Tock Market Forecasting Based on Machine Learning Approach of ARIMA Model

Lingfeng Ren[1,†] Chenhao Zhao[1,†,*]

[1] *SWJTU-Leeds joint school, Southwest Jiaotong University, Chengdu, 611756, China*
[*] *Corresponding author. Email: sc19c2z@leeds.ac.uk*
[†] *These authors contributed equally*

ABSTRACT

As an important manifestation of national economic and financial activities, the stock market plays an important role in the economic development of various countries. If we can grasp the trend of the stock market in advance, it will be beneficial to both investment institutions and investors. By training a good ARIMA model, this paper applies this model to the stock prediction of five different listed companies: Alibaba, Baidu, Tencent, Pinduoduo, and WangYi. The predicted results are evaluated by MSE, MAE, RMSE, and MAPE. The results show that the accuracy of this model in the stock forecast of Alibaba, Baidu, Tencent, PingDuoDuo, and WangYi is between 96.0% and 99.4%, among which WY has the highest accuracy of 99.4% and PDD has the worst accuracy of 96.0%. The ARIMA model has high accuracy for stock prediction, but it needs to rely on the reliability of its AR and MA models, and it is highly dependent on the judgment of residual sequence. At the same time, because the stock market is also affected by other factors, a framework that may be accurate for all types of stock markets will be included in the future development plan, thus providing investors and related investment institutions with reference for stock investment decisions.

Keywords: *ARIMA model, Time Series, Stock forecast*

1. INTRODUCTION

As a crucial manifestation of national economic and financial activities, the stock market plays a role in the economic development of various countries and has an important impact on enterprises and individuals. If the trend of the stock market can be grasped in advance, it will be beneficial to both investment institutions and investors. However, due to the uncertainty and volatility of the stock market, if investors invest blindly, it will bring immeasurable losses. Therefore, the stock forecast is always a challenging and important task.

In the field of financial markets, as a kind of time-series data, stock data implies a series of operation rules. Through the analysis of these rules, we can make the corresponding trend forecast of stocks in the market and provide decision support for investors to invest in stocks. It is of great significance for investors and regional economic development. However, the fluctuation of the stock price depends on many factors, such as stock policy, panic, equity, interest rate, options, warrants, mergers, ownership of large financial companies, etc. Scientific and effective stock market forecasting methods can not only provide basic information for trading regulators to formulate policies to stabilize the financial market, but also be important parameters for investors to make profits and avoid risks.

In the previous time, the fundamental method of stock analysis was to analyze the trend of stock price to forecast. For a long time, most research has used combination forecasting and regression analysis to predict the trend of the stock and the possible price in the future, but the predicted values of these methods all had great errors. ARIMA model has been well developed in the analysis and application of time series and can accurately predict stock price data. ARIMA model can not only fit the stable time series but also has good fitting characteristics for the non-stationary time series.

Sima and Akbar think about the exactness of ARIMA and LSTM as agent methods for gauging time-series information. These two strategies are executed and applied to a bunch of monetary information, and the outcomes show that LSTM outflanks ARIMA. Even more explicitly, LSTM-based calculations further developed expectations by a normal of 85% contrasted with ARIMA. Moreover, the paper reports no improvement when the quantity of ages is changed. The

Y. Jiang et al. (Eds.): ICEDBC 2022, AEBMR 225, pp. 233–237, 2022.
https://doi.org/10.2991/978-94-6463-036-7_34

work portrayed in this paper advocates the advantages of applying profound learning-based calculations and procedures to monetary and monetary information. A few other gauging issues in money and financial matters can be formed utilizing profound learning. The creators intend to explore the enhancements accomplished through profound advancement by applying these procedures to a few different issues and datasets with various quantities of highlights [1]. Paulo et al. assess the exhibition of the model ARIMA in Ibo Vespa time series anticipating. The examination strategy utilized is numerical displaying, following the Box-Jenkins technique. To contrast the outcomes and other smoothing models, the assessment boundary MAPE (Mean Absolute Percentage Error) was utilized. The outcomes show that the model utilized accomplishes lower MAPE values, demonstrating more noteworthy pertinence. Thusly, this shows that the ARIMA model can be utilized for time-series records connected with financial exchange file estimating [2]. Zhou and Wu laid out the ARIMA model given recorded information and made a transient gauge of the everyday shutting cost of the Shanghai Composite Index. Through the forecast outcomes, it tends to be seen that the model impact is great, the distinction between the expected results and the genuine worth is little, and the model laid out in this paper has specific ideas for corporate share. In any case, the financial transaction mode is not only interfered by many complicated factors in essence, but also has the characteristics of time shift, irregularity, and nonlinearity. Therefore, the model presented in this paper is only a substantive description of instantaneous expectations, and its center is the general model focusing on specific stocks. Financial supporters can provide reference for speculation by judging the model and strengthening the investigation of key and special parts of the stock. At the same time, they should also consider the expectation of the medium and long-term model of the stock exchange. th[3].

Rachel and Gilbert utilized MLPNN and ARIMA models to estimate month-to-month homegrown antibody interest in the Philippines from January 2014 to December 2019. Then, utilize the RMSE and MAE precision measures to choose a fitting gauging strategy to find a proper model to foresee suitable antibody stocks to keep away from deficiencies and oversupply. The outcomes show that the MLPNN model beats the ARIMA model in anticipating month-to-month immunization needs. The expectations of this study can assist policymakers with settling on better conclusions about expanding the inoculation inclusion. In ongoing investigations, further analyses can be directed by applying this technique to a more extensive scale and utilizing other determining strategies, particularly blended models [4]. Ayodele et al. have a broad strategy for building an ARIMA model for stock cost anticipating. Distributed stock information got from the New York Stock Exchange (NYSE) and the Nigerian Stock Exchange (NSE) was utilized with the created stock cost gauging model. The trial results acquired with the best ARIMA model exhibit the capability of the ARIMA model to foresee palatable stock costs temporarily. This can direct financial backers in the securities exchange to settle on beneficial speculation choices. With the got results, the ARIMA model can contend sensibly well with arising gauging strategies in transient determining [5]. Hiransha and Gopalakrishnan utilize four DL models to foresee the stock costs of NSE and NYSE. Because of the stock cost of NSE Tata Motors, it prepared four organizations MLP, RNN, LSTM, and CNN, and utilized the acquired models to foresee from NSE financial exchange. It is presumed that direct models, for example, Arima are univariate time-series expectations. In the proposed work, CNN performs better compared to the next three organizations. At long last, it is reasoned that the DL model is superior to the ARIMA model [6].

Chairman applied LSTM to establish a robust prediction model and established the prediction model of Tehran stock price. The accuracy and error were compared with the ARIMA model. The results showed that LSTM is better than Arima in prediction. Although both models could work well in the short term, the prediction accuracy of both models would decrease with the increase in prediction time [7]. Gaurav and Sidnal discussed various technical indicators, challenges faced by stock market prediction, classification types, and ML algorithm analysis. The ML algorithm forms the basis of complex problems related to big data analysis such as stock market prediction. Its analysis can be used as EML, which is the correct combination of all prediction models, to better make correct decisions [8].

Shah et al. put forward the classification of calculation methods of stock market analysis and prediction, makes detailed literature research on the latest algorithms and methods commonly used in stock market prediction, and discuss some continuous challenges that need more attention in this field, which provides opportunities for future development and research. It is concluded that the market is like a weighing machine, with less noise and stronger predictability. The hybrid method combining statistics and machine learning technology may prove to be more useful for stock forecasting [9]. Moghar and Hamiche proposed an RNN model based on LSTM to predict the future value of Google and NKE assets. The results of the model show some promising results. The test results show that it can track the evolution of the opening price of these two assets [10].

1.1. Objective

This paper focuses on the ARIMA model to predict the stocks of different companies. Firstly, the accuracy of the ARIMA model is shown by inputting an equal proportion of the training set and test set. After that, this ARIMA model is applied to the stock price prediction of different companies, which shows the accuracy and error of this model in the prediction. To show whether the ARIMA model has high accuracy in stock prediction.

In the method part of this paper, the steps of the prediction process and the internal principle of the ARIMA model will be introduced. In the result part, the accuracy and error of the prediction results of five different listed companies' stocks by using the constructed ARIMA model will be displayed. In the evaluation part, the reasons for this result will be analyzed and evaluated. In the summary part, through the accuracy and error of this model in the real stock market, suggestions on how investors and companies invest in the stock market will be given.

2. METHOD

As an algorithm in machine learning, the ARIMA model can be used for a stock price forecast. The process of forecasting in this paper follows the following processes: data receiving, data preparation, model selection, model training, and forecasting. After the prediction is finished, the accuracy of the model prediction can be obtained by comparing the prediction results with the original data. After completing each step of the above process, based on different purposes and requirements, the target model can be trained into a model with excellent stock prediction ability.

In an ARIMA model (P, D, q), AR is "autoregressive", which is based on the Equation (1):

$$X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \epsilon_t \quad (1)$$

and P is the number of autoregressive terms. Ma is the moving average, q is the number of moving average terms, and d is the differences (order) that make it a stationary sequence. Although the word "different" does not appear in ARIMA's name, it is a crucial step. Research mainly analyzes stock data according to the Equation (2).

$$\left(1 - \sum_{i=1}^{p} \emptyset_i L^i\right)(1 - \mathrm{L})^{\mathrm{d}} \mathrm{X_t} = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right)\epsilon_{\mathrm{t}} \quad (2)$$

The acquisition of time series was obtained through experimental analysis. Pre-processing of time series the pre-processing of time series includes two aspects of testing, namely, stationary testing and white noise testing. Checking the stationarity of data is one of the important steps in time series analysis. Generally, the stationary of time series was tested through a time series diagram and correlation diagram. Model identification is to choose a model that was consistent with the given time series process from known models.

Determination of model order After determining the type of model, the model order still needs to be selected. Then, the parameters of the model were estimated to obtain more accurate data. Model validation was mainly to validate the fitting effect of the model. After preparing the model and data, by analyzing the stock data and subdivision data of Alibaba and Huawei, we train the model with the training data set, predict with the model, and compared the predicted results with the test data set. The prediction accuracy of the corresponding model is obtained by analyzing the curve fitting degree and the deviation between nodes. In the experiment, the ARIMA model is compared with linear regression and learning models. By comparing the coincidence degree of graph curves and the deviation degree of final nodes, the advantages and disadvantages of the model are analyzed.

3. RESULTS

After the training of the data set, the ARIMA model predicts the stock of the next period of the time series, couples the prediction results with the test set, obtains the final accuracy, and evaluates the results with four evaluation indexes: MSE, MAE, RMSE and MAPE The accuracy of this model in the stock prediction of Alibaba, Baidu, Tencent, PDD, and WY is between 96.0% and 99.4%, of which WY is the most accurate, with an accuracy of 99.4%, and PDD is the worst, with an accuracy of 96.0%.

In the table, MSE stands for Mean square error, RMSE stands for Root mean square error is the addition of a root sign to MSE, which is intuitive in order of magnitude, MAE stands for Mean absolute error and MAPE stands for Mean absolute percentage error.

Table 1. The calculation results of MSE, MAE, RMSE, and MAPE

	Alibaba	Baidu	Tencent	PDD	WY
MSE	0.02263898	0.01239289	0.01091623	0.04063916	0.00059732
MAE	0.13144390	0.06991626	0.08359703	0.13792756	0.02044451
RMSE	0.15046256	0.11132335	0.10448077	0.20159157	0.02444020
MAPE	0.02830114	0.01445468	0.02190189	0.03900679	0.00560615

4. EVALUATION

The trained model needs to be tested to prove whether it can work properly in the real world. Therefore, part of the dataset created for evaluation is used to check the proficiency of the model. This will put the model in a scenario, in which case it will encounter some situations that are not included in the training content. In this test, this may mean trying to identify completely unfamiliar model types. However, through training, the model should have sufficient information reasoning ability and the ability to predict future stocks.

By inputting an equal proportion of training set and test set, the ARIMA model shows a certain degree of accuracy This result shows that AR and ma have good data structure when the previous model is established, and the data set may have been fully optimized to eliminate unnecessary noise so that the model may have enough matching properties to couple and predict the data

However, the results still show that the ARIMA model has different degrees of accuracy for the prediction of different types of data sets, which may be related to the optimization of the model and data when establishing the model, which makes the model show different coupling degrees when predicting, and finally leads to different accuracy.

5. CONCLUSION

The economic systems of some developing countries are unstable, making it difficult to establish robust forecasting models. With the help of new approaches to RNN, the possibility to obtain better estimates emerges (without differing and linearizing ignoring the nature of the data). In this paper, the ARIMA model is applied to this objective and a predictive model for stock prices is built. And collected different data sets for experiments and compared the accuracy and error. The results show that the ARIMA model has a very high accuracy for stock forecasting, but it needs to rely on the reliability of its AR and MA models and has a high dependence on the judgment of the residual sequence.

Because the stock market is influenced by other factors, such as government and consumer decisions. In the future, the ARIMA model will be worked on to develop a framework that is likely to be accurate for all types of stock markets.

REFERENCES

[1] S.N. Sima, S.N. Akbar, FORECASTING ECONOMIC AND FINANCIAL TIME SERIES: ARIMA VS. LSTM, Texas Tech University, 2018.

[2] R.J. Paulo, L.R.S Fernando, ARIMA: An Applied Time Series Forecasting Model for the Bovespa Stock Index, Applied Mathematics, vol. 5, 2014, pp. 3383-3391.

[3] X.L. Zhou, Z.X. Wu, Short Term Forecast of Daily Closing Price of Shanghai Composite Index—Empirical Analysis Based on ARIMA Model, Advances in Social Sciences, vol.5, no.4, 2016, pp. 630-637.

[4] R.T. Alegado, G.M. Tumibay, Statistical and Machine Learning Methods for Vaccine Demand Forecasting: A Comparative Analysis, Journal of Computer and Communications, vol.8, 2020, pp. 37-49.

[5] A.A. Adebiyi, A.O. Adewumi, C.K Ayo, Stock Price Prediction Using the ARIMA Model, International Conference on Computer Modelling and Simulation, 2014, pp. 106-112.

[6] M. Hiransha, E.A. Gopalakrishnan, V.K. Menon, K.P. Soman, NSE Stock Market Prediction Using Deep-Learning Models, Procedia Computer Science, vol. 132, no. 2018, pp. 1351-1362.

[7] G. Abdoli, Comparing the Prediction Accuracy of LSTM and ARIMA Models for Time-Series with Permanent Fluctuation, vol. 9, no.2, 2020.

[8] U. Gurav, N. Sidnal, Predict Stock Market Behavior: Role of Machine Learning Algorithms. In Intelligent Computing and Information and Communication. Advances in Intelligent Systems and Computing, vol 673, 2018.

[9] D. Shah, H. Isah, F. Zulkernine, Stock Market Analysis: A Review and Taxonomy of Prediction Techniques. International Journal of Financial Studies, vol. 26, no.7(2), 2019.

[10] M. Adil, H. Mohamed, Stock Market Prediction Using LSTM Recurrent Neural Network, Procedia Computer Science, vol 170, 2020, pp 1168-1173.