

Research on the Prediction Effect of Stock Closing Prices Based on Machine Learning

Zimeng Hao*

School of Artificial Intelligence and Data Science, Hebei University of Technology, Tianjin, China

*Corresponding author: 205215@stu.hebut.edu.cn

Abstract. In recent years, with the rapid development of the global economy and stock market, stock investment has become one of the most commonly used financial management methods. The main goal of this paper is to use the Python programming language to build a stock forecast model based on machine learning technology, select the model whose forecast effect is most in line with the real value through the forecast results of various models, and use this model to analyze my country's stock market. with the forecast. Using LSTM, KNN, moving average model, and linear regression model to predict and analyze the closing price of Ping An Bank stock, and finally, detect the value of the RMSE of each model. Machine learning technology has high research value in stock analysis and can effectively reduce the impact and impact of some stock market disasters on the economy.

Keywords: machine learning; stock prediction; neural network.

1. Introduction

In recent years, due to practical needs, the research and application of machine learning have received more and more attention. Machine learning can continuously optimize and improve prediction models based on the comparison of old and new data during the application process. Apply machine learning to stock market forecasts to mine important hidden information from historical stock data. This can provide not only theoretical support for shareholders but also decision support for the company.

LSTM neural networks are commonly used in natural language processing, It has also been researched in time for serious prediction of the stock by people., In our case, the results show that LSTM can do it. Currently, encoding and decoding networks are using LSTM networks, even for related gating units. In the aspect of translation, it performs well. LSTM inherits the characteristics of most of the RNN models, and it improves many of the RNN retention problems. It can solve the gradual reduction of the gradient backpropagation process that leads to the Vanishing Gradient problem. [1] Specifically, for problems that are highly related to time series and language processing tasks, LSTM is very suitable for dealing with them. It includes item prediction, dialogue generation, machine translation, and so on. The cognitive processes of behavior in humans, the development of logic, and the organization of neural networks are those that LSTM more precisely represents or simulates.

This paper conducts a simple strategic analysis of the stock data of Ping An Bank in recent years and selects the real closing price and the mean square error of the forecast results as the baseline for evaluation. The results show that the LSTM model has the strongest fitting effect on stock forecasts. The research conclusion shows that the application of machine learning to stock analysis and prediction can improve the efficiency of stock price information prediction and ensure the processing efficiency of massive data.

2. An Introduction to the Economic Situation

2.1 The general trend of China's economic market in these years

In recent years, the progress of China's economic market has been reflected in various aspects. Industrial production has gradually picked up, while equipment manufacturing has grown rapidly. The designated size increased by 4.2% yearly at the added value of industries, and the percentage is

0.4 points higher than the last month., and increased by 0.32% month-on-month. The equipment manufacturing industry increased by 9.5% per annum, 1.1 percentage points over the last month; the service industry continued to recover, and the modern service industry had good growth momentum. In August, the production of the national service industry index increased by 1.8% year-on-year growth. The percentage points are 1.2 over the previous month. From the perspective of market expectations, the service industry business activity expectation index is 57.6%; the market's sales recovery has accelerated, The total retail sales of social consumer goods growth rate have changed, and the positive growth replaced the negative growth. National online and offline retail sales totaled 8.429.5 billion, up 3.7% from the previous year. Among them, the retail sales of goods on the Internet were nearly 7.241.4 billion, which increased 5.8% over the previous period, accounting for 25.6% of the total retail sales of social consumer goods; investment in fixed assets rose steadily; High-tech industries were invested more and more money; scientific and technological achievements were transformed into service industries; R&D and design Investment in the service industry increased by 20.0% and 16.9%, respectively. The social sector experienced an increase of 14.1%; the import and export of goods continued to grow, and the proportion of general trade increased; The general trend of employment is stable, and the unemployment rate has declined not only in the city but also in the countryside [1-3]. China's overall economic development tends to rise steadily, various fields are developing in an orderly manner, and China's new capital is accelerating its entry into fast-growing industries.

However, due to the current situation and the influence of force majeure factors in recent years, both at home and abroad still have many unstable and uncertain factors. The situation in China will turn complicated. From the perspective of the external situation, the international situation, supply chain shortages, global inflation, and other problems that cannot be completely solved in the short term, economic recovery and growth are still facing uncertainty. From the perspective of the domestic situation, my country's economic development is facing pressures such as lower expectations, shrinking demand, and oversupply, resulting in certain fluctuations in economic development. Therefore, it is even more necessary for us to find the best stocks in the stock market to minimize our economic losses and the country as a whole, to adapt to the current unpredictable domestic and foreign environment [4].

2.2 Four Stock Market Crashes' Effects in China

China has experienced four big hits in the equity market in the 20th and 21st centuries. The stock market is the national economy's direct reflection, and the occurrence of the crush of the stock market signals the beginning of an economic recession [5]. A significant impact on the entire stock market and financial markets will occur because of the stock market crash. The non-performing assets of banks will increase, and people's investment opportunities will decrease, resulting in an outflow of funds, causing currency devaluation, and then impacting the entire financial market, completing a serious production boycott, a sharp drop in national income, and the economy will fall into a vicious circle.

The first stock market crash was on December 16, 1996, and was called "Black Monday" by investors. Before 1996, the Chinese stock market had experienced a bull market. That bull market happened in 1993, and there will be a bear market after the bull market. At that time, many mechanisms in the Chinese stock market were not perfect, and junk stocks were rampant. With the increase in junk stocks in the market, the stock price rose sharply, but it lacked a support point, so it eventually suffered a stock market crash.

The second stock market crash occurred on July 30, 2001. After the baptism of the last stock market crash, the Chinese stock market also underwent rectification. After many listed companies entered the stock market, there were fewer and fewer junk stocks and more and more high-quality stocks, so the performance of the Chinese stock market in those five years was very stable, but at the end of 2001, there were some fluctuations in the Chinese market. A-shares had another serious stock market crash.

The third stock market crash occurred on February 27, 2007, a year before the arrival of the Asian financial crisis. At that time, the mechanism of the stock market was perfect, but there was a lot of confusion. Behind the stock market, there were various people who manipulated the stock market. Manipulate the stock market behind it. Therefore, the normal development of the stock market was affected, and the decline of the Shanghai Composite Index was constantly breaking records and finally ushered in a stock market crash.

The fourth stock market crash happened in 2015. In just one year, the Chinese stock market plummeted three times, so the stocks at that time all fell very badly. Therefore, the market value of the stock market evaporated by 3 trillion on January 19 and May 28. The Japanese stock market hit shareholders hard again. A-shares fell 6.5% in one day, and even some "technical" stocks fell severely. Throughout 2015, China's stock market has been hit hard [6-7].

3. Analysis of Method and Result

Given the changes in the closing price of Ping An Bank of China in the past ten years, four models were selected to predict the closing price of the stock in the first half of this year. When compared with the actual data, the prediction model that is closest to the true value is selected.

3.1 Model of prediction

The four prediction models are as follows:

1. LSTM is a time-series recurrent neural network and also a long-term memory network, which aims to solve the problem of gradient explosion and disappearance during long-sequence training. It has better performance than other models, and it can complete many projects that other models cannot. LSTM already has a variety of applications in science and technology. Nowadays, LSTM systems can control artificial intelligence, translate different languages, perform text summarization, image and text analysis, and can also be used for tasks such as handwriting recognition, image recognition, speech recognition, predicting diseases, predicting stock closing prices, and synthesizing multi-track music [8].

2. KNN is one of the simplest machine learning algorithms proposed by Cover and Hart in 1968 and belongs to the classification algorithm in supervised learning. Only training samples need to be saved, no parameter estimation is required, and it is not susceptible to small errors. A sample also belongs to a class if most of the K nearest neighbor samples in the feature space belong to that class.

3. Simple moving averages. The current value of the time series is a model formed by the random error term and the linear function of the lag error term.

4. Linear regression. A mathematical model that determines the correlation between variables shows a linear or non-linear relationship between the independent variable and the dependent variable, which continuously changes the value of the coefficient to make the deviation between the predicted value and the actual value smaller and smaller.

To compare how well they fit, I will use the RMSE to test the predictions, judge the accuracy of the predicted data, and finally determine the feasibility of the model. Choose the optimal stock forecasting model.

3.2 Results Discussion

In terms of accuracy, LSTM performed the best, followed by linear regression; moving average and KNN algorithm models performed the worst; moving average and linear regression performed slightly better than KNN in predicting future stock price trends. It is because they remove the volatility of the data and show the long-term trend of the data. The moving average only relies on the previous time data to predict the stock price on the next trading day. Linear regression and KNN rely on date features to fit regression models. Neither traditional linear models nor machine models can mine the implicit information in the data. It is difficult to achieve satisfactory performance on nonlinear data with large fluctuations and strong noise.

Table 1. Four models RMSE comparing

Model	Moving Averages	Linear Regression	KNN	LSTM
RMSE	8.8579	4.7651	12.9326	0.4424

Through RMSE, it can be seen that the value of LSTM is the smallest, and the value of KNN is the largest. It can be seen that the LSTM is more consistent with the ground truth. Therefore, LSTM is the most accurate stock prediction model among these models (Table 1).

4. Discussion

Some features of the deep learning model make it more suitable for processing predictive financial time series data [9].

1) The deep learning model is not limited by dimensions, and all feature data related to the dependent variable can be incorporated into the model to obtain a more complete representation.

2) It has a good nonlinear fitting ability and is suitable for fitting time series data with irregular fluctuations.

3) It is difficult to fall victim to model over-fitting and local optimum.

4) It does not require manual feature construction, extracts data hidden features through multi-layer neural networks, and has the stronger expressive ability of deep networks when compared to linear regression and traditional machine learning. And the LSTM network selected in this paper has the function of long-term memory, so the prediction effect of high-noise financial stock data will be better.

But the same LSTM also has disadvantages:

1) There are disadvantages to parallel processing. Compared with some of the latest networks, the relative effect is average.

2) LSTM has made some progress on the gradient problem, but there is still a gap from the ideal state. Because the length of the sequence it can handle is limited, it can still not predict accurately when it encounters a longer sequence.

3) Calculation requires time. There are four MLPs per LSTM cell. If the time series span of LSTM is very large and the network is very deep, then this calculation amount will be very large for the computer [10].

5. Conclusion

Under the current economic situation, more and more people are starting to invest in the stock market. Choosing the best stocks as the latest financial management method based on the stock price trend and certain forecasts is a key step before investing in stocks. The adjustment and fusion of models in machine learning and data features make it possible to more accurately predict stock trends. And its application in the stock market can minimize the stock market crash has a significant impact on people's lives and will cause economic panic.

In the complex stock market environment, neural network algorithms have been well practiced for stock prediction. With its good learning performance and strong simulation ability, compared with the traditional econometric method, the neural network can be used in financial time. It is more advantageous in sequence prediction.

In terms of stock prediction, the four models each have their characteristics. Although LSTM performs the best among the four models, it still cannot be completely solved for sequences with larger magnitudes, and the longer time consumption is also an important factor. A major drawback is that in future projects, it is hoped that LSTM can be more optimized for the shortcomings of LSTM to make the prediction results fit better.

References

- [1] Jiang Bolin, Yue Jie, and Lan Shengjie[J] Investigation of stock prediction systems based on LSTM neural networks.Hebei Institute of Architecture and Engineering Journal, 39(4): 165-170, 2021.DOI: 10.3969/j.issn.1008-4185.2021.04.030
- [2] Shen Liping, Wu MingmingAn Empirical Study on the Development and Economic Growth of China's Stock Market [J]. 2009(1):71-72,99. Technical Economics and Management ResearchDOI: 10.3969/j.issn.1004-292X. 2009.01.022.
- [3] Chen Shiwei. The Influence of Stock Market Prosperity on China's Economic Development—A Cold Reflection on the Current Stock Market Upsurge [J]. 143-146 in Jiangxi Social Sciences, 2007(11).DOI: 10.3969/j.issn.1004-518X.2007.11.026,
- [4] Zhou Jianbo, Chen Hao, and Sun Shubin.From the standpoint of the Literature History Review [J].Journal of Shandong University(Philosophy and Social Sciences Edition), 2021(4):50-64. DOI: 10.19836/j.cnki.37-1100/c.2021.04.005.
- [5] Yin Shuhao, Zeng Wei.Analysis of China's stock market disaster relief and preventive measures [J]. North China Finance, 3(3), pp. 26-29.DOI: 10.3969/j.issn.1007-4392.2016.03.005.
- [6] Teng Yiyang. Analysis of the Causes and Influence Countermeasures of China's Stock Crash in 2015 [J]. 2017(36):76-77. China Collective Economy.DOI: 10.3969/j.issn.1008-1283.2017.36.042.
- [7] Wu Xianbo and Hui Xiaofeng.Research on the regional interdependence of China's stock market and its dynamic evolution: the analysis background of the 2015 stock market crash [J]. Complex Systems and Complexity Science, 2020, 17 (2), 1-10. DOI: 10.13306/j.1672-3813.2020.02.001.
- [8] Li Guicheng, Xu Li, and Zhang Li. LSTM-Based Stock Price Prediction Analysis [J].12 (5): 123-128 in Intelligent Computer and Applications, 2022.DOI: 10.3969/j.issn.2095-2163.2022.05.023.
- [9] Wang Bo and Cheng Fuyun.Application of the KNN Algorithm in Stock Prediction [J]. Science and Technology Entrepreneurship Monthly, 2015, 28(16): 25-26. DOI: 10.3969/j.issn.1665-2272.2015.16.011.
- [10] Fu Hangcong and Zhang Wei. Application of Machine Learning Algorithms in Stock Trend Prediction [J]. DOI: 10.11907/rjdk.171549. Software Guide, 16(10), 31-34, 46, 2017.