree models are automatic and do not require statistical model assumptions. Finally, regression tree models do not require a large amount of task delivery data to be used. However, regression tree models can have high variance.

The Bagging (Bootstrap Aggregating) model is a solution to the problem of high variance generated by regression tree models. The Bagging model is also a simple analysis method, but with powerful ideas. It uses averaging to reduce variance and bootstrap to ensure a large training data set [5]. However, since the Bagging process involves the random selection of a subset of observations, interpreting the results can be difficult. This problem can be solved by using relative impact plots. This paper is mainly focusing on the U.S. stock market. The follow-up research will consider more local stock markets.

The random forest model and the boosted tree model are only models that build on the bag method. Partial dependency graphs and relative importance graphs are important ways to interpret these models.

With these machine learning methods, the author can make better forecasts of stock prices, and one can have a deep understanding of the connections between index prices in different countries. After the prediction, the author will use the naive heuristic to do portfolio optimization. How to obtain valuable information from the massive stock history data? The author researched stocks' buying and selling

points, combined deep learning methods to predict stock price trends, dug out stocks with their investment value, and assisted stock investors in making investment decisions. These researches are of theoretical and practical significance.

The rest of the paper will proceed as follows. The author reviewed historical literature in Section 2. In Section 3, the author introduced the methodology and analysis used in the paper. The author discussed the influence of the forecast and the result of portfolio optimization in Section 4. In Section 5, the author concluded the paper.

2. Literature Review

2.1. Machine Learning and Optimization. Machine learning techniques and optimization mathematics are interactive. Machine learning techniques are a solid foundation for optimization—however, machine learning and optimization focus on different areas of development. Machine learning focuses on more straightforward mathematics and generates robust general optimization codes [6]. But optimization focuses more on accuracy, speed, and robustness.

LASSO is a handy model selection tool for large-scale predictors. Traditional methods such as OLS regression and stepwise regression are subject to random errors. Moreover, when real-world datasets are analyzed in R , the results show that LASSO performs better and more accurately than other traditional methods. On the other hand, based on the forecast curve fit and the mean squared error of the forecast results, the highest accuracy is obtained using the MRDL_4 model to forecast the 30-day trend of stocks. Next, the author compared the MRDL_4 model with the traditional multiple regression model (MLRM). The experiments show that the MRDL_4 model fits the prediction curve better than the multiple regression model, which verifies the effectiveness of the proposed method. However, the different parameter settings in the MRDL model have an impact on the prediction results. The next step will be to adjust the model's parameters and try to train the MRDL model using different optimization functions to improve the prediction accuracy [7].

2.2. Determinants of Stock Price Movements. Jothimani used regression methods to analyze the SSE Composite Index and predict stock prices [8]. Asghar used partial least squares to make a simple prediction of stock prices [9]. Cao et al. used the least-squares trained regression model to select the price of the gold spot as an influencing factor to predict the trend of gold stocks [10]. Livieris used the one-dimensional linear regression method combined with least-squares training regression coefficients to analyze and predict the movements of per capita GDP and per capita consumption in 31 provinces [11]. Atkins et al. used the least-squares training regression model to analyze the trend of rebar prices [12].

Predicting stock price movements is a central and challenging issue in the financial world. There are thousands of factors that can affect the direction of stock prices. The cash flow of a company is an essential factor in predicting stock price movements [13]. The second significant predictor

is diversification. Qing Jiang et al. proposed using an extended short-term memory network (LSTM) to predict stock prices [14]. Gao et al. predicted future market prices based on a deep belief network consisting of multiple layers of random hidden variables [15]. Huck et al. predicted stock prices based on the structure of the restricted Boltzmann machine algorithm in deep belief networks [16]. Ghazanfar et al. predicted stock prices based on recurrent neural networks with a multifactor training model [17]. Parikh et al. expect stock prices based on multiplex recurrent neural networks combined with extracted stock-related news text features [18]. Weihua Chen et al. combined deep learning methods with stock forum data to study stock market volatility accuracy prediction [19].

2.3. Predicting Stock Prices by Machine Learning. A basic approach is to focus on the patterns generated in the stock market and extract knowledge from these patterns to predict the future behavior of the stock market. A necessary process is to make the data easily classifiable. All methods in machine learning can be used to predict the stock market, and most of them are adequate and easy to do analysis.

Due to the numerous factors affecting stock price fluctuations and the high complexity of financial markets, most scholars in past academic studies of stock markets have chosen to use complex techniques or methods to predict stock price movements to judge stock buying and selling points [20]. These theoretically involved stock buying and selling trading models have primarily enriched the theoretical knowledge in the field of finance, but as the complexity of the models increases, the time consumed by their training models increases. If investors are not familiar with stock movements, they can easily suffer losses from the stock market. The use of machine learning for stock price prediction is becoming more critical as increased stocks are entering the market. Artificial neural networks can perform better in predicting stock prices. And decision trees can provide some rules to describe the prediction. Combining these two methods can give us a comprehensive knowledge of stock price prediction.

3. Methodology and Analysis

3.1. Collecting and Processing Data in Excel. All prices for these thirty-five stocks and four smaller companies have been collected from Yahoo Finance. The timeline is from March 7, 2018, to March 5, 2021. When the prices were downloaded from Yahoo Finance, there were some stocks and small companies with null values, and these null values were replaced with the average of the respective stock prices. In addition, in this project, the author has used the percentage change as a predictor, that is, the following formula:

$$y = h(x_i^2), \quad i = m + 1, \dots, M. \quad (3)$$

In addition, the author created a lagged variable of percentage change to eliminate the effect of time differences between countries. Each layer of the DenseNet network is

connected to any other layer in the form of feedforward, and the input of any layer is the output of all the previous layers, and the output of the layer itself is the input of all the subsequent layers so that each layer is connected to the input data, which reduces the error of input information transfer between multiple layers and optimizes the gradient and information transfer [21]. This optimizes the gradient and the flow of information and enhances the transmission of data features. More importantly, the DenseNet network has the effect of regularization, which alleviates the problem of overfitting the data set to a certain extent and utilizes the data features more effectively, as shown in Figure 1.

In addition, the DenseNet network is different from the residual network (ResNet) where each layer has its weight and the number of parameters is huge. In addition, the DenseNet network does not acquire new network architecture by deepening the number of network layers but improves the utilization of parameters by reusing features, so it requires fewer parameters and is easier to train the network. In both forward and backward propagation neural networks, the ReLU function has only a linear relationship, so it takes less time to train the model. On the other hand, the ReLU function does not produce gradient saturation if the input z is a real number greater than 0. Therefore, the ReLU function is chosen as the activation function of the neural network in this paper.

The depth regression model MRDL objective function is to minimize the mean squared error between the predicted and true values of the closing price and is calculated as follows:

$$\text{loss} = \frac{1}{k} \sum_i^{i+k} \frac{1}{2} (y_i + h(x_i^2)). \quad (4)$$

Figure 2 shows the time required to train the MRDL_4 model with the different number of neurons in the hidden layer and its training loss. Therefore, in this section, the MRDL_4 model uses 64 neurons in the implicit layer, and the output of 64 neurons in MRDL_4 implicit layer H1 is used as the input of implicit layer H2, and the implicit layer H2 is adjusted by the small-batch gradient descent method. The output of 64 neurons of implicit layer H2 is used as the input of the output layer, and the weight of each input data is also determined using the small-batch gradient descent method.

DMI is a medium to long-term indicator used to analyze the trend of stock prices. Most of the existing stock analysis indicators are calculated by using the closing price of each day of the stock to calculate each index, ignoring the real difficulties of the stock on that day. For example, when a stock's opening and closing prices for the day are the same as the previous day's opening and closing prices, but the highest (lowest) price for the day is different, then this stock's up and down quotes for the two days are not the same. In most other indicators, this is very difficult to reflect. The DMI indicator is composed of two sets of four parameters: a long/short indicator that includes upward movement +DI and

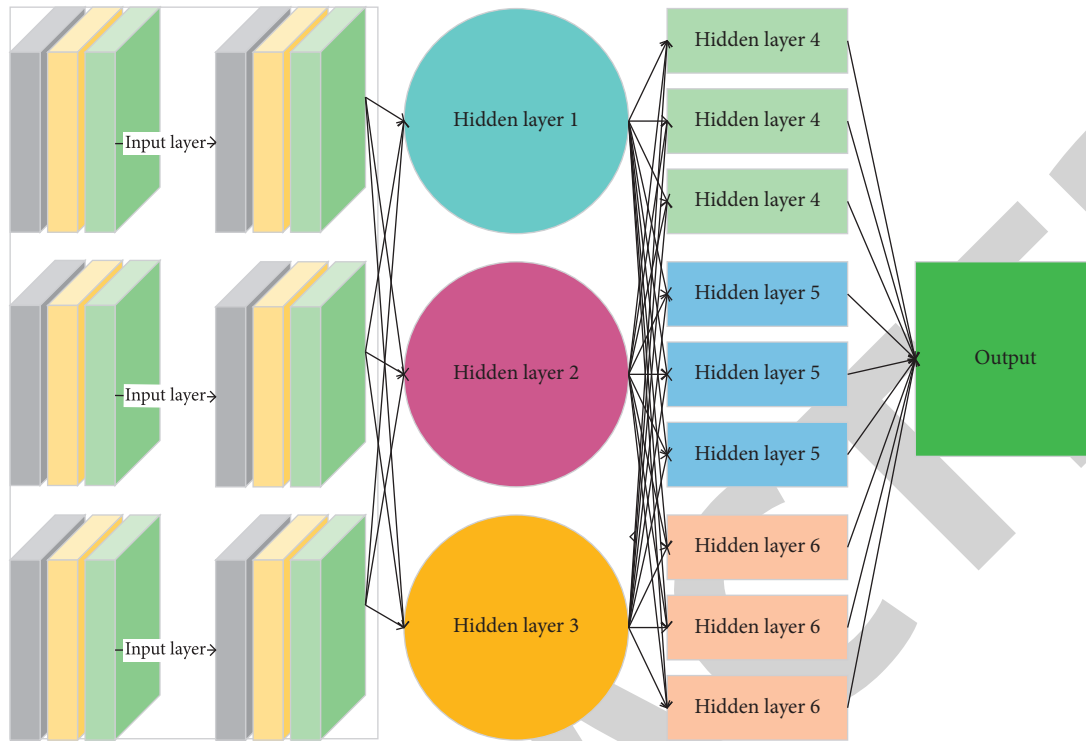


FIGURE 1: DenseNet network structure.

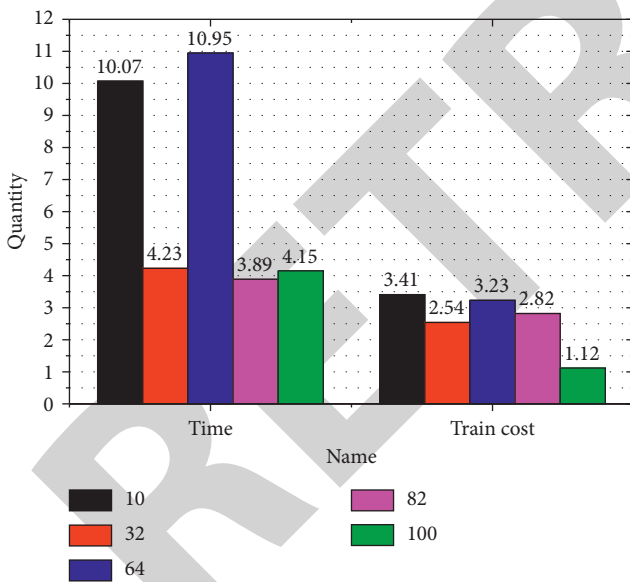


FIGURE 2: Training loss and time required for MRDL_4 model with different number of neurons in the hidden layer.

downward movement $-DI$ and a trend indicator that includes ADX and $ADXR$. Given a stock X , the parameters of the DMI indicator are defined and calculated as follows. This method is a compressed estimation. It obtains a more refined model by constructing a penalty function, which makes it compress some regression coefficients, that is, force the sum of absolute values of coefficients to be less than a certain fixed value; at the same time, set some regression coefficients to zero. Therefore, the advantage of

subset shrinkage is retained, and it is a biased estimate for processing data with multicollinearity.

$$\begin{aligned} +DM_i &= H_i + H_{i-1}, \\ -DM_i &= L_i - L_{i-1}. \end{aligned} \tag{5}$$

The long/short indicators $+DI$ and $-DI$ represent the strength of the upward and downward trend of the stock price. A larger $+DI$ means a stronger uptrend, while a larger $-DI$, on the contrary, means a stronger downtrend. If $+DI$ rises and $-DI$ falls, and if $+DI$ crosses $-DI$, then the stock price will have an upward wave and the buyer's power will be increasing; on the contrary, if $+DI$ falls and $-DI$ rises, and $+DI$ crosses $-DI$, then there will be a downward wave and the seller's power will be increasing, representing a partial fall in the stock price.

In general, the movement indicators $+DI$ and $-DI$ are most accurate in predicting short-term stock buying and selling operations, and when the stock is in an oscillating uptrend, because in an oscillating downtrend, the rally up is short, and the movement indicators $+DI$ and $-DI$ take longer to respond, so it is impossible to accurately predict whether the uptrend can continue at this time, and the same problem exists in consolidation trends. In addition, it should be noted that when the rising indicator $+DI$ rises from 20 or below to above 50, the stock is likely to have an intermediate upward wave, and similarly, when the falling indicator $-DI$ rises from 20 or below to above 50, the stock is likely to have an intermediate downward wave. If both $+DI$ and $-DI$ fluctuate above and below the benchmark line of 20, then the stock is mostly in a box, and the stock market is balanced between long and short forces.

3.2. Descriptive Analysis. To better understand the data set, the author has divided the thirty-five stocks into four categories. Airline-related stocks: DAL, UAL, ALK, SKYW, ALGT, SAVE, CPA, JBLU; transportation and coordination: UPS, FDX, NM, EGLE, TK, EXPD, HUBG, DSX; transportation stocks: KEX, CHRW, ODFL, KNX, ASTG, ASIA, KSU, LSTR, R, OSG, JBHT, STNG, RAIL, NSC, CSX, UNP, MRTN, and others. After the classification step, the distribution, correlations, linear relationships, and seasonal effects are analyzed.

First, the distribution of the percentage changes in stock prices and Indigo prices was seen. For some stocks and small-cap stocks, the percentage changes varied greatly. For example, stocks related to aviation changed more than other stocks. It can be seen in Figure 3 that we chose four stocks related to aviation and plotted box plots to see the distribution. The percentage change ranges from -40% to 40%.

If the ADX crosses the ADXR, the cross formed at this point is called a golden cross, as marked by the solid line in Figure 3. This indicates a period of upward movement for the stock. If the ADX and ADXR move up to above 50 at the same time, the stock market will have an intermediate or higher upward movement, and if it moves up to above 80, the stock market will have a more than doubling of the market. If the ADX and ADXR move down to about 20, the stock market is in a consolidation phase and there is no market. When the distance between ADX, ADXR, +DI, and -DI lines is shortened, the stock market is also in consolidation, but the difference is that the method of using the DMI index to determine the stock market is distorted.

Furthermore, it was found that the percentage change in the price of indices is smaller than the percentage change in the price of stocks. Most of the percentage changes in indices prices are between -10% and 20%.

For the daily data of a given stock, the DMISV buy-sell strategy algorithm is used to calculate the output buy-sell points, and the buy-sell operation is performed based on the buy-sell points, and the stock return is calculated by equation (6), thus verifying the effectiveness of the DMISV buy-sell strategy.

$$D_b = \{DM_i, DM_{i+1}, DM_{i+2}, \dots, DM_{i+k}\}. \quad (6)$$

Second, the correlation and linear relationship between the percentage changes in stocks were checked. The result is that there is a positive relationship between stocks in the same classification. However, this relationship may not be linear. Let me use the shipping and coordination category as an example. Figure 4 shows a scatter plot of these stocks; we can see that there are some positive linear relationships, such as UPS and FDX.

In addition, the relationship between indices from different countries is not clear through the scatter plot. It can be seen from Figure 5 that only SPX and NDX have a strong positive relationship because they are both indices from the United States.

The scope of stock market linkage can be the tendency to have common movements between stock indices of various

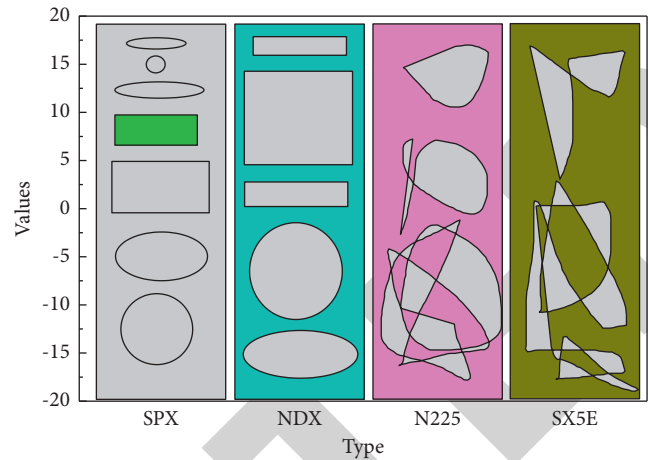


FIGURE 3: Box plot of percentage change in airline-related stocks.

countries or between various segments of a country's stock market and between the prices of various stock assets within each segment.

Based on the principle of the regional scope and market scope of stock indices and stocks from large to small, the main four aspects of stock market linkage are described in terms of linkage among stock indices in the world, linkage among different stock indices between the same countries, linkage among various sections of the same stock market and different industries, and linkage among individual stocks in the same section [22]. Thirdly, the seasonality effect was analyzed. The result is that only some stocks related to airlines have a strong seasonality effect. We can see in Figure 6 that the stock price of Delta Airlines increased from October to December 2018. The seasonal effect may be caused by the holiday season. The economic fundamentals hypothesis is based on the efficient market hypothesis, which asserts that the linkage between stock markets comes from the linkage of fundamental factors between economies. Economic fundamentals as intrinsic causes drive the transmission of shocks between different markets mainly include the microstructure of the market, economic system, industrial structure, macroeconomic policies, and cultural background.

In addition, it is found that airline-related stocks are most affected by COVID-19. As can be seen in Figure 7, the price of Delta Air Lines is relatively stable until COVID-19, and there is significant volatility from COVID-19 onward. When one country's stock market is hit by a capital shock, investors in another country's stock market cannot accurately determine whether the capital shock is the result of an economic risk outbreak or a systemic error based on available information. In addition, coupled with geographical differences, policy differences, and institutional differences, this leads to irrational decisions at the economic level, that is, portfolio adjustments, resulting in stock market volatility in the invested stock market and so contagion to the next stock market. This is reflected in the fact that the opening prices of the stock markets with different opening times have an impact on the opening prices of the stock markets that opened earlier; moreover, the closing prices of

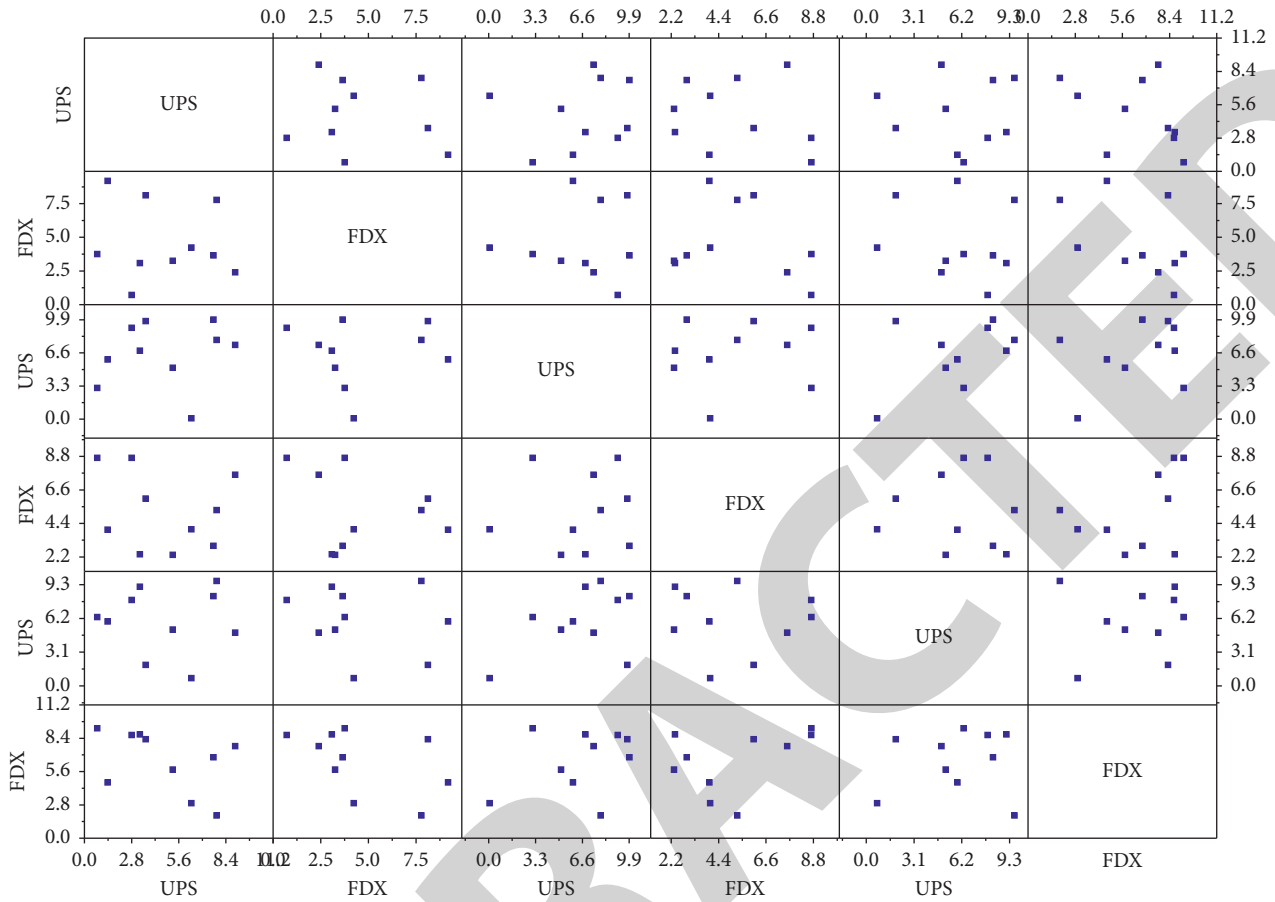


FIGURE 4: Scatter plot of shipping and coordination categories.

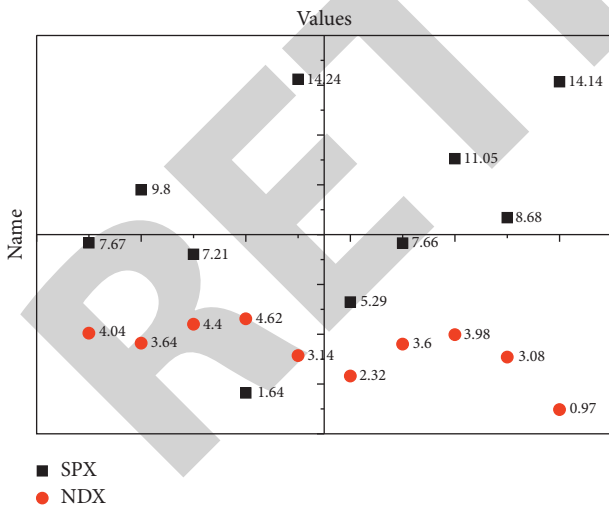


FIGURE 5: Scatter plot of indices.

the stock markets on the same day have an impact on the opening prices of the stock markets on the following day. The linkage effect of stock market contagion is more pronounced during the financial crisis.

Financial liberalization is becoming increasingly widely accepted in the economics world, and at the same time, a series of deregulation measures are gradually being

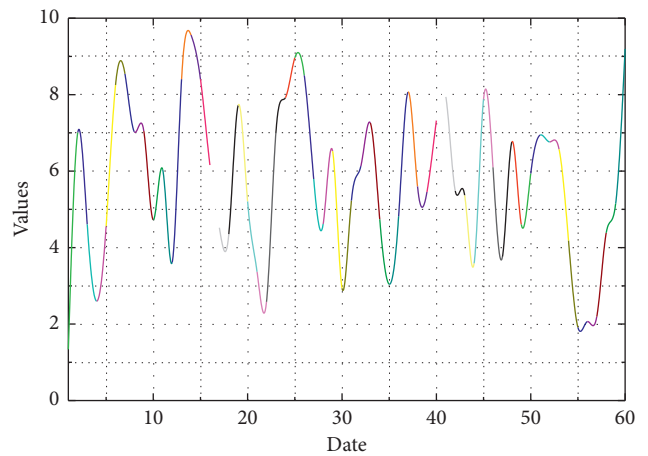


FIGURE 6: Time series plot of DAL stock prices.

implemented in many countries, further breaking down the barriers between financial markets, in which investors can allocate their assets to multiple markets as they wish, and cross-investment scenarios are gradually emerging. The stock and money markets of a country are influenced by the capital flows of these trading agents in the stock and money markets, thus creating a linkage between the two markets. At the same time, due to the gradual breaking down of the barriers to the flow of funds between the various financial

markets, trading entities can interoperate across financial markets for the financing of funds. In the current situation in China, the main trading entities in the stock market, such as securities companies, trust companies, and fund companies, can use the interbank lending market or the interbank bond market for short-term financing, and such financing activities will lead to a certain degree of capital flow from the money market to the capital market, and it is this flow of funds in the financial market that makes the intermarket. It is this flow of funds in the financial markets that make the linkage between the markets even closer.

4. Result Analysis

4.1. Forecast Analysis of Stock Prices and Printing and Dyeing Prices. Each stock and indices were analyzed by using Linear Regression, LASSO, Regression Trees, Bagging, Random Forest, and Boosted Trees. For each stock, the author used the percentage change in U.S. indices and the lagged variables of indices, as well as the lagged variables of percentage change for all stocks. Let me use the KNX stock as a specific example to illustrate my predictive analysis.

The linear regression model is the simplest. From the linear regression model, only the lagged variables of percentage change of SPX, a percentage change of ODFL, lagged variables of percentage change of SKYW, lagged variables of percentage change of MRTN, and lagged variables of percentage change of ALGT are significant. The R-squared is only 0.3105 and the MSE is 4.2346. When there are too many predictors, the linear regression is inadequate and difficult to interpret [23].

The LASSO model will give a more convenient result because it allows some coefficients to be zero. In addition, the cross-validation of the LASSO model shows that when we obtain the lowest MSE, the regression tree model results in 17 terminal nodes. We can see this result in Figure 8.

Most of the terminal nodes are on the right side of the tree, that is, when the percentage change in SPX closing price is greater than -0.64185 . In the first two levels, the regression tree model uses only the percentage change in the closing price of SPX. Finally, Bagging, Random Forest, and Boosted Trees are used. There are too many predictors in the model; so to understand the results, one can look at the significance plot. Figure 9 shows the importance plot for Boosted Trees, as shown in Figure 9.

The first two predictors are the percentage changes in the SPX and NDX. In fact, for most stock prices, the percentage changes in SPX and NDX are the most important predictors.

The study of stock market linkage theory considers stock market linkage mainly as a common trend of change in the prices of multiple stock assets in the stock market. The study of this common trend of stock market changes is usually examined from the two perspectives of return and volatility of stock indices. The specific scope of the study refers to four aspects of stock market linkage: linkage among stock indices in the world, linkage among different stock indices between the same countries, linkage among sectors and different industries in the same stock market, and linkage among individual stocks in the same sector. The most authoritative definition of market

contagion theory from a methodological point of view refers to the significant increase in linkages between stock markets in a country or a region after a financial crisis. This theory is consistent with the theory of increased stock market linkages in the wake of a financial crisis. Market contagion theory suggests that the crisis is mainly transmitted between stock markets of different countries through spillover effects, monsoon effects, and net contagion effects, and the corresponding mechanisms of contagion are trade spillover, financial spillover, industrial linkage, and net contagion. The relationship between linkage and market contagion is complementary; that is, the linkage between stock markets is established through different contagion channels, while the occurrence of financial crisis breaks the original linkage between markets, and the crisis strengthens the linkage between stock markets in the process of market contagion.

4.2. Naïve Heuristics and Portfolio Optimization. This paper determines the lag order of Johansen cointegration based on the SC and AIC criteria and uses the Pantula principle to determine whether the tested model has a deterministic trend term, a linear deterministic trend term, and a quadratic trend term. The Johansen cointegration test shows that there is a long-run equilibrium relationship between European and US stock markets, both in the prefinancial crisis period and during the subprime and European debt crises. This indicates that mature and developed capital markets such as the European and American stock markets have gradually developed economic base level linkages in the context of economic and financial globalization, and although the contagion effect caused by the international financial crisis will weaken the economic base level linkages, it is impossible to completely offset them, so the European and American stock markets will show long-term linkages. In addition, the widespread and rapid spread of technology has gradually increased the interdependence of European and American economies and strengthened the international economic coordination mechanism, and the globalization of the economy has continuously led to the globalization of finance; thus, the linkage between European and American stock markets has rapidly increased.

For portfolio optimization, the Naïve Heuristics method is used. Naïve Heuristics is based on stock price forecasts, ranking stocks based on potential returns and allocating an equal percentage of capital to each stock. Moreover, the stock price forecast is calculated from the formula generated by LASSO. And the potential return is today's closing price minus the previous day's closing price. Here are some recommendations for 10 stocks from March 8 to March 12, 2021, as shown in Figure 10.

For the daily data of a given stock, the KDJSV buy/sell strategy algorithm is used to calculate the output buy-sell points and perform buy and sell operations based on the buy and sell points. The effectiveness of the KDJSV buy/sell strategy is verified by calculating the stock return. The highest stock return is achieved when the long-short indicator growth rate was equal to the threshold $c = 4$. The experimental results are shown in column R8. Using the

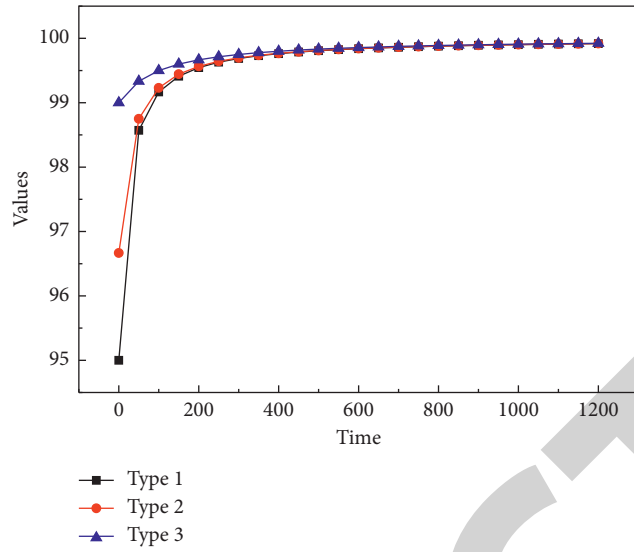


FIGURE 7: Time series plot of the percentage change of DAL.

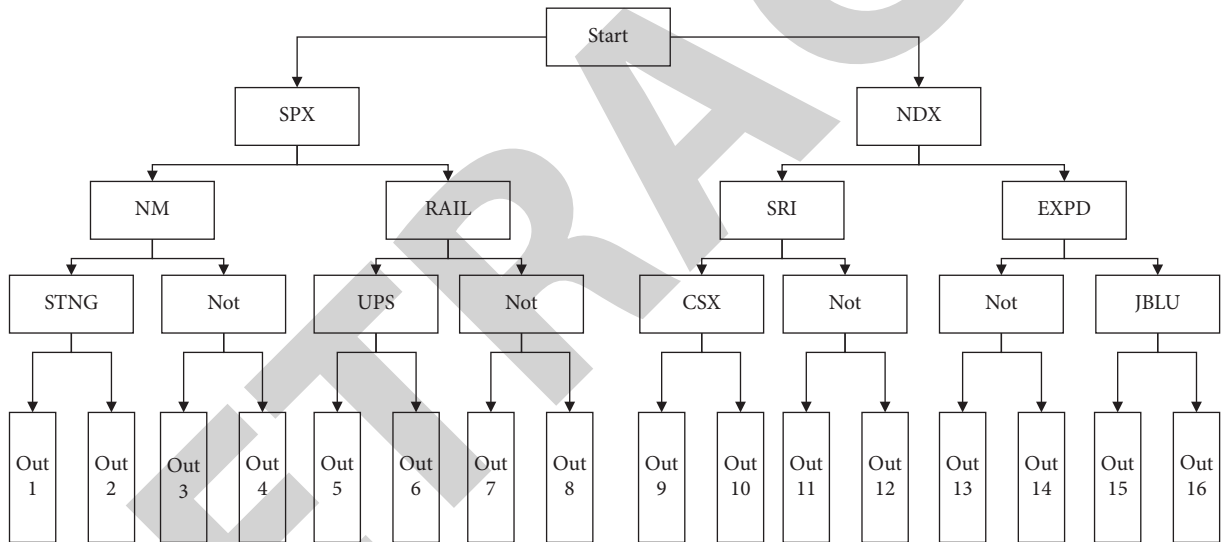


FIGURE 8: Regression tree model.

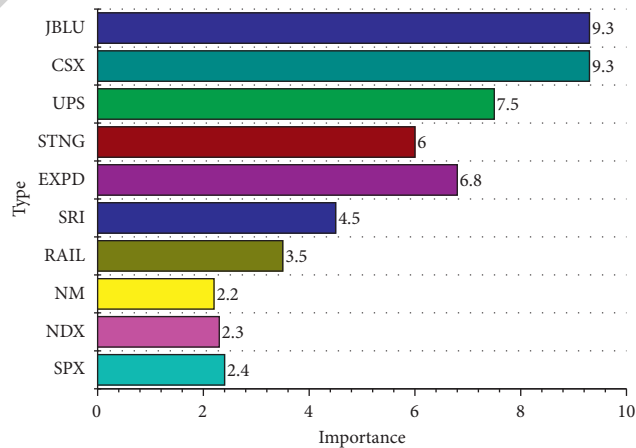


FIGURE 9: Importance of lifting tree diagram.

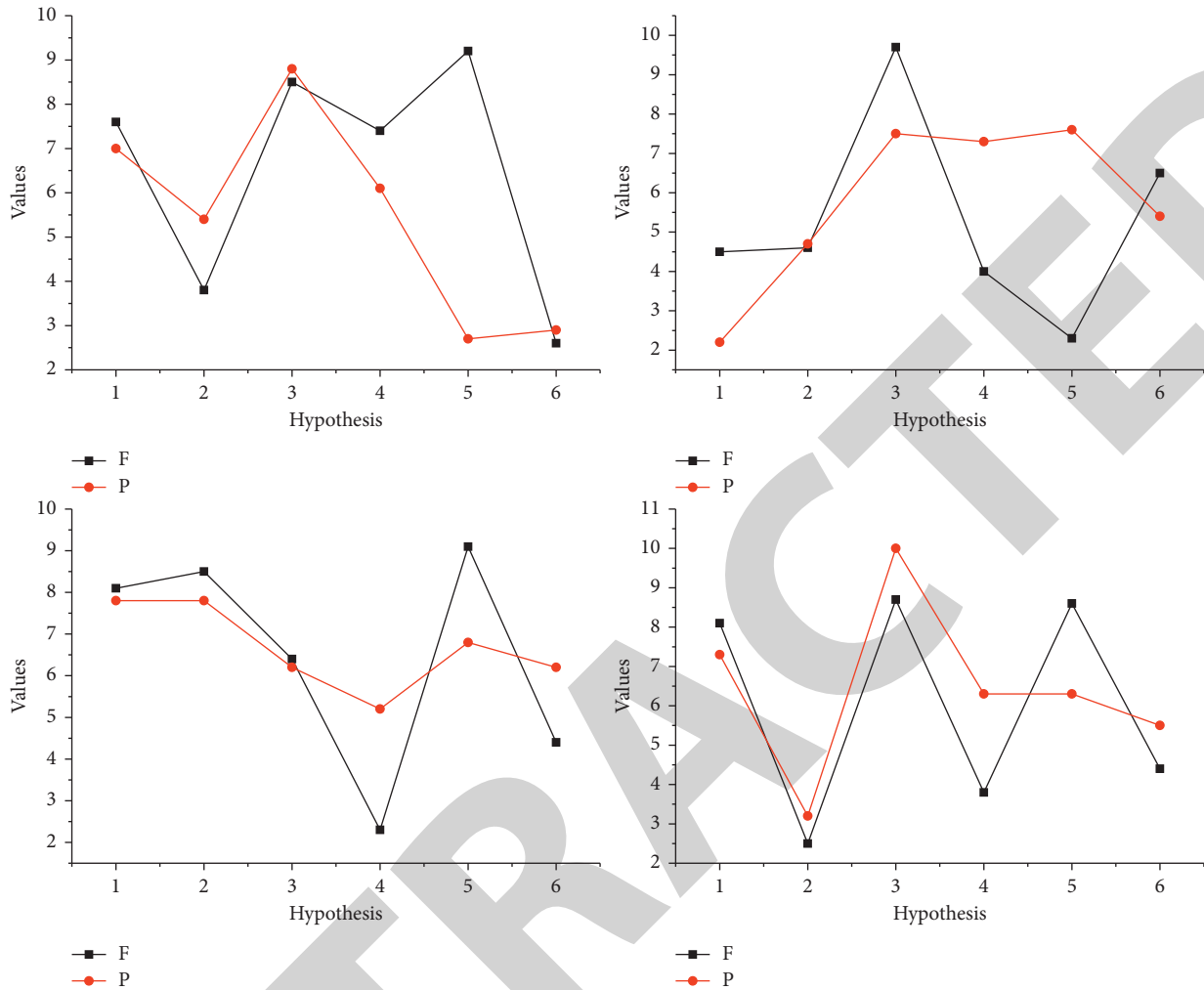


FIGURE 10: Causality test results.

KDJSV buy and sell strategy algorithm, 78.86% of the stocks had a return greater than 30% when $c=4$ and 5.76% of the stocks had a negative return.

5. Conclusion

Since the stock examples are all from the US market, the results show that their prices are highly correlated with SPX and NDX prices. The link between different stocks is not obvious. It is difficult to use other stocks to predict prices. However, stocks in the same area tend to move in the same direction. In addition, stock prices and indices prices tend to move in the same direction. However, the movement of indices prices will be smaller than the movement of stock prices. Stock returns are highest when the long-short indicator grows at a rate iDK threshold. Focusing on 52 representative stocks in different sectors of the Shanghai and Shenzhen stock markets, buying and selling points were calculated using daily data from 2013.10.1 to 2018.10.17, and the KDJSV buying and selling strategy algorithm was used. 78.86% of stocks with $c=4$ had returns greater than 30%, and 5.76% of stocks had negative returns. The DMISV buying and selling strategy is based on the

DMI indicator. The KDJSV buying and selling strategy is based on the KDJ indicator. The MACDV buying and selling strategy is based on MACD indicator. The DKB buying and selling strategy is proposed based on DMI, KDJ, and MACD indicators. The stock system is complex, and there are many influencing factors. In this paper, only the stock opening price, closing price, and other factors of the stock itself are selected, but other economic factors related to the stock such as macroeconomic and financial policies are not considered; therefore, the selection of other factors affecting the stock price fluctuation as independent variables to judge the stock price trend is one of the contents to be studied in the future.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest regarding the publication of this paper.

References

- [1] P. Yu and X. Yan, "Stock price prediction based on deep neural networks," *Neural Computing and Applications*, vol. 32, no. 6, pp. 1609–1628, 2020.
- [2] E. Akyildirim, A. Goncu, and A. Sensoy, "Prediction of cryptocurrency returns using machine learning," *Annals of Operations Research*, vol. 297, no. 1, pp. 3–36, 2021.
- [3] M. Z. Asghar, F. Rahman, F. M. Kundi, and S. Ahmad, "Development of stock market trend prediction system using multiple regression," *Computational and Mathematical Organization Theory*, vol. 25, no. 3, pp. 271–301, 2019.
- [4] A. Atkins, M. Niranjani, and E. Gerding, "Financial news predicts stock market volatility better than close price," *The Journal of Finance and Data Science*, vol. 4, no. 2, pp. 120–137, 2018.
- [5] Y. Song, J. W. Lee, and J. Lee, "A study on novel filtering and relationship between input-features and target-vectors in a deep learning model for stock price prediction," *Applied Intelligence*, vol. 49, no. 3, pp. 897–911, 2019.
- [6] S. C. Nayak, B. B. Misra, and H. S. Behera, "ACFLN: artificial chemical functional link network for prediction of stock market index," *Evolving Systems*, vol. 10, no. 4, pp. 567–592, 2019.
- [7] J. Cao and J. Wang, "Exploration of stock index change prediction model based on the combination of principal component analysis and artificial neural network," *Soft Computing*, vol. 24, no. 11, pp. 7851–7860, 2020.
- [8] W. Chen, H. Xu, L. Jia, and Y. Gao, "Machine learning model for Bitcoin exchange rate prediction using economic and technology determinants," *International Journal of Forecasting*, vol. 37, no. 1, pp. 28–43, 2021.
- [9] E. O. M. Ebadati and T. M. Mortazavi, "An efficient hybrid machine learning method for time series stock market forecasting," *Neural Network World*, vol. 28, no. 1, pp. 41–55, 2018.
- [10] T. Gao and Y. Chai, "Improving stock closing price prediction using recurrent neural network and technical indicators," *Neural Computation*, vol. 30, no. 10, pp. 2833–2854, 2018.
- [11] F. García, F. Guijarro, J. Oliver, and R. Tamosiuniene, "Hybrid fuzzy neural network to predict price direction in the German DAX-30 index," *Technological and Economic Development of Economy*, vol. 24, no. 6, pp. 2161–2178, 2018.
- [12] M. A. Ghazanfar, S. A. Alahmari, S. A. Alahmari et al., "Using machine learning classifiers to predict stock exchange index," *International Journal of Machine Learning and Computing*, vol. 7, no. 2, pp. 24–29, 2017.
- [13] S. Gu, B. Kelly, and D. Xiu, "Empirical asset pricing via machine learning," *Review of Financial Studies*, vol. 33, no. 5, pp. 2223–2273, 2020.
- [14] H. Sheng, S. Wang, Y. Zhang et al., "Near-online tracking with co-occurrence constraints in blockchain-based edge computing," *IEEE Internet of Things Journal*, vol. 8, no. 4, pp. 2193–2207, 2021.
- [15] B. M. Henrique, V. A. Sobreiro, and H. Kimura, "Stock price prediction using support vector regression on daily and up to the minute prices," *The Journal of finance and data science*, vol. 4, no. 3, pp. 183–201, 2018.
- [16] N. Huck, "Large data sets and machine learning: applications to statistical arbitrage," *European Journal of Operational Research*, vol. 278, no. 1, pp. 330–342, 2019.
- [17] D. Jothimani and S. S. Yadav, "Stock trading decisions using ensemble-based forecasting models: a study of the Indian stock market," *Journal of Banking and Financial Technology*, vol. 3, no. 2, pp. 113–129, 2019.
- [18] G. Kou, X. Chao, Y. Peng, F. E. Alsaadi, and E. Herrera-Viedma, "Machine learning methods for systemic risk analysis in financial sectors," *Technological and Economic Development of Economy*, vol. 25, no. 5, pp. 716–742, 2019.
- [19] I. E. Livieris, E. Pintelas, and P. Pintelas, "A CNN-LSTM model for gold price time-series forecasting," *Neural Computing and Applications*, vol. 32, no. 23, pp. 17351–17360, 2020.
- [20] N. Xiao, R. Xinyi, Z. Xiong et al., "A diversity-based selfish node detection algorithm for socially aware networking," *Journal of Signal Processing Systems*, vol. 93, no. 7, pp. 811–825, 2021.
- [21] A. Ntakaris, M. Magris, J. Kannianen, M. Gabbouj, and A. Iosifidis, "Benchmark dataset for mid-price forecasting of limit order book data with machine learning methods," *Journal of Forecasting*, vol. 37, no. 8, pp. 852–866, 2018.
- [22] M. Risse, "Combining wavelet decomposition with machine learning to forecast gold returns," *International Journal of Forecasting*, vol. 35, no. 2, pp. 601–615, 2019.
- [23] R. B. Parikh, C. Manz, C. Chivers et al., "Machine learning approaches to predict 6-month mortality among patients with cancer," *JAMA network open*, vol. 2, no. 10, Article ID e191597, 2019.