Frontiers in Business, Economics and Management
ISSN: 2766-824X | Vol. 6, No. 3, 2022

Stock Price Forecasting in Real Estate Industry Based on Investor Sentiment

Xiaoyuan Fan[1,*], Jiashuo Chen[2]

[1] School of Statistics, Shandong University of Finance and Economics, Jinan, Shandong, 250014, China
[2] Consulting Department, Shanke Digital Economy Research Institute Co., Ltd, Jinan, Shandong, 250014, China
[*]Corresponding author's e-mail: fxyuan902@163.com

Abstract: Stock price fluctuations are unstable, and investors' decision-making is also subject to the influence of their feelings. This study suggests a real estate stock price prediction approach based on investors' moods in response to the present phenomena that investors' desire to participate in the real estate development business is falling and risk aversion is increasing. Firstly, we quantify the stock bar comment scores combined with the Baidu search index to construct a composite sentiment index. Next, we combine additional stock price influencing factors and perform functional principal component analysis and dimensionality reduction to address the issue of multiple covariances. Finally, we input the index into CNN's prediction model to determine when the stock price will rise or fall.

Keywords: Sentiment analysis, Functional principal component analysis, CNN stock price prediction model.

1. Introduction

The real estate sector, a crucial component of the economic system and has a significant impact on the stock market, is essential for the orderly and healthy growth of China's capital market. China has experienced a meteoric rise in the number of Internet users as a result of the rapid development of mobile Internet technologies. According to the 49th Statistical Report on the Development Status of China's Internet, there are 1.032 billion Internet users in China. Investors in stocks frequently rely on their investment decisions on the information they may find online. It has been a hot topic for academic and commercial studies to use investor behavior as one of the stock prediction indicators.

2. Overview of Related Studies

The real estate sector plays a unique function in the stock market as a fundamental component of the economic system. The stock price predictions made by Zhu Yongming and Shao Gengyun using BP neural networks demonstrated the viability of neural networks in the field of stock price forecasting in the real estate business, but the input indicators are overly limited to micro-level indicators [1]. Li Junzhao, Guo Kun, and others accurately forecasted the trend of the real estate sector stock price index based on the regression model of the Markov chain. They also demonstrated the influence of the banking sector on the real estate sector [2]. Bai Yunfei and Wang Qi applied the robust regression model to track and predict the real estate index, and performed well in the extrapolation prediction, but did not consider macro factors [3]. Mao Zengli conducted a research on the prediction of the real estate stock price index based on the SSA-GRU recurrent neural network, which proved that the model has a strong reference significance for the trend prediction of the real estate stock price index, and provided a certain basis for formulating the real estate industry stock investment strategy [4].

In addition to the distinctive structure of Chinese stockholders, investor mood is also a factor affecting stock prices. Wysocki has shown that the quantity of online stock reviews can predict stock trading volume and return by analyzing more than 3000 stocks and forum comments on Yahoo! using cross-sectional data analysis and time series data analysis[5]. Yang Tou et al. revealed that the prediction accuracy of stock indexes using SVM-LSTM models containing sentiment polarity data is superior to that of conventional prediction methods [6]. Hongrui Zhao and Lei Xue employed a CNN model to efficiently extract deep features from the data and implemented an attention method to create accurate predictions for the SSE index [7].

This paper decides to use a combination of subjective and objective approaches, combining investor sentiment with convolutional neural networks and introducing functional principal component analysis to identify the primary characteristics of real estate industry stock data and forecast its movement trend in order to improve the accuracy of the forecast.

3. The Theoretical Basis of Functional Principal Component Analysis

In the context of the era of big data, with the increase in data volume, it can be noticed that the data all possess function-type properties. The fundamental principle of functional data analysis is to consider the observed data as a whole, instead of a single individual. After processing the research data functionally, more original data features can be mined, and the patterns of the research data and the significant sources of data changes can be investigated.

Functional data analysis involves fitting discrete raw data into overall functional data. Common fitting methods include basis function fitting, B spline basis function, Fourier basis function, and polynomial basis function. The basis function is usually chosen depending on the data's characteristics. Functional data analysis involves turning discrete data into function-based data. Since market indicators are strongly connected, and data changes are irregular, this paper uses functional principal component analysis to reduce dimensionality. This approach compresses several variables into a few variables that synthesize and reflect much of the information for later analysis. The integrated variables after PCA are not correlated, which overcomes the problem of

multicollinearity in the original data. Functional principal components can retain more data information than discrete principal components.

4. Experimental Procedure and Analysis of Results

4.1. Data sources

Stock price fluctuations are impacted by a variety of factors, and in order to improve forecasting accuracy, it is vital to collect data from multiple angles. Investor sentiment data and stock market data make up the majority of the data selected for this article.

4.1.1. Construction of investor compound emotion index

From eight stocks in the real estate industry with a total market capitalization of more than 50 billion yuan, three stocks are randomly selected: Wanke A (000002), Hua Qiaocheng A (000069), and Greenland Holdings (600606). The authors then crawl the stock bar comment data of the three selected stocks on the Oriental Fortune website, obtain the daily stock bar sentiment score by the method of text sentiment analysis, and combine it with the Baidu search engine. The index is developed to quantify investors' emotional traits.

4.1.2. Stock trading data

The stock trading data is selected from the 11 indicators of opening price, closing price, high price, low price, change amount, change rate, turnover rate, volume, transaction amount, total market capitalization, and market capitalization in circulation of the three stocks. This raw data is obtained from NetEase.com, starting on January 1, 2021, and ending on December 31, 2021, for a total of 243 trading days.

4.2. Function-based principal component analysis

4.2.1. Descriptive analysis of data

Figure 1 is a line graph of 11 indicators of Wanke A after standardization, which illustrates that there is no obvious periodicity in the trend of each indicator and that the change trends of several indicators are similar and correlated. Therefore, this paper employs the functional principal component analysis method for dimensionality reduction to extract the features of the original data of functional type for the functional principal component prediction below, and uses B spline basis functions for the functional principal component prediction. A functionalized fit is conducted to produce accurate data of the functional type.

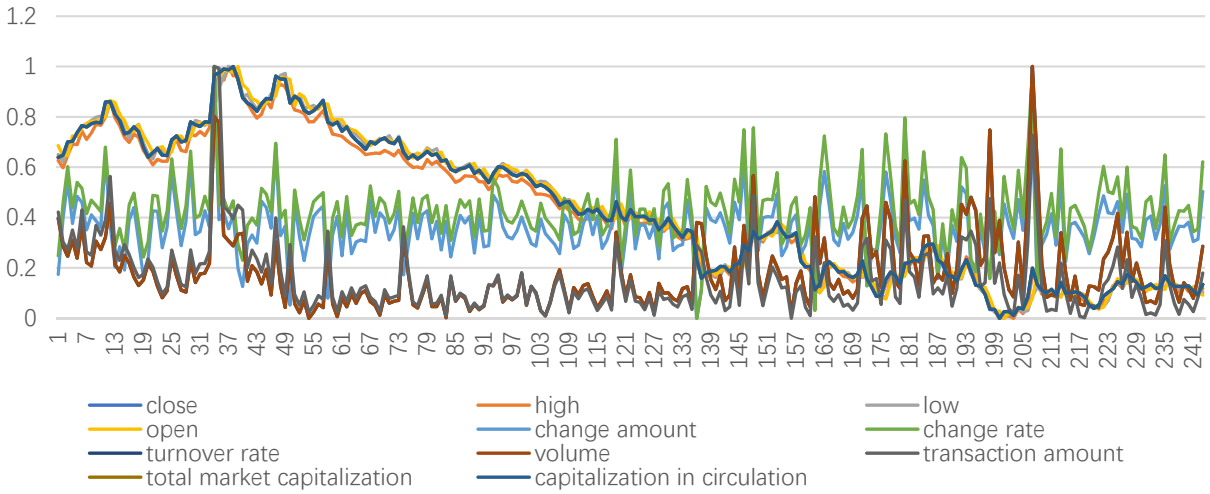

Figure 1. Wanke A line chart

Using Wanke A as an example, Figure 2 depicts the results of each indicator's data fitting following a rough penalty. The results demonstrate that the trend of eleven indicators over 243 trading days may be derived more accurately by employing the B-spline basis function. Change information and provide adequate raw materials for functional data.

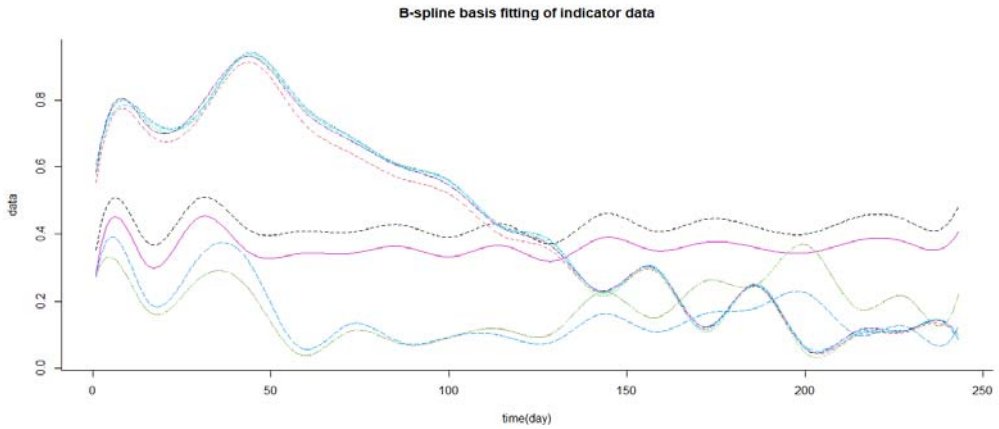

Figure 2. B spline basis function fitting results of Wanke A

4.2.2. Function-based principal component extraction

To analyze the factors that contribute to the trend of stock price changes over time, R software is used to generate a perturbation plot of the principal components against the mean function, which reflects the pattern of changes represented by each principal component, as well as the degree of variation in the variance explained by the principal components and the fluctuations of the weight function curve. The following are the perturbation plots of the principal components on the mean function.

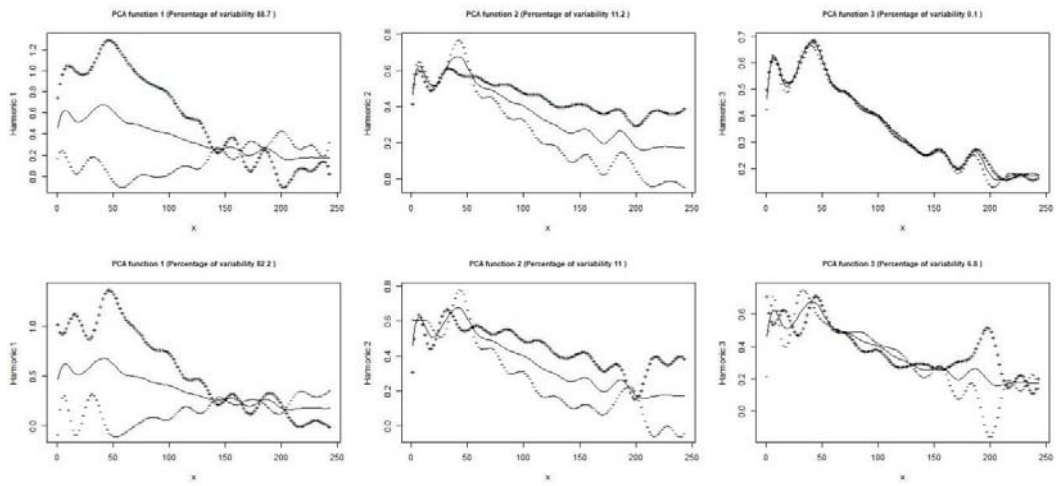

Figure 3. Main component disturbance chart of Wanke A without investor sentiment

As seen in the upper portion of Figure 3, the contributions of the three primary components are 88.7%, 11.2%, and 0.1%, respectively, with a total contribution of 100%, which reflects all of the data's features. Figure 3 bottom half depicts the function-type principal component curves following orthogonal rotation using the maximum variance method. The contribution rates of the three primary components are 82.2%, 11.0%, and 6.8%, respectively, and the cumulative contribution rate is 100%, which encompasses all of the fitted data's information. To avoid overfitting, the first two principal components are chosen for the subsequent prediction process.

Figure 4 shows the three functional principal components obtained by decomposition, in which the first principal component is high and stable in the first half of the year, and continues to decline in the middle of the year until the end of the year, and the first principal component is regarded as the stock price factor; the second principal component fluctuates significantly but the overall trend increases, and the second principal component is regarded as the stock trading situation factor.

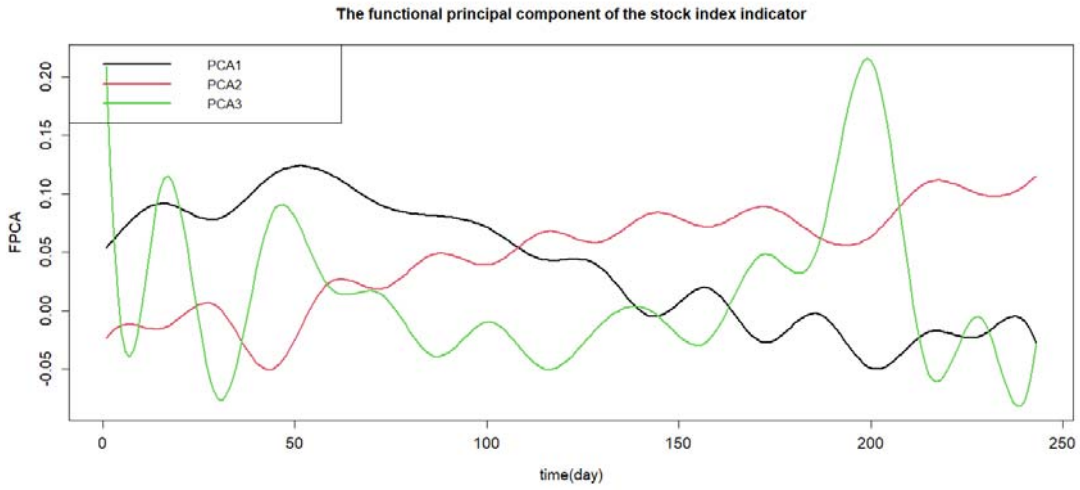

Figure 4. Wanke A Functional Principal Component

The investor sentiment index was not included in the processing of functional primary components described in the preceding section. The investor composite sentiment index was introduced as a new indication to the stock market data, and then the functional-type principal components were extracted using the same procedures as described previously. According to the main component perturbation plots, the functional principal components changed when the investor sentiment index was incorporated. The first three principal components can represent 99.2% of the characteristics of the indicator data, and the third principal component contributes 36% of the variance and contains more characteristics that cannot be eliminated, so it is reasonable to select the first three functional principal components for the indicator data of Wanke A with addition of the investor sentiment index.

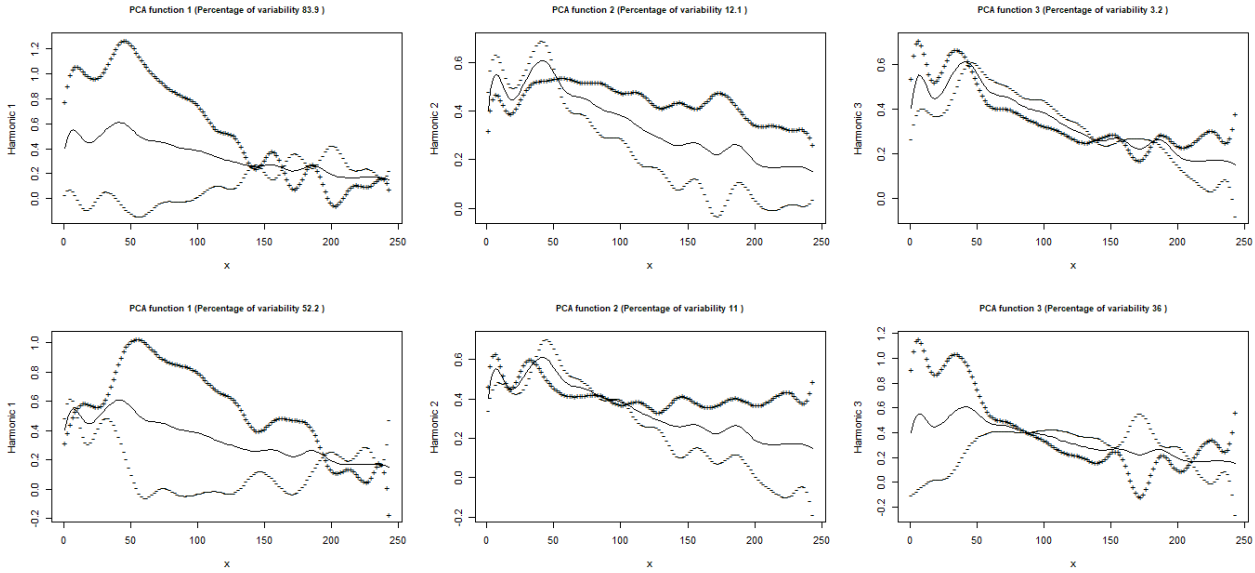

Figure 5. Main component disturbance chart of Wanke A with investor sentiment

Figures 6 depict the obtained functional main components: The first principal component is larger at the beginning of 2021 and decreases gradually beginning in the second month; the first principal component is regarded as a factor influencing stock trading data; the second principal component has occasional fluctuations during the year, but the overall trend is upward, similar to the movements of total market capitalization, market capitalization in circulation, and investor sentiment index; and the second principal component is regarded as a factor influencing stock trading data. The third primary component is a "U" shaped trend with highs at both ends of the year and a low in the center, and it is considered a trading element.

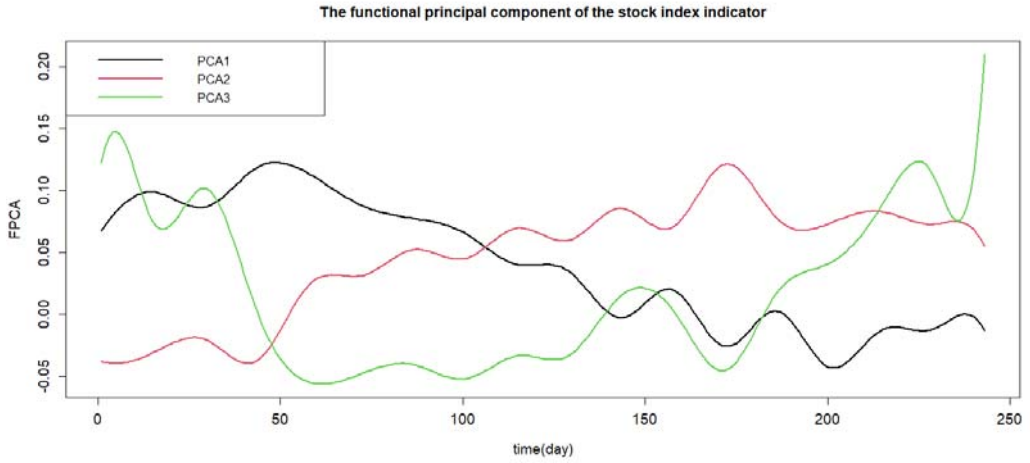

Figure 6. Functional principal component chart of Wanke A with investor sentiment

4.3. CNN prediction model

4.3.1. Model description

A CNN prediction model based on investor sentiment indices is constructed based on the closing price indicator series, and 243 trading days of data from January 1, 2021 to December 31, 2021 using a total of 13 indicators are used to predict the daily closing indices of Hua Qiaocheng A (000069), and Greenland Holdings (600606). In addition, a grid search technique is utilized to fine-tune the model's parameters in order to improve its predictive accuracy. The empirical test evaluates the efficacy of the CNN approach in predicting stock prices in the real estate business. Python 3.9 is the programming environment, and the extension libraries consist of Pandas, Numpy, Sklearn, Torch, Sys, etc.

4.3.2. Constructing the evaluation system of prediction results

The evaluation system consists of root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and goodness of fit (R^2).

4.3.3. Comparative analysis of stock price forecast results

Figure 7 depicts the loss of the training set and test set after each iteration of the model, using the variation of the loss function of the training set and test set of Wanke A stock as an illustration. As the number of model iterations rises, the loss function first decreases and eventually tends to become smooth; if interrupted prior to this smoothness, underfitting occurs. The number of iterations in this study is 100, and the loss functions of both the training and test sets have been smoothed out, so this amount is suitable.

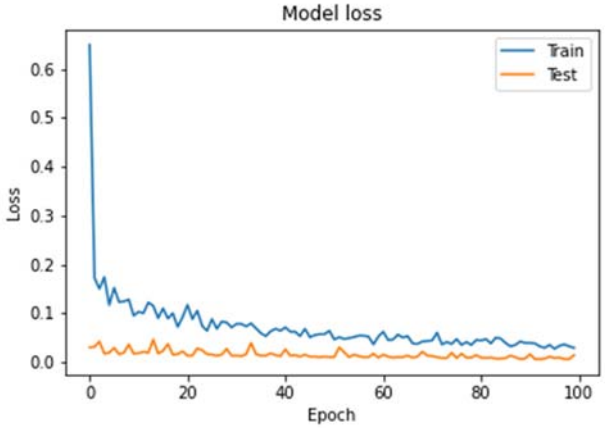

Figure 7. Losses in the training and test sets of Wanke A

Table 1 displays the findings of the model-based evaluation indices for the three equities. Following the addition of the sentiment index, the error class evaluation indexes are reduced to varying degrees, and all R^2 have improved. Therefore, it can be concluded that the addition of the investor sentiment index has contributed to the improvement in the accuracy of stock closing price forecasts. After adding investor sentiment to the model, the MAE, RMSE, and MAPE are rather tiny. All of them are greater than 89%, and all of the indicators are within a tolerable range, indicating that the model prediction results are accurate.

Table 1. Evaluation system of model prediction results for three stocks

	Wanke A		Hua Qiaocheng A		Greenland Holdings	
Investor sentiment	0	1	0	1	0	1
R^2	85.59%	94.79%	84.61%	90.39%	85.43	89.32%
MAE	1.41	0.95	0.38	0.32	0.20	0.16
MAPE	5.27%	4.37%	4.77%	3.89%	4.03%	3.22%
RMSE	3.73	1.35	0.24	0.15	0.06	0.05

The comparison between the three specific predicted values and the true value is shown in Figure 8, from left to right, Wanke A, Hua Qiaocheng A and Greenland Holdings, except for the inflection point prediction of the true value change, there is a certain error, the prediction effect is close to the true value at other time points, and the error is small.

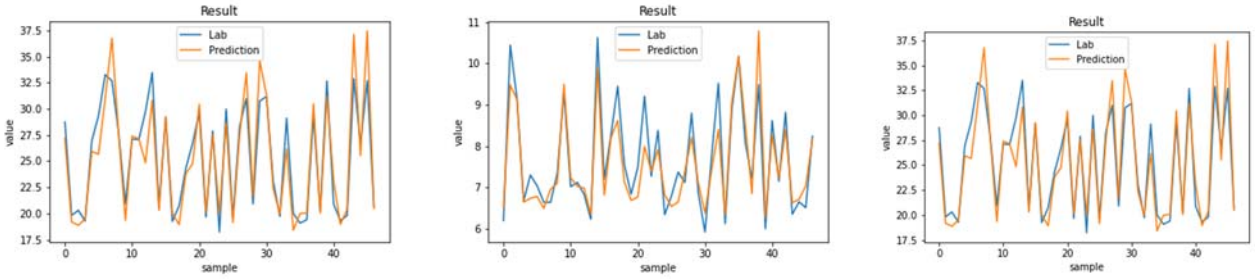

Figure 8. Graph of specific forecast results for three stocks

5. Conclusion

In the era of online big data, it is not scientific to base financial judgments just on stock trading data or internet information. To increase the accuracy of their predictions, investors must examine and filter the gathered investment information and combine it with the stock's own data. Based on investor sentiment indexes, this study provides a mechanism for predicting stock prices in the real estate business. Through text analysis of stock bar comments, quantified investor sentiment, combined with the Baidu search index to build a composite sentiment index, using the method of functional principal component analysis for dimensionality reduction, and inputting functional principal components into CNN for prediction, a good prediction effect is finally achieved, presenting a good idea and research direction for this field.

References

[1] Zhu YM, Shao GY. (2013) Stock price trend prediction based on BP neural network--a case study of listed real estate development companies. Finance and Accounting Monthly. 14:76-79.

[2] Li Junzhao, Guo Kun, Yao Hongliang, Wang Hao, Fang Shuai. (2014) Real estate sector index forecasting based on Markov blanket time series regression model. Systems Engineering Theory and Practice. 34(04):817-825.

[3] Bai Yunfei, Wang Qi. (2018) Application of robust regression models for tracking and forecasting of real estate indices. Journal of Chongqing University of Technology (Natural Sciences). 32(04):230-240.

[4] Mao Zengli. (2021) Research on real estate stock price index prediction based on SSA-GRU recurrent neural network. Harbin Institute of Technology.

[5] Wysocki P D. (1998) Cheap Talk on the Web: The Determinants of Postings on Stock Message Boards. University of Michigan Business School Working Paper. (98025).

[6] Yang T., Li W. L., Zheng S. H. (2020) Fusion of sentiment analysis and SVM-LSTM model for stock index prediction. Software Guide. 19(08):14-18.

[7] Zhao Hongrui, Xue Lei. (2021) Research on stock prediction based on LSTM-CNN-CBAM model. Computer Engineering and Applications. 57(03):203-207.