Check for updates

Research on Stock Market Investment Model Based on Time Series Forecasting and Dynamic Programming

XuDong An, Xuan Yi, BeiQi Zhou, * QinJuan Zhang

China Telecommunications Terminal Laboratory, China Academy of Information and Communications Technology, Beijing, China

* Corresponding author: zhangqinjuan@caict.ac.cn

Abstract. Forecasting the value of a stock population has always been attractive and challenging for shareholders due to its inherent dynamics, nonlinearity and complexity. In this paper, we propose a stock investment model based on time series forecasting and dynamic programming. The time series forecasting model is utilized for next day high- and low-price prediction, combined with the dynamic programming model to formulate stock trading strategies. The study conducted simulated back tests on 200 stocks randomly selected from the Chinese Shanghai and Shenzhen stock markets, and the results show that the scheme proposed in this study can achieve a return of more than 12% after more than a 3-month investment cycle.

Keywords:time series forecasting; dynamic programming; stock prediction; data mining.

1 Introduction

The stock market has a high yield and high-risk coexisting characteristics, if handled properly, the stock market buying and selling can make people rich, may also make people a pot of money [1]. In order to profit and avoid risk, investors have to pay attention to the stock market, analyze the stock market, explore its inherent laws, looking for effective prediction methods and tools, so as to find the factors affecting the change of stock market sentiment, and try to predict the development trend of the stock market. However, the stock market is a complex system with many influencing factors and various uncertainties, and its price fluctuations often show strong non-linear characteristics.

Stock market forecasting must first recognize that there are certain patterns in the stock market. And these laws are fully affected in the stock market's historical data, the purpose of continuing stock market forecasting is to movement some feasible way to find out the laws from the data, and use these laws to the development trend of the stock market to research and judgement, which can help to guide the investment decision.

The process of predicting stock values has always been a challenging problem [2] because of its long-term unpredictability. The outdated market hypothesis suggests

Y. Jiao et al. (eds.), *Proceedings of the 3rd International Conference on Internet Finance and Digital Economy (ICIFDE 2023)*, Atlantis Highlights in Economics, Business and Management 1,
https://doi.org/10.2991/978-94-6463-270-5_46

that it is impossible to predict stock values and that the behavior of stocks is random, but recent technical analysis has shown that most stock values are reflected in previous records; therefore, movement trends are crucial for effective prediction of values [3]. In addition, stock market groups and trends are influenced by several economic factors such as political events, general economic conditions, commodity price indices, investor expectations, other stock market trends, and investor psychology [4]. The ambiguous nature of stock value movements makes investors' investments risky. In addition, it is often difficult to detect the market position of governments. Indeed, stock values are usually dynamic, nonparametric, and nonlinear; therefore, they often lead to poor performance of statistical models that fail to predict accurate values and movements [5,6].

Machine Learning (ML) is the most powerful tool that contains different algorithms that can be efficiently developed for their performance in specific case studies. It is widely recognized that ML has the remarkable ability to identify valid information and detect patterns from datasets [7]. In contrast to traditional methods in the field of ML, integrated modeling is a machine learning based approach in which some of the commonly used algorithms are used to solve specific problems and have been shown to outperform each method in predicting time series [8,9,10]. For prediction problems in the field of machine learning, boosting and bagging are effective and popular algorithms among the integrated methods. Tree-based models have recently made some progress with the introduction of gradient enhancement and XGBoost algorithms, which have been heavily used by top data scientists in competitions. In fact, a modern trend in ML, Deep Learning (DL), can be considered as a deep nonlinear topology in its specific structure and has a remarkable ability to extract relevant information from financial time series [11]. Contrary to simple artificial neural networks, recurrent neural networks (RNNs) have achieved great success in the financial field due to their superior performance [12,13]. All researchers agree that stock price prediction and modeling has been a challenging problem for researchers and speculators due to the noisy and non-stationary nature of the data.

This study proposes a stock finance model that integrates time series forecasting and dynamic planning, which predicts the highest and lowest prices of stocks on the following day by introducing weighted lightGBM, and combines the available funds in hand and the investment strategy of dynamic planning to achieve a more stable return in longer-term stock market investment.

2 Material and Methods

2.1 Dataset

The purpose of this study is to provide short-term forecasts of the Chinese stock market, with data derived from stock market data from January 2019 to December 2022 (4 years), including open, close, low-high, and price to calculate 10 technical indicators. The data are mainly collected using AKShare [14], which is a Python-based financial data interface library, aiming to realize a set of tools for fundamental data, real-time and historical ticker data, and derivative data of stocks, futures, options, funds, forex, bonds, indices, cryptocurrencies, and other financial products, from data collection, data cleansing, to data landing, which is mainly used for academic research AKShare is characterized by the acquisition of raw data published by relatively authoritative financial data websites, and by utilizing the raw data for cross-validation between various data sources, and then reprocessing it to draw scientific conclusions.

2.2 Time Series Forecasting Model

The widespread use of GBDT in machine learning tasks for tabular data in recent years is largely due to open source GBDT tools such as XGBoost, LightGBM and CatBoost. These tools are meticulously optimized in terms of performance (accuracy) and efficiency. LightGBM is an efficient open-source implementation of GBDT, which is currently used in many applications in industry. The parameters of the LightGBM algorithm are relatively complex, and can be broadly classified into the core parameter, the learning control parameter, the IO parameter, the target parameter, the metric parameter, etc., and the ones that generally need to be tuned are the core parameters, learning control parameters, and metric parameters. Considering that the price of the stock has a stronger correlation with the recent stock market situation and a weaker correlation with the past stock market, the study adopts the data weighting scheme.

2.3 Investment Strategies Based on Dynamic Planning

Dynamic Programming is a branch of operations research that solves for the optimization of a decision-making process.

In order to maximize the benefits of stock investment, we need to choose the right time point and price to buy and sell according to the stock market situation. To fully describe the recursive problem, you need to first define the variables of the recursion, i.e., the control variables of the recursive process. Before defining the control variables, it is first necessary to define which variables are the relevant quantities for the recursion. First of all, we defined the stock can be purchased time interval period, time interval period can be sensibly understood as an array of subscripts index, each line standard up to the standard trading day of the day; Secondly, we need to define the daily trading status, trading status there are three, the trader can choose not to carry out any transactions, the choice to buy, the choice to sell a total of three status (as shown in Fig. 1). Due to the title of the limit can be up to two operations per day, so we need to introduce the variable k to record the number of transactions, a round of trading is defined as the success of a buy plus the success of a sell, a round of trading is completed, we need to add the variable k to (k + 1).

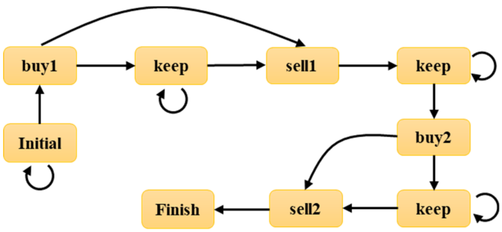


Fig. 1. Change in status of stock trading

If there are no buys on the day, then that means no operations are performed on the day, then it is equivalent to keeping the initial value unchanged, which is equivalent to self-adding to index and keeping the other variables unchanged.

A new day from the purchase of shares to start buy1, after the purchase of shares, hold the current stock without any operation can get the maximum profit, then the day's trading is over, waiting for a day when the highest stock price and then sell; or buy, you can immediately sell, assuming that the day's stock price fluctuations did not occur, it is equivalent to the premise of no gain, directly throw off the hands of the stock; the first time after the sell1. There can be two choices, choose to hand no longer allotment, or immediately buy buy2. the back and the front of the same reason, when the second sell completed, due to the limitations of the number of transactions on the day, the two-trading quota has been used up, the end of the stock trading activities.

2.4 Evaluation Measures

Mean Absolute Percentage Error (MAPE) is often employed to assess the performance of the prediction methods. MAPE is also a measure of prediction accuracy for forecasting methods in the machine learning area, it commonly presents accuracy as a percentage,

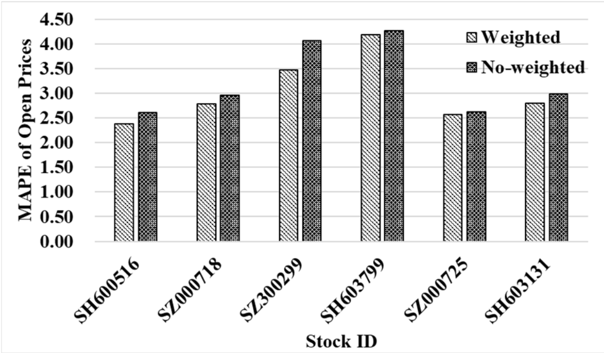
$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \times 100 \quad (1)$$

where A_t is the actual value and F_t is the forecast value. In the formula, the absolute value of the difference between those is divided by A_t. The absolute value is summed for every forecasted value and divided by the amount of data.

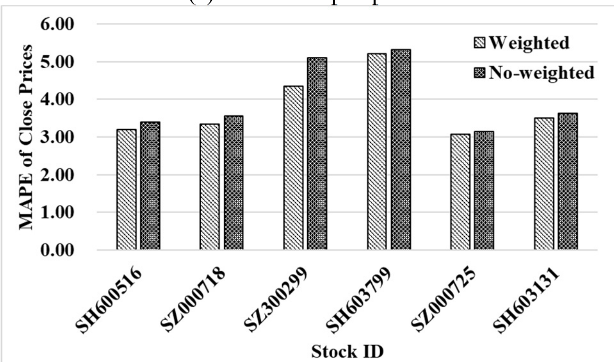
3 Results

3.1 Accuracy of Time Series Forecasting

The study analyzed the prediction results of opening and closing prices of different stocks using weighted and unweighted LightGBM. The accuracy of the prediction results was evaluated using the MAPE metric (Fig. 2), and the results show that low MAPE results can be obtained using the weighted LightGBM, i.e., the weighted LightGBM yields more accurate fitting results.



(a) MAPE of open prices



(b) MAPE of close prices

Fig. 2. Results of price prediction using weighted and unweighted LightGBMs

Figure 3 shows the results of predicting stocks using weighted and unweighted LightGBM.

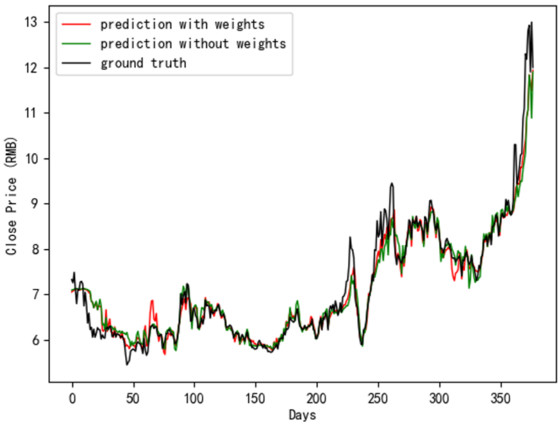


Fig. 3. Results of close price prediction using weighted and unweighted LightGBMs

3.2 Evaluation of the Effectiveness of Benefits

The study utilizes weighted LightGBM to predict the high and low price of a stock on the next day, combined with a dynamic programming algorithm for buying and selling stocks. The study randomly selected 200 stocks and randomly selected the time when the stocks were opened with a principal of ￥200,000 RMB. The returns after 1 month, 2 months, 3 months, 3-6 months, 6-12 months and 1 year after reducing the position were analyzed (Fig. 4). It can be seen that using the algorithm proposed in this study, an average return of more than 12% can be obtained after 3 months, more than 20% after 6 months, and more than 55% after 1 year.

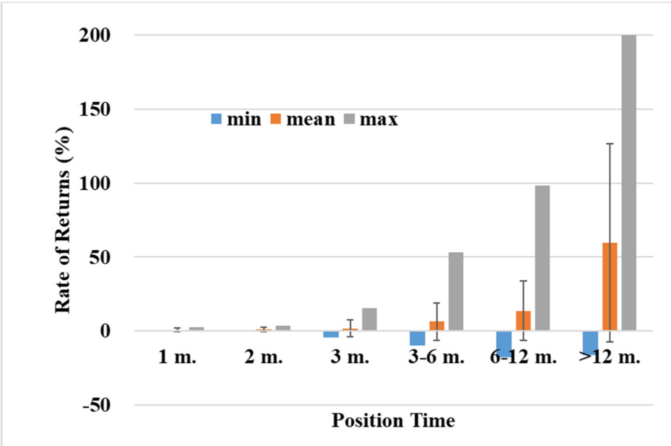


Fig. 4. Average return on investment for 200 stocks

Figure 5 shows the buying and selling operations, positions and returns for a stock using the algorithm proposed in this study for a period of 2 years from the time the position was opened. It can be seen that the algorithm proposed in this study does not achieve a profit every time it sells, but it can achieve a better return result that can be obtained in long term investment.

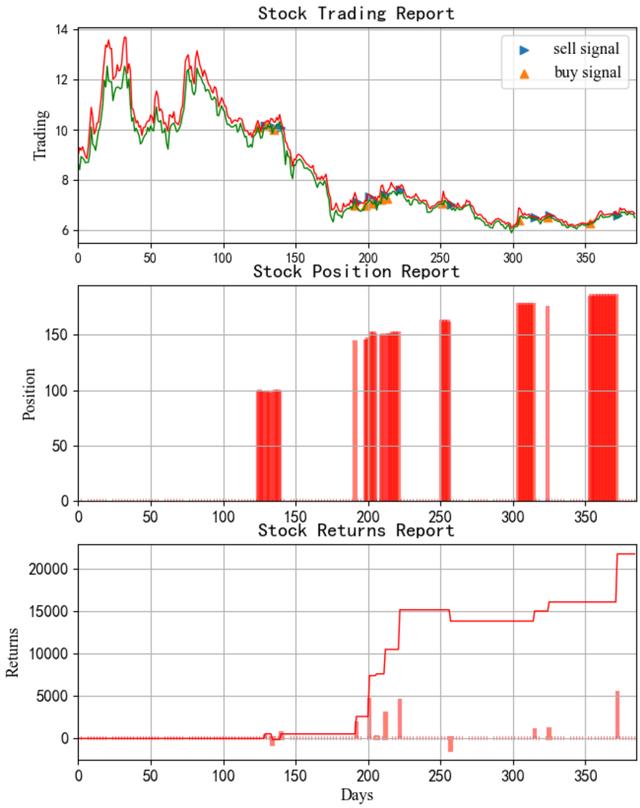


Fig. 5. Buying and selling operations, positions, and returns for a stock over a 2-year period from the time the position was established

4 Conclusions

This study proposes a stock finance model that integrates time series forecasting and dynamic planning, which predicts the highest and lowest prices of stocks on the following day by introducing the weighted LightGBM, and combines the available funds in hand and the investment strategy of dynamic planning to achieve more stable returns in longer-term stock market investments. The experimental results prove that a better return performance can be realized after 3 months of position building using the scheme proposed in this study. Of course, there are some shortcomings in this study, for example, when the stock market is in bear and bull markets, the stock market returns are affected by the inaccurate prediction of the stock prices, resulting in the failure to achieve immediate buy and sell operations, which may require the inclusion of more quantitative strategies to optimize the buy and sell operations. In order to obtain more stable return results. Later on, we will conduct a more in-depth study on this issue to obtain a more stable return on investment.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (No. 2022YFF0604300).

References

1. J. Zou, Q. Zhao, Y. Jiao, et. al, "Stock Market Prediction via Deep Learning Techniques: A Survey," arXiv:2212.12717, 2022.
2. S. Asadi, E. Hadavandi, F. Mehmanpazir, et. al, "Hybridization of evolutionary Levenberg–Marquardt neural networks and data pre-processing for stock market prediction," Knowl.-Based Syst., vol. 35, pp. 245–258, 2012.
3. S. Akhter, M. A. Misir, "Capital markets efficiency: Evidence from the emerging capital market with particular reference to Dhaka stock exchange," South Asian J. Manag., pp. 12–35, 2005.
4. K. Miao, F. Chen, Z. Zhao, "Stock price forecast based on bacterial colony RBF neural network," J. Qingdao Univ. (Nat. Sci. Ed.), vol. 2, pp. 210–230, 2007.
5. M. P. Naeini, H. Taremian, H. B. Hashemi, "Stock market value prediction using neural networks," Proceedings of the 2010 international conference on computer information systems and industrial management applications (CISIM), pp. 132–136, 2010.
6. B. Qian, K. Rasheed, "Stock market prediction with multiple classifiers," Appl. Intell., vol. 26, pp. 25–33, 2007.
7. E. S. Olivas, Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques: Algorithms, Methods, and Techniques. IGI Global: Hershey, PA, USA, 2009.
8. M. Ballings, D. V. D. Poel, N. Hespeels, et. al, "Evaluating multiple classifiers for stock price direction prediction," Expert Syst. Appl., vol. 42, pp. 7046–7056, 2015.
9. M. M. Aldin, H. D. Dehnavi, S. Entezari, "Evaluating the employment of technical indicators in predicting stock price index variations using artificial neural networks (case study: Tehran Stock Exchange)," Int. J. Bus. Manag. vol.7, pp. 25, 2012.
10. C.-F. Tsai, Y.-C. Lin, D. C. Yen, et. al, "Predicting stock returns by classifier ensembles," Appl. Soft Comput., vol. 11, pp. 2452–2459, 2011.
11. R. C. Cavalcante, R. C. Brasileiro, V. L. Souza, et. al, "Computational intelligence and financial markets: A survey and future directions," Expert Syst. Appl., vol. 55, pp. 194–211, 2016.
12. S. Selvin, R. Vinayakumar, E. Gopalakrishnan, et. al, "Stock price prediction using LSTM, RNN and CNN-sliding window model," Proceedings of the 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp. 1643–1647, 2017.
13. A. Sachdeva, G. Jethwani, C. Manjunath, et. al, "An Effective Time Series Analysis for Equity Market Prediction Using Deep Learning Model," Proceedings of the 2019 International Conference on Data Science and Communication (IconDSC), pp. 1–5, 2019.
14. Albert King，AKShare 2019, GitHub repository，https://github.com/akfamily/akshare