Proceedings of the 2nd International Conference on Financial Technology and Business Analysis
DOI: 10.54254/2754-1169/46/20230353

Amazon Stock Price Prediction During COVID-19 Based on the LSTM Model and Linear Regression Model

Jing Tan[1,a,*]

[1]*Institute of Problem Solving Korean(Business),Beijing Foreign Study University, Zizhuyuan Street, Beijing, China*
a.20080026@bfsu.edu.cn
**corresponding author*

Abstract: With COVID-19 sweeping the world in 2019 and the ensuing turmoil in the stock market, the need for accurate stock forecasting becomes particularly crucial during exceptional periods. Utilizing Amazon stock price data from 2002–2023, this study uses a machine learning approach to create a stock price prediction model in order to more accurately anticipate the trend of the stock price movement during COVID-19. The linear regression model and LSTM model are two models commonly used in forecasting studies, and this paper uses both models to forecast stock prices separately and evaluates the models by RMSE. The linear regression model is constructed directly with all data as independent variables. According to a 9:1 ratio, the LSTM model is segmented into a training set and a test set in this paper, with an overall structure of three layers and an optimizer of Adam optimizer. The test results reveal that the LSTM model's error is significantly lower than the linear regression model's error, proving that the LSTM model has a superior forecasting capacity to that of the linear regression model. As a result, it can be widely applied to the task of predicting stock prices for COVID-19 and offering decision-making advice to the appropriate decision-makers.

Keywords: Covid-19, stock price prediction, machine learning methods, LSTM, linear regression

1. Introduction

A global fast rise in COVID-19 infections, a respiratory condition brought on by the SARS-CoV-2 coronavirus, was observed in later 2019 [1]. This sudden global outbreak, which continues to this day, has not only posed a huge challenge to the public health, research and medical communities with which it is directly linked [2], but has also dealt a severe blow to global markets, slowing economic activity and plunging financial markets into panic [3, 4]. Among them, the e-commerce industry has received the most serious impact. The vast majority of e-commerce trade depends on the supply chain, but the supply chain has been paralyzed by countries' policies such as mandatory quarantine and working from home to stop the widespread spread of COVID-19. Major corporations, as market participants, were even more directly affected by COVID-19, with consumption and economic output declining sharply [5], corporate stock prices generally falling, showing many characteristics such as high volatility and unpredictability. The global stock market has become turmoiled these years. The stock price of Amazon was affected by COVID-19 during the outbreak and its changes were

DOI: 10.54254/2754-1169/46/20230353

representative. As a globally renowned multinational company, predicting the growth trend of Amazon's stock and enabling relevant decision makers to react to it in a timely manner is conducive to maintaining the stability of financial markets. Therefore, it becomes particularly important to forecast stocks during the special period of the New Crown epidemic. A more accurate forecasting model would facilitate decision making by governments, companies and investors, contributing to maintain stability in financial markets.

In the past two decades, Artificial Intelligence (AI) has developed rapidly in solving a number of tasks such as knowledge representation, reasoning, language understanding, translation, theorem proving, and associative memory [6]. In addition, classical and modern AI techniques are increasingly being used in a wide range of applications in finance. AI is fundamentally transforming the value and meaning of finance and driving smart fintech innovation at the application level in terms of mechanisms, products, services and systems [7]. For example, in 2008, Nakamoto et al. first proposed the application of blockchain technology to protect cryptocurrencies traded online [8]; in November 2022, as a generative AI software developed and released by OpenAI, ChatGPT has been proven to perform tasks such as stock price prediction, investment management and asset management [9]. As AI continues to evolve in the financial sector, deep learning techniques are also being widely applied to many predictive problems. For instance, decision tree models [10] and Bayesian vector autoregression [11, 12] were once solved using traditional learning methods.

Stock forecasting is one of the major research areas of forecasting work in finance. Forecasting of stock time series is heavily influenced by an elevated degree of uncertainty and a commonly acknowledged insufficient type of market efficiency [13], leading to the complexity of stock changes. Therefore, the accuracy of the forecasting results has become an important pursuit for researchers. Forecasting models have been iterated many times. Linear algorithms like the Autoregressive Integrated Moving Average (ARIMA) model [14], non-linear algorithms such as the Generalized Auto Regressive Conditional Heteroskedasticity (CARCH) model [15], and deep learning models such as Recurrent Neural Natworks (RNN) model and Long-Short Term Memory (LSTM) model [16] are all commonly used to perform stock forecasting. However, the researchers chose stock prediction periods with generally flat stock change trends. After the COVID-19 outbreak, stocks entered a period of significant volatility, making the accuracy of the model need further study. In order to determine whether stock forecasting models can still have more accurate predictions, this study will use stock prices before and after the outbreak as the study sample, and use the linear change model and the deep learning model LSTM to predict stock trends respectively. This study will compare their performance to determine the better-fitting stock forecasting model and provide advice to investors, and according to the prediction results of the model, propose suggestions for decision makers.

To ascertain that this finding is still valid during COVID-19, this study downloaded AMZN's 2002-2023 stock price data from the Yahoo platform, train the model using linear regression model and LSTM with the first 18 years of the dataset respectively, and test the accuracy of the prediction results with the remaining data. The RMSE for these models were 28.1825 and 0.0364 separately, demonstrating that the LSTM was a better predictor than the linear regression model.

2. Methods

2.1. Dataset Preparation

The source of the dataset for daily stock price AMZN comes from Yahoo Finance comprising Microsoft stock price information [17]. It contains 5,377 entries from January 2002 to May 2023. To improve the precision of forecasting the daily changes in the stock, four indicators “open price”, “highest price”, “lowest price” and “close price” are selected for model training and testing. The specific data changes in terms of the open price and close price are shown in Figure 1 and Figure 2.

DOI: 10.54254/2754-1169/46/20230353

This provides a solid basis for this paper to predict stock price changes during the new crown epidemic.

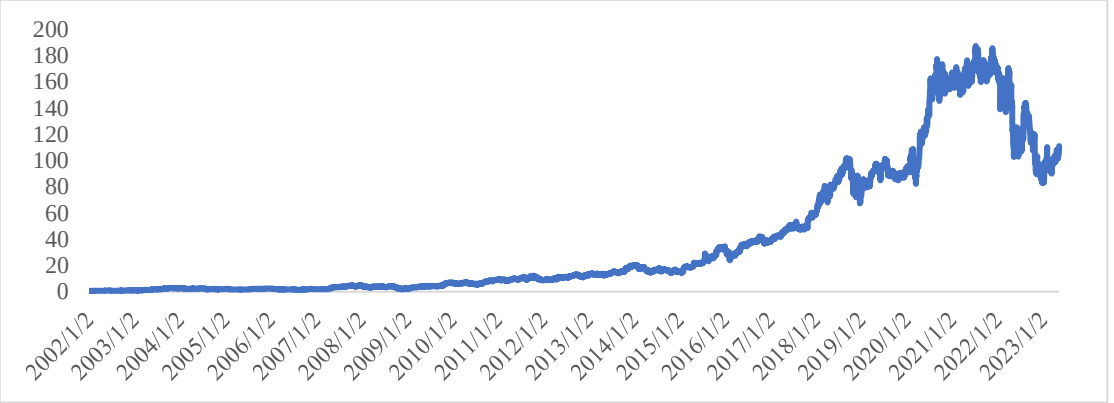

Figure 1: AMZN stock opening price curve (Photo/Picture credit: Original).

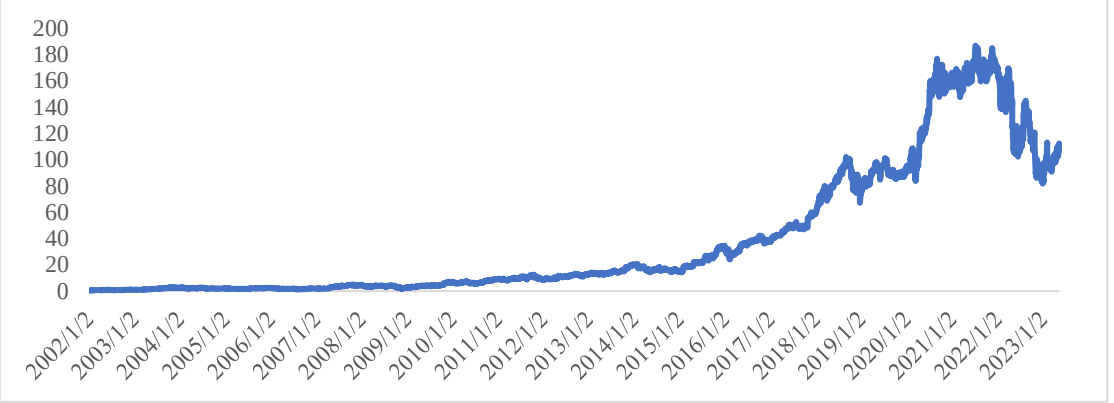

Figure 2: AMZN stock closing price curve (Photo/Picture credit: Original).

The data preprocessing is subsequently implemented. Firstly, a thorough examination was conducted to identify any instances of missing values or erroneous entries within the dataset. The training set and test set are then split into two groups in a 9:1 ratio. Finally, data normalization is performed on the data to reduce the training error as shown in equation (1) and equation (2).

$$X_{std} = \frac{X - X.\min(axis=0)}{X.\max(axis=0) - X.\min(axis=0)} \quad (1)$$

$$X_{scaled} = X_{std} * (max - min) + min \quad (2)$$

2.2. Machine Learning Models

This paper wishes to investigate the suitability of linear and non-linear models for accurately capturing stock price dynamics. To achieve this, both linear regression models and LSTM models have been selected as the forecasting techniques to be employed in this study. They will be utilized for comparing the performance of these two distinct modeling approaches in terms of their ability to effectively predict stock prices.

DOI: 10.54254/2754-1169/46/20230353

2.2.1. Introduction of Linear Regression

In this paper, a one-dimensional linear regression model was chosen and the opening and closing prices, highs and lows of AMZN stock prices were separately regressed using EXCEL. Assuming that the number of days is x and the daily stock price is y (where the opening stock price is y_1, the closing stock price is y_2, the high stock price is y_3 and the low stock price is y_4), let the one-dimensional equation be:

$$y = \beta_1 x - \beta_0 \tag{3}$$

2.2.2. Introdiction of LSTM

LSTMs are a type of recurrent neural network (RNN). Three gates control an LSTM's memory cell. They are the input gate, the forget gate, and the output gate.LSTMs are deep learning neural network algorithms that can handle temporal data and are widely used in different tasks. They are able to learn complex patterns in sequential data because they can selectively remember or forget information from previous inputs and states.

The LSTM structure proposed in this paper has 3 layers, the first two are LSTM layers, each possessing 128 neurons, and the output layer is the final one, a Fully connected layer (FCL), with 4 number of neurons, representing 4 different prices of the output.

2.3. Implementation Details

The optimizer used in this paper is Adam optimizer, learning rate is 0.01, epochs equal to 100, and the loss function is portrayed by the mean square error (MSE).

3. Results and Discussion

In the linear regression model prediction, based on the regression analysis's findings, the respective regression equations were:

$$\text{Opening price linear regression equation:} y_1 = 0.027253x - 34.8325 \tag{4}$$

$$\text{Closing price linear regression equation:} y_2 = 0.02758x - 34.2568 \tag{5}$$

$$\text{High price linear regression equation:} y_3 = 0.026892x - 34.3603 \tag{6}$$

$$\text{Low price linear regression equation:} y_4 = 0.027239x - 34.8074 \tag{7}$$

The results of fitting the predicted stock price to the actual stock price are shown in following Figure 3, Figure 4, Figure 5 and Figure 6.

DOI: 10.54254/2754-1169/46/20230353

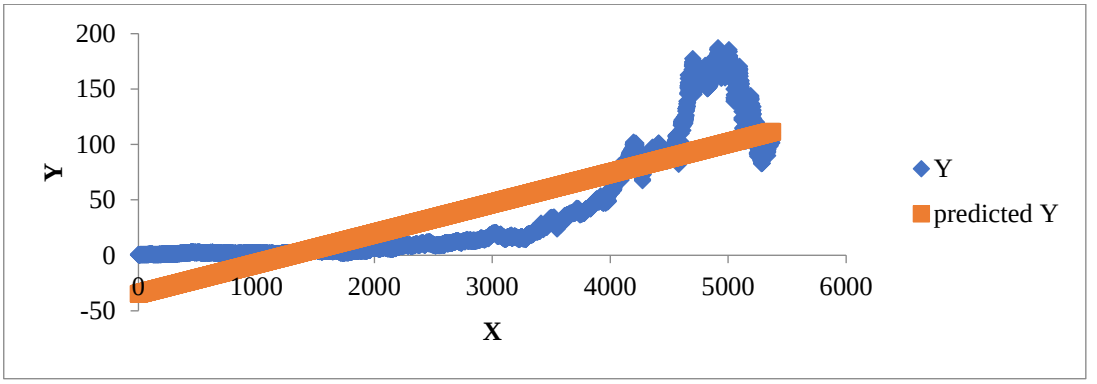

Figure 3: The predicted opening price based on the linear regression (Photo/Picture credit: Original).

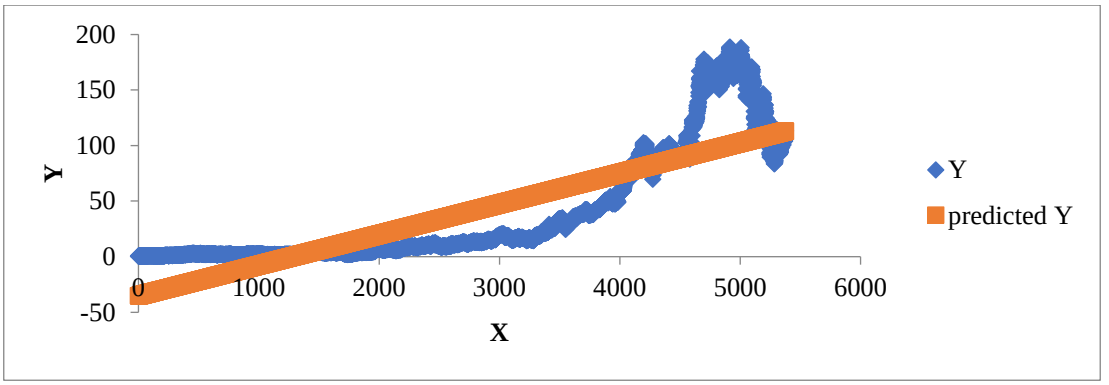

Figure 4: The predicted closing price based on the linear regression (Photo/Picture credit: Original).

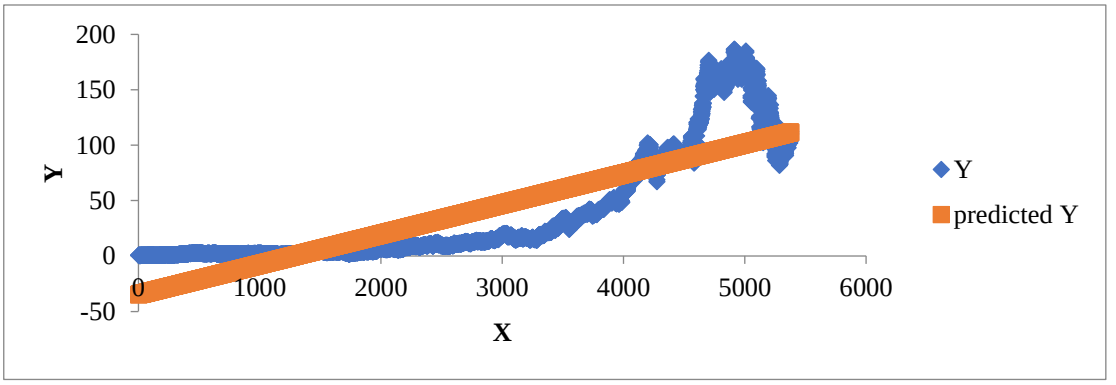

Figure 5: The predicted high price based on the linear regression (Photo/Picture credit: Original).

DOI: 10.54254/2754-1169/46/20230353

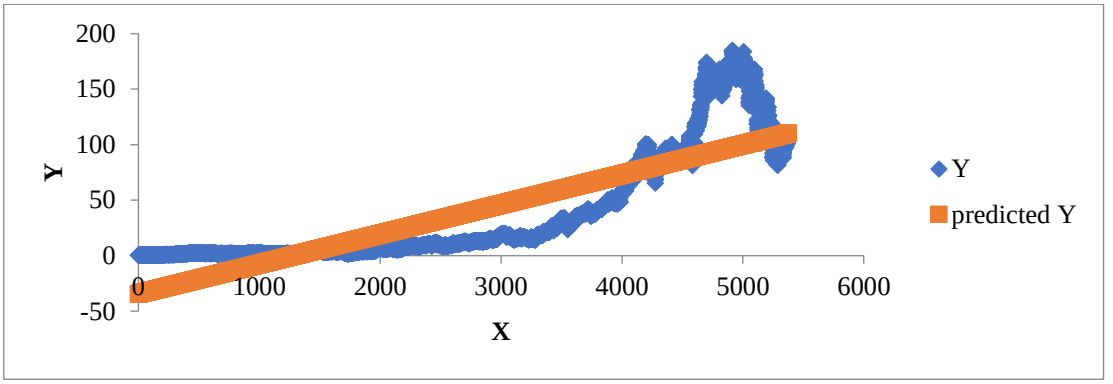

Figure 6: The predicted low price based on the linear regression (Photo/Picture credit: Original).

The statistics above demonstrate that the linear regression model does not accurately predict the stock price on a given day; it merely describes the movement of a stock. Actually, only responding to the forecasts' outcomes is insufficient; it is also important to assess the predictability and accuracy of the model. The correctness of the findings is evaluated in this research using the root mean square error.

From the results of the regression analysis, it can be observed that the RMSE of the above four one-dimensional regression equations are 28.20, 28.53, 27.83 and28.17, respectively. Additionally, the mean value is 28.1825.

In the LSTM model prediction, this paper used past data to predict the next trading day's earnings. The results of fitting the predicted stock price to the actual stock price for the test set are shown in Figure 7, Figure 8, Figure 9 and Figure 10.

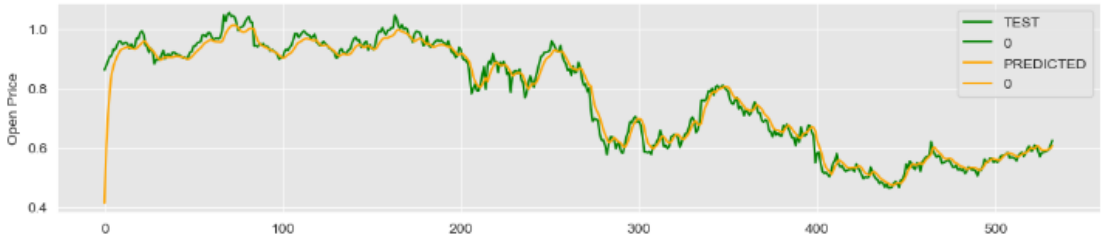

Figure 7: The predicted open price based on the LSTM model (Photo/Picture credit: Original).

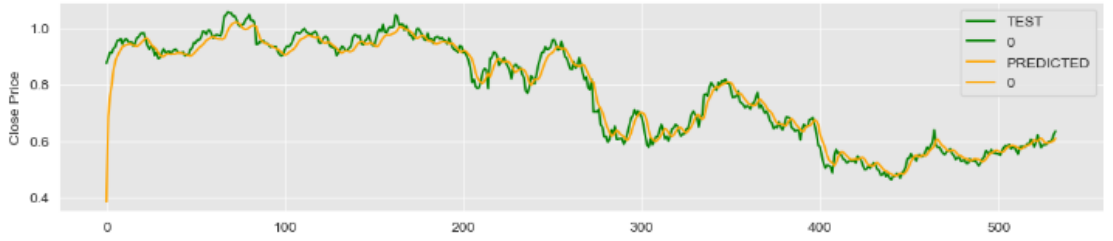

Figure 8: The predicted close price based on the LSTM model (Photo/Picture credit: Original).

DOI: 10.54254/2754-1169/46/20230353

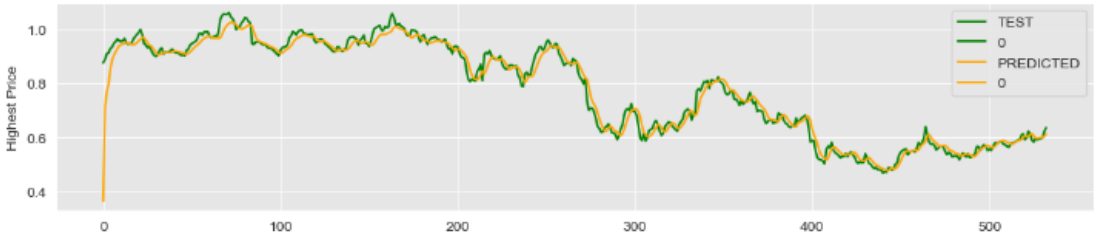

Figure 9: The predicted highest price based on the LSTM model (Photo/Picture credit: Original).

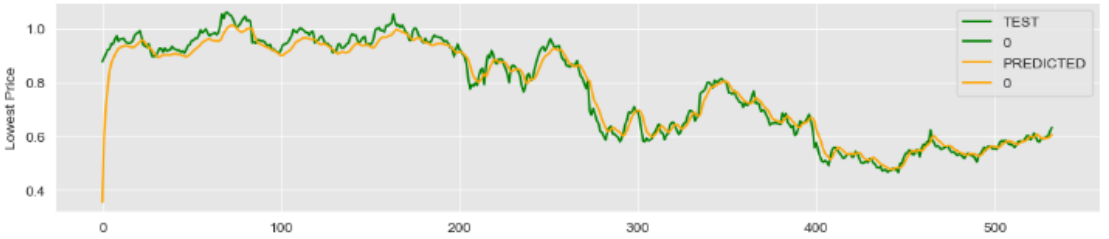

Figure 10: The predicted lowest price based on the LSTM model (Photo/Picture credit: Original).

The analysis of test results reveals that the employed LSTM model exhibits remarkable accuracy not only in predicting stock price movements but also in accurately estimating specific values, exhibiting a high degree of goodness of fit in the process. This study calculated the RMSE of the LSTM model's forecasting outcomes with a specific value of 0.0364, significantly lower than the RMSE of the linear regression model, to further confirm the model's accuracy. This illustrates that for predicting stock price, the LSTM model is more reliable than the linear regression model. By observing at the statistical results, it can be found that Amazon's share price showed a trend of decreasing volatility after the outbreak of the COVID-19. Amazon stock price volatility stabilises going into 2023. The LSTM model fits better and can effectively provide advice to decision makers.

4. Conclusion

In order to provide suggestions for pertinent decision makers, this study suggests a strategy for forecasting Amazon stock prices during the COVID-19. This paper accomplishes this by utilizing a machine learning methodology to train a stock price forecasting model on the basis of Amazon stock price data from 2002–2023. To anticipate stock prices, a linear regression model and a LSTM model are created. The suggested model is assessed by RMSE. The test results showed that the LSTM model beats the linear regression model and has an error that is substantially smaller than that of the latter, which can be widely used for stock price prediction during COVID-19. In the future, further study plans to apply this machine learning approach to serve more stock price prediction tasks, and continue to develop other non-linear models for stock price forecasting accurately.

References

[1] *Yang, L., Liu, S., Liu, J., et al. (2020) COVID-19: immunopathogenesis and Immunotherapeutics. Signal transduction and targeted therapy, 5(1): 128.*

[2] *Fauci, A. S., et al. (2020) Covid-19—navigating the uncharted. New England Journal of Medicine, 382(13): 1268-1269.*

[3] *Brodeur, A., Gray, D., Islam, A., et al. (2021) A literature review of the economics of COVID-19. Journal of Economic Surveys, 35(4): 1007-1044.*

[4] *Li, W., Chien, F., Waqas, K. H., et al. (2022) The nexus between COVID-19 fear and stock market volatility. Economic research-Ekonomska istraživanja, 35(1): 1765-1785.*

DOI: 10.54254/2754-1169/46/20230353

[5] Mazur, M., Dang, M., Vega, M. (2021) COVID-19 and the march 2020 stock market crash. Evidence from S&P1500. Finance research letters, 38: 101690.
[6] Buchanan, B. G. (2005) A (very) brief history of artificial intelligence. Ai Magazine, 26(4): 53-53.
[7] Cao, L. (2020) AI in finance: A review. Available at SSRN 3647625.
[8] Yaga, D., Mell, P., Roby, N., et al. (2019) Blockchain technology overview. arXiv preprint arXiv:1906.11078, 2019.
[9] Ali, H., Aysan, A. F. (2023) What will ChatGPT Revolutionize in Financial Industry?. Available at SSRN 4403372.
[10] Elmachtoub, A. N., Liang, J. C. N., McNellis, R. (2020) Decision trees for decision-making under the predict-then-optimize framework, International Conference on Machine Learning. PMLR, 2858-2867.
[11] Karlsson, S. (2013) Forecasting with Bayesian vector autoregression. Handbook of economic forecasting, 2: 791-897.
[12] Xu, F., Uszkoreit, H., Du, Y., et al. (2019) Explainable AI: A brief survey on history, research areas, approaches and challenges, Natural Language Processing and Chinese Computing: 8th CCF International Conference, NLPCC 2019, Dunhuang, China, October 9–14, 2019, Proceedings, Part II 8. Springer International Publishing, 2019: 563-574.
[13] Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. The Journal of Finance, 25(2), pp. 383–417. doi:10.2307/2325486.
[14] Mondal, P., Shit, L., Goswami, S. (2014) Study of effectiveness of time series modeling (ARIMA) in forecasting stock prices. International Journal of Computer Science, Engineering and Applications, 4(2): 13.
[15] Franses, P. H., Van, D. D. (1996) Forecasting stock market volatility using (non-linear) Garch models. Journal of forecasting, 15(3): 229-235.
[16] Selvin, S., Vinayakumar, R., Gopalakrishnan, E. A., et al. (2017) Stock price prediction using LSTM, RNN and CNN-sliding window model, 2017 international conference on advances in computing, communications and informatics (icacci). IEEE, 1643-1647.
[17] Yahoo (2023) Retrived from https://tw.stock.yahoo.com/quote/AMZN.