Proceedings of the 2023 International Conference on Management Research and Economic Development
DOI: 10.54254/2754-1169/20/20230224

Stock Price Prediction Based on OLS and LSTM: Evidence from JPMorgan Chase, Tesla and Apple

Ronglei Xiao[1,a,*]

[1]Wuhan University, Wuhan, 430072, China
a. 2020302011112@whu.edu.cn
**corresponding author*

Abstract: Contemporarily, the stock price prediction remains a challenging topic that puzzles many economists and researchers. With the development of science and technology, many types of methods have been utilized to predict the stock price. The paper first reviews the popular stock price prediction methods in recent years and then uses the OLS and LSTM models and the stock market data including 12 Technical indicators from 2018 to 2022 to predict the stock prices of the three major companies: JPMorgan Chase, Tesla and Apple. According to the analysis, the regression and the test results show that LSTM has better prediction ability. On this basis, this study gives suggestions for three companies to predict the stock price in the future. Finally, the shortcomings of the methods used in the article are proposed, and the future prospects are also discussed. These results offer a deeper and clearer understanding of the role and effect of OLS and LSTM models in stock price forecasting.

Keywords: stock price prediction, Ordinary Linear Regression, LSTM

1. Introduction

Stock price prediction is an attractive yet challenging topic. The prediction can help governments and intergovernmental organizations supervise the financial market and make timely and necessary adjustment. Meanwhile, the entrepreneurs and independent investors intend to get the best portfolio strategy by predicting the trend of stocks. However, the unstable environment, the complex market mechanism and the competition for interests make it difficult to make reliable and accurate prediction. According to the well-known efficient market hypothesis (EMH), Fama holds that the market if efficient so technical analysis or fundamental analysis would not generate any persistent above-average return to investors [1]. However, many economists and researches disagreed with this view and they devote half of their lives to finding the proper method and the rule of it. With their long-term efforts, scholars have more advanced and reliable method to make stock prediction now.

Previously, researchers have mainly used the following five techniques to predict stock price: Statistical, Pattern Recognition, Machine Learning and Sentiment Analysis and Hybrid [2]. Time series analysis and regression is the foundation of stock price prediction. Therefore, people can find these statistical techniques in most relevant researches [3]. However, in most cases, scholars will combine them with other methods to get more reliable predictions [4]. Nowadays, the existing technology enables scientists to do more sophisticated algorithms. They tend to use machine learning and pattern recognition to predict trend in highly volatile and noisy environments [5]. On one hand, the accuracy and utility of the prediction have been improved a lot by these techniques. On the other

DOI: 10.54254/2754-1169/20/20230224

hand, the selection of proper models and the ambiguity of frequency of retraining the model remain to be questions [6]. Meanwhile, investors' emotion and expectation are one of the sources of the fluctuation of the market. So, some researchers just took these factors into consideration when predicting stock price [7]. Nevertheless, it is hard to quantify the human beings' feeling or reaction so the relevant researches need much data sup-port. The integration of multiple methods has become a major trend in stock price prediction [8, 9]. How to combine multiple methods reasonably and scientifically and give full play to the advantages of different methods in dealing with different situations will become the focus of researches in the next stage.

The purpose of this paper is to evaluate the performance of Ordinary Linear Regression (OLS) and Long Short Term-Memory (LSTM) on stock price prediction. The former model is more basic while the latter one combine Pattern Recognition and Machine Learning. The paper will compare their performance and help investors to find out whether the more advanced and complicated model can gain ad-vantages in different situations. In order to make more robust conclusions, the paper Select three typical stocks in different industries in the US stock market: JPMorgan Chase (JPM), Tesla (TSLA) and Apple (AAPL). JPM ranks first among American banks in terms of asset size [10]. TSLA is a pioneer in the electric vehicle industry. Its boss, Elon Musk, has become the focus of the world because of SpaceX and the acquisition of Twitter. AAPL is a very innovative and popular company, and its global market share reached 19% in 2022, ranking second in the world [11].

The reminder of the study is organized as follows. In part Two, the Source and selection of data is mentioned. Meanwhile, it intro-duces the basic principle of OLS and LSTM, and lists the indicators to evaluate the regression results. Part Three establishes two models in python, then com-pares the different results of the three companies, and finally gives some suggestions for future price prediction. In part Four the limitations of this study are put forward and the prospects for future research are given. In part Five, a brief summary for full content is given.

2. Data & Method

The paper utilized three stock price datasets from Yahoo Finance, which respectively contain the basic stock information about JPM, TSLA and AAPL. These datasets were collected from 2018 Jan. to 2022 Dec. in daily option. Meanwhile, the datasets consist of 1507 stock prices as train set and 302 stock prices in test set when making LSTM analysis.

OLS is the most fundamental form of regression analysis, which requires the least model conditions. The model's goal is to minimize the sum of squares of the distances from all observations on the scatter plot to the regression line (MSE). When making stock price predictions, the price is often chosen as the dependent variable Y, and the volume price combination and technical indicators as independent variables X. The model involves partial derivation and matrix operation when solving problems.

LSTM is a variant of Recurrent Neural Network (RNN) and its core concepts are cell state and "gate" structure. Compared with traditional neural network, RNN can ensure the continuous existence of information by continuously cycling information. Standard RNN has certain memory function and can be used to solve many problems, such as speech recognition, machine translation. However, the limitation of standard RNN is very obvious. When the prediction point is far away from the dependent information, standard RNN cannot memorize the information. LSTM are designed to solve long dependencies problem.

The unit in the standard RNN network has only one network layer. On the contrary, there is four network layers in the unit of the LSTM network. One can view these units as cells. The cell state runs through the entire cell with only a few branches, which ensures that information remain unchanged after flowing through the entire cell. The cell contains three gate structures: forget gate, input gate and forget gate. A typical gate structure in LSTM consists of a sigmoid layer and a pointwise

DOI: 10.54254/2754-1169/20/20230224

multiplication operation. These gates determine whether information is retained or forgotten. With the help of these gate structures, even information from earlier time steps can be transferred to subsequent cells. In this paper, LSTM is implemented through Tensorflow 2.0 in python. The whole network is composed of the LSTM layer and the Dense layer. The input dimension of the LSTM layer is 1, i.e., the price of the stock and its output dimension is 50. The Epoch and the Batch Size of LSTM network training is respectively 200 and 50. In the training, MAE is the objective function and the optimizer is Adam. Others are all system default parameters.

The paper uses the following four metrics to evaluate the results of regression and training: MAE, MAPE, MSE, R-Square with the expression as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{1}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|\hat{y}_i - y_i|}{y_i} \times 100\% \tag{2}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{3}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(\bar{y}_i - y_i)^2} \tag{4}$$

For MAE, the range is from 0 to $+\infty$. The smaller MAE indicates that the prediction model has better accuracy. For MAPE, the range is from 0 to $+\infty$. The smaller MAPE indicates that the prediction model has better accuracy. Regarding to MSE, the range is from 0 to $+\infty$. The smaller MSE indicates that the prediction model has better accuracy. As for R-square, the range is from 0 to 1. The bigger R-square indicates that the prediction model has better accuracy.

3. Results & Discussion

3.1. Price Action

The stock price of JPM experienced two big declines in the spring of 2020 and 2022, which were most likely caused by the COVID-19 pandemic and the Russia-Ukraine war. The relative low prices lasted for about six or eight months (seen from Fig. 1). The stock price of TSLA generally showed an upward trend from 2018 to 2021, and only suffered a decline in the spring of 2021. However, after entering 2022, the price fluctuation increased and the overall trend was downward. The stock price of AAPL also showed an upward trend from 2018 to 2021. In 2022, the price experienced two major fluctuations.

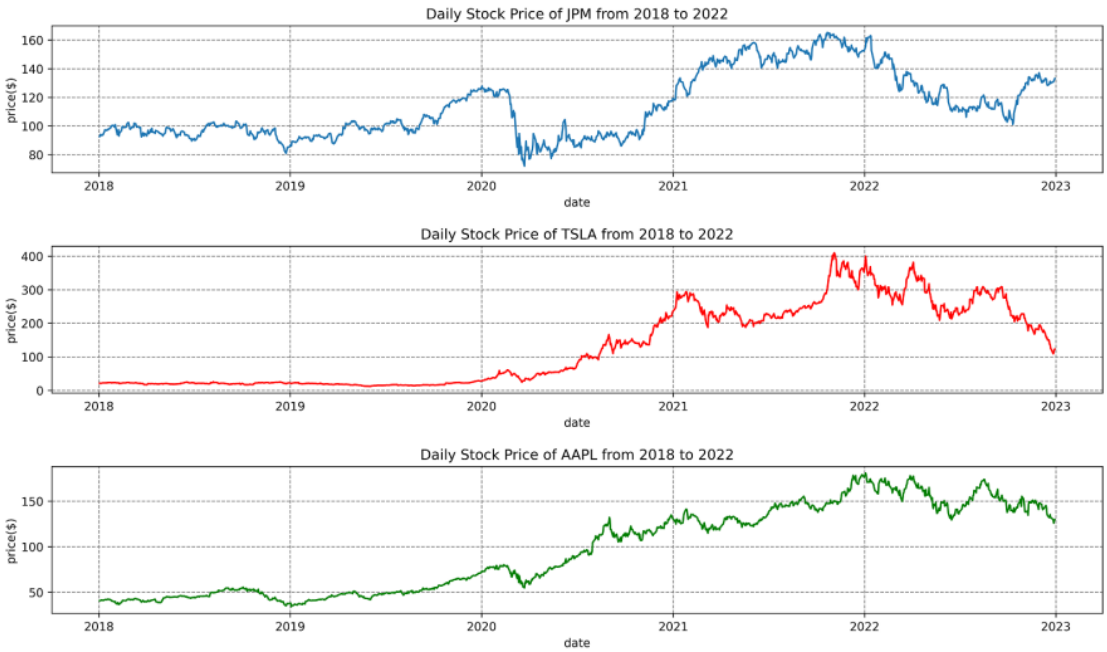

Figure 1: Daily stock price from 2018 to 2022.

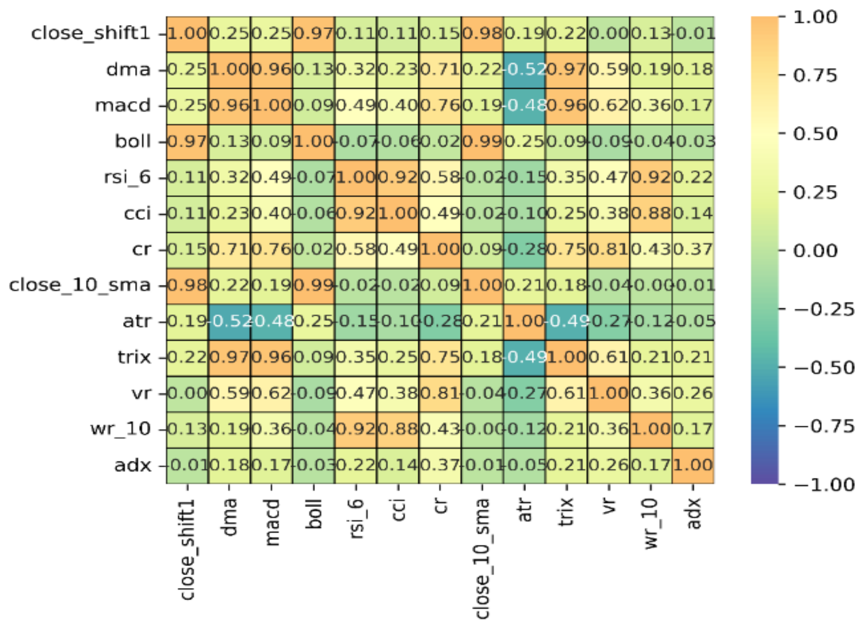

Figure 2: The pearson correlation coefficient between the indicators and stock prices of JPM.

3.2. OLS

This paper uses OLS model to make linear regression of each company's stock price and multiple technical indicators. Firstly, 12 technical indicators are calculated by using StockStats module. They are DMA, MACD, BOLL, RSI (6 DAYS), CCI, CR, SMA (10 DAYS), ATR, TRIX, VR, WR (10 DAYS), ADX. Secondly, the Pearson Correlation Coefficient between each indicator and the stock price is worked out and ranked according to the value. Then, the paper selects the three indicators with the highest correlation coefficient, which must be greater than 0.8. Finally, linear regressions are made and the results are shown. As shown in Fig. 2, the stock price of JPM has a great positive

DOI: 10.54254/2754-1169/20/20230224

correlation with BOLL and SMA (10 DAYS). Therefore, these two indicators are selected to make linear regression on the price. The fitted curve is Price=-0.528×BOLL+1.522×SMA (10 DAYS) +0.841. Other results of the regression are shown in Fig. 3 and Table. 1. According to the results, Linear regression has achieved excellent results in this case.

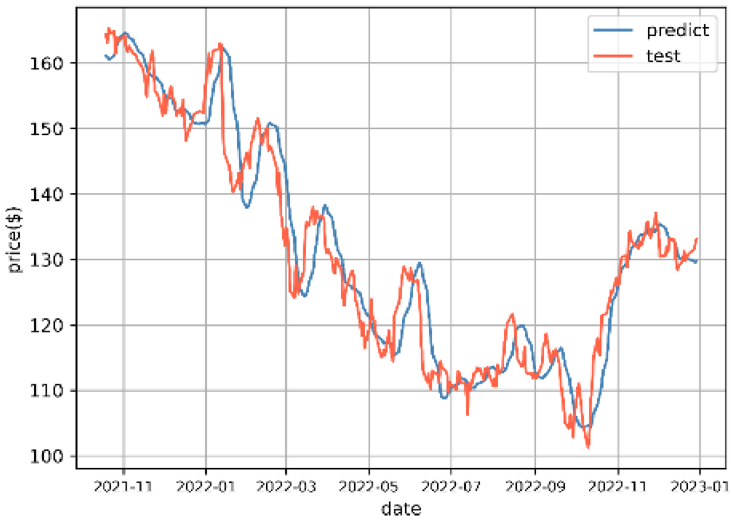

Figure 3: The linear regression effect on the JPM's stock price.

Table 1: The metrics of the linear regressions.

	MAE	MSE	MAPE	R-square
JPM	3.695	24.483	2.878	0.918
TSLA	18.143	505.239	6.582	0.874
AAPL	4.725	33.593	3.068	0.795

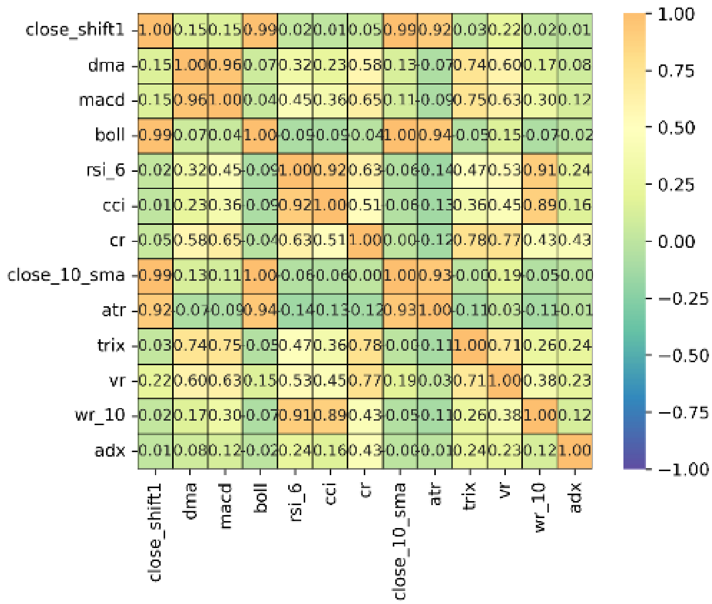

Figure 4: The pearson correlation coefficient between the indicators and stock prices of TSLA.

The stock price of TSLA has a great positive correlation with BOLL, SMA (10 DAYS) and ATR (as depicted in Fig. 4). Therefore, these three indicators are chosen to make linear regression on the price. The fitted curve is Price=-0.603×BOLL+1.601×SMA (10 DAYS) +0.104×ATR+0.398. Other results of the regression are shown in the following Fig. 5 and Table. 1. The result of linear regression in this case is not as good as that of JPM: the regression has large MAE, MSE, MAPE though the R-square seems good.

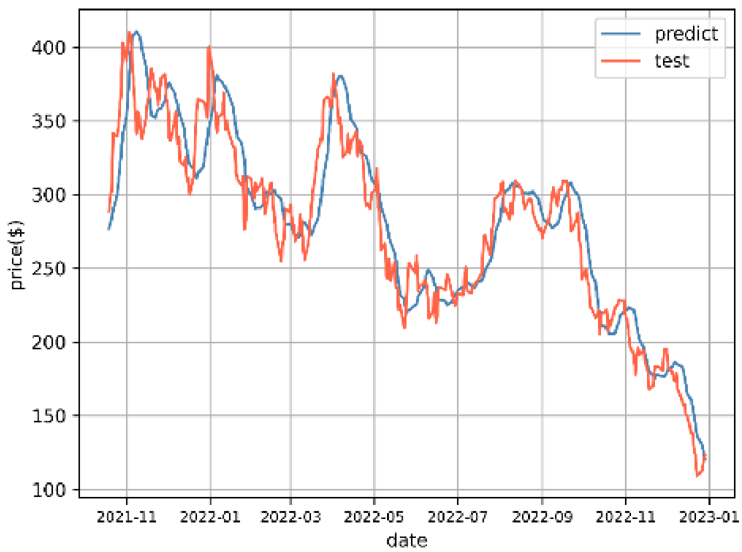

Figure 5: The linear regression effect on the TSLA's stock price.

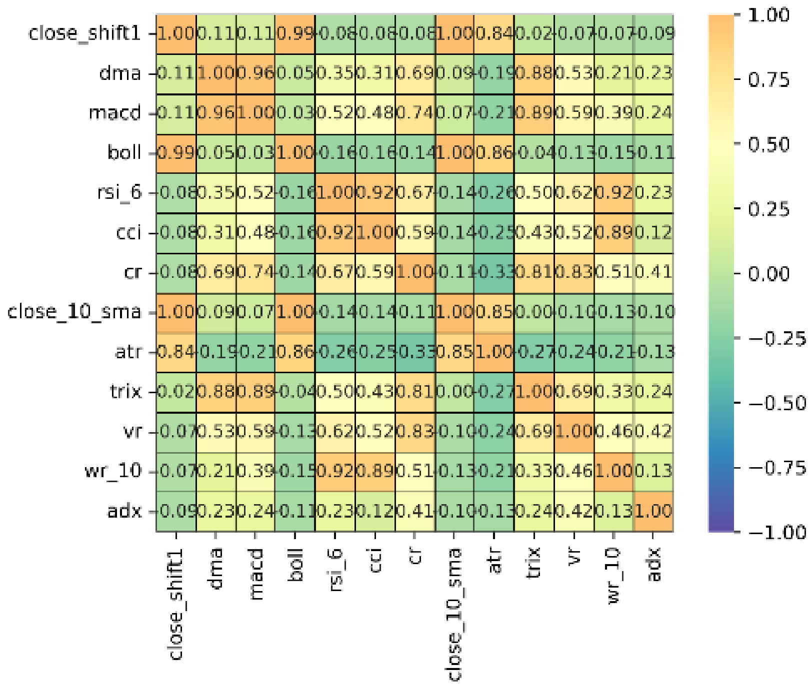

Figure 6: The pearson correlation coefficient between the indicators and stock prices of AAPL.

DOI: 10.54254/2754-1169/20/20230224

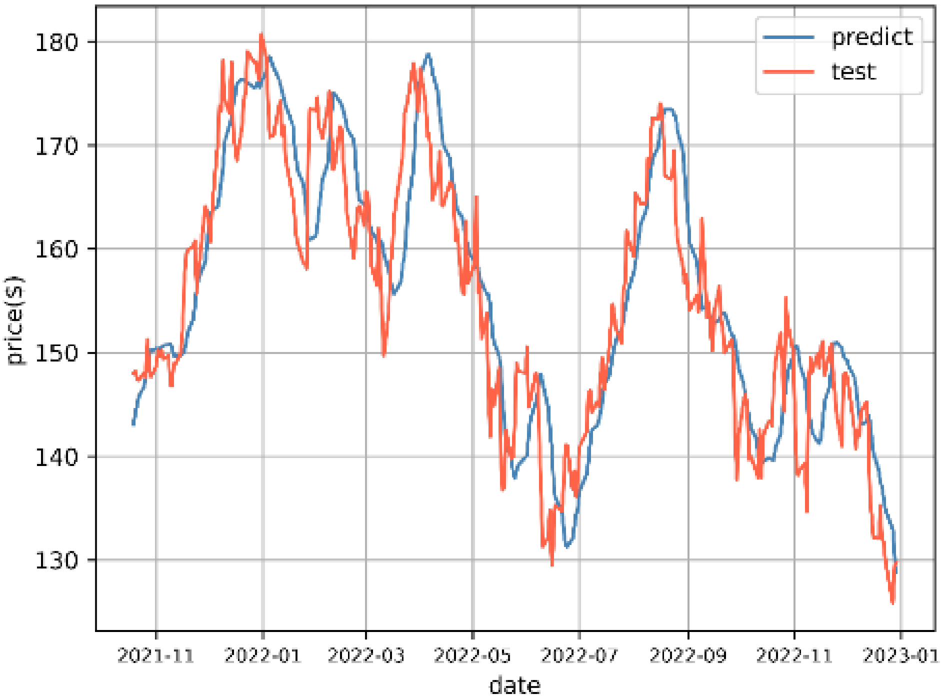

Figure 7: The linear regression effect on the AAPL's stock price.

The stock price of AAPL has a great positive correlation with BOLL, SMA (10 DAYS) and ATR (seen in Fig. 6). Therefore, these three indicators are chosen to make linear regression on the price. The fitted curve is Price=-0.517×BOLL+1.524×SMA (10 DAYS) -0.212×ATR+0.198 Other results of the regression are shown in the Fig. 7 and Table 1. Linear regression in this case has a relatively small MAE, MSE, MAPE, but a poor score of R-square.

In general, when using technical indicators to make linear regression of price, only the regression of JPM has achieved an ideal result.

Figure 8: The result of using LSTM to predict the price of JPM.

DOI: 10.54254/2754-1169/20/20230224

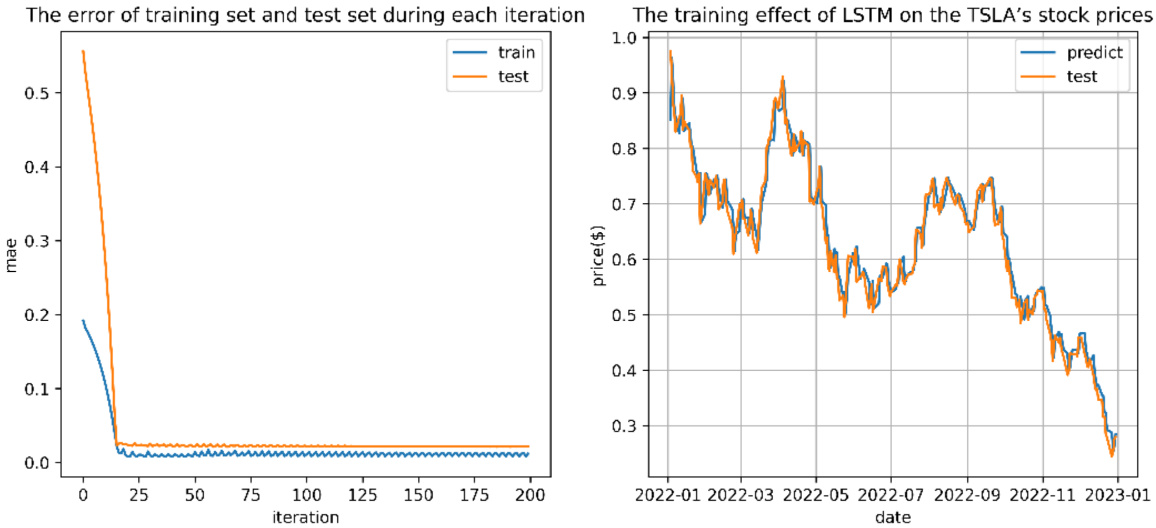

Figure 9: The result of using LSTM to predict the price of TSLA.

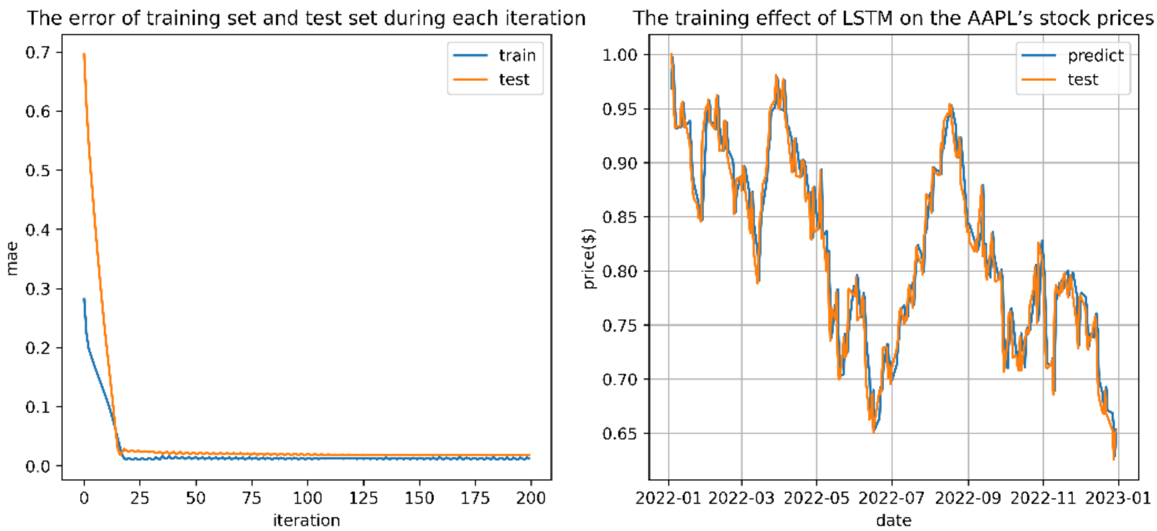

Figure 10: The result of using LSTM to predict the price of AAPL.

3.3. LSTM

After training the stock price of JPM by using LSTM model, its test set has achieved excellent results as illustrated in Fig. 8 and Table 2. After training the stock price of TSLA based on LSTM model, its test set has also achieved excellent results (seen from Fig. 9 and Table 2). After training the stock price of AAPL in terms of LSTM model, it has achieved good results, though the results are slightly inferior to the first two stocks, (as demonstrated in Fig. 10 and Table. 2). Overall, the LSTM network has perfect training effect with the price of these three stocks.

Table 2: The metrics of the LSTMs.

	MAE	MSE	MAPE	R-square
JPM	0.020	0.001	29.689	0.968
TSLA	0.021	0.001	29.036	0.962
AAPL	0.018	0.001	12.384	0.929

DOI: 10.54254/2754-1169/20/20230224

3.4. Comparison & Suggestion

The prediction of the price of three stocks from 2018 to 2022 has achieved good results when using LSTM model to make analysis. When using OLS to forecast, except for the good effect of predicting the price of JPM, the prediction of TSLA and AAPL is not good. Therefore, when predicting stock prices in the future, JPM can use OLS or LSTM, while for TSLA and AAPL, only LSTM is recommended.

4. Limitations & Prospect

This article's research methodology has certain drawbacks. Some limitations are caused by the model itself, and some subjective factors can be improved in the follow-up study. For OLS, although the principle of multiple-choice linear regression is simple, the calculation is fast. Besides, it is easy to add new associated variables, the stock market data is not linear, so some hidden factors in the market may be erased because of linearity and the model's fitting effect will not be well. Meanwhile, this paper only selects 12 technical indicators from the stock market, and does not consider other technical indicators such as MTM or conformance of volume and price. The ensuing study on utilizing linear models to predict stock prices can include research on these aspects. For LSTM, the paper only uses the stock price as the input of the network. In the follow-up researches, more relevant variables should be added for training. Additionally, even though LSTM networks have addressed the issue of long dependencies, they require a lot of computation time and resources when the time span of data is broad.

The research discussed in this paper is limited to historical data; neither the causes of data variations nor non-quantitative aspects are examined. The reason why stock prices are difficult to predict is that an emergency event may change all the rules obtained in the past and the researches cannot take all the factors into consideration. Therefore, future researches on stock price prediction should focus more on hybrid method, because it means that people can use more laws that they have mastered to make prediction. Hence, there will be a greater chance of winning. Moreover, one needs to improve the sensitivity of the model so that the model can quickly catch the arrival of the turning point and make corresponding analysis.

5. Conclusions

In summary, this study employs OLS and LSTM to predict the stock prices of JPM, TSLA and AAPL from 2018 to 2022. When using OLS to make regressions, only JPM's price is predicted well and there are some issues with the other two firms' price predictions. On the contrary, the outcomes are great when LSTM is used to predict the prices of the three companies. In general, LSTM achieves better results in prediction than OLS. However, this paper only investigates the past stock market data, and ignore the non-quantitative factors, which also have an important impact on the stock price. Meanwhile, there is still room for further optimization of data processing, e.g., increasing the selected technical indicators and the input dimension of LSTM. This study describes the basic principles of OLS and LSTM, and uses python to verify the process. According to the prediction of the stock prices of three leading companies in different industries, it compares and evaluates the effects of the two models on stock price prediction. These results help understand the role and effect of OLS and LSTM in stock price prediction more clearly and intuitively.

References

[1] *Jiang, W.: Applications of deep learning in stock market prediction: recent progress. Expert Systems with Applications, 184, 115537 (2021).*

[2] *Shah, D., Isah, H., Zulkernine, F.: Stock market analysis: A review and taxonomy of prediction techniques. International Journal of Financial Studies, 7(2), 26 (2019).*

[3] *Alotaibi, R.: ARIMA Model for Stock Market Prediction. In Proceedings of the 2022 8th International Conference on Computer Technology Applications (pp. 1-4) (2022, May).*

[4] *Prasad, V. V., Gumparthi, S., Venkataramana, L. Y., Srinethe, S., Sruthi Sree, R. M., Nishanthi, K.: Prediction of stock prices using statistical and machine learning models: a comparative analysis. The Computer Journal, 65(5), 1338-1351 (2022).*

[5] *Joshi, A., Deshmukh, P. K., Lohokare, J.: Comparative analysis of Vanilla LSTM and Peephole LSTM for stock market price prediction. In 2022 International Conference on Computing, Communication, Security and Intelligent Systems (IC3SIS) (pp. 1-6). IEEE (2022, June).*

[6] *Patel, R., Choudhary, V., Saxena, D., Singh, A. K.: Review of stock prediction using machine learning techniques. In 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI) (pp. 840-846). IEEE (2021, June).*

[7] *Bhavsar, H., Jivani, A., Amesara, S., Shah, S., Gindani, P., Patel, S.: Stock Price Prediction Using Sentiment Analysis on News Headlines. In ICT with Intelligent Applications: Proceedings of ICTIS 2022, Volume 1 (pp. 25-34). Singapore: Springer Nature Singapore (2022).*

[8] *Trawinski, A., Wimmer, H., Oliver, D.: Sentiment Based LSTM for Stock Price Prediction: Congress vs General Public. In 2022 International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT) (pp. 885-893). IEEE (2022, October).*

[9] *Bose, A., Hsu, C. H., Roy, S. S., Lee, K. C., Mohammadi-Ivatloo, B., Abimannan, S.: Forecasting stock price by hybrid model of cascading multivariate adaptive regression splines and deep neural network. Computers and Electrical Engineering, 95, 107405 (2021).*

[10] *BankRate, https://www.bankrate.com/banking/biggest-banks-in-america/, last accessed 2023/2/16.*

[11] *Statcounter, https://gs.statcounter.com/vendor-market-share/mobile/worldwide/#yearly-2022-2022-bar, last accessed 2023/2/16.*