ISSN 2616-5902 Vol. 7, Issue 4: 39-43, DOI: 10.25236/AJBM.2025.070405

# Stock Price Prediction Using LSTM: A Case Study of Maotai

**Jiahui Teng[1,a,#], Jun Feng[1,b,#], Yukun Huang[1,c,#], Xinyi Li[1,d,#], Yiyang Tao[1,e,#], Jinming Xu[1,f,#]**

*[1]Beijing 21st Century School, Beijing, 100142, China*
*[a]jiahui8202021@163.com, [b]2947706478@qq.com, [c]2037450364@qq.com, [d]Lixiny8587@163.com, [e]taoyiyang0508@outlook.com, [f]13811569995@163.com*
*[#]Co-first author*

***Abstract:*** *Although the stock market is complex and fluctuates, making prediction challenging, accurately forecasting stock prices is crucial for investors to make informed decisions based on past performance. Hence, a suitable prediction model can significantly help investors take action in a prudent manner. In this paper, we design a stock price prediction model using LSTM (Long Short-Term Memory) and compare its performance on Maotai stock with other models based on the RMSE value. In comparison, the LSTM model outperforms other models with an RMSE value. Therefore, this model can help investors forecast stock prices more accurately to some extent.*

***Keywords:*** *Stock, Maotai, Machine learning, Prediction, LSTM*

## 1. Introduction

The stock market holds economic significance and the prediction of stock price has been a concern for investors and researchers. The accurate prediction of financial market fluctuations is very difficult, given the numerous unpredictable events that can influence stock market performance. In contrast to the simplistic view that past performance is a sole indicator of future outcomes, it is obvious that a multitude of variables, both predictable and unpredictable, contribute to the complexity of market dynamics. There are numerous financial indicators and information that need to be analyzed in order to get accurate prediction results. Among so many models, LSTM, which belongs to deep learning system can better handle with this information since it remembers long-term historical information and ignore irrelevant data, using past stock price trends to predict future price changes.

This paper aims to use the LSTM model to predict the price of Maotai stock and evaluate the prediction results. The research method involves selecting historical Maotai stock data, followed by data cleaning and feature engineering. The LSTM model is then used for training, and its predictive power is evaluated.

This paper will introduce the designed model in several sections. The second part will discuss the work of other researchers that is related to the subject of the current study. The third part will describe the data that the computer studied, the feature engineering and the overview of the algorithm. The fourth and final part will evaluate the result based on a specific standard.

## 2. Related Work

Researchers have explored the use of models such as LSTM or GRU to predict stock prices such as Fischer, T., & Krauss, C. The paper " Enhancing Stock Price Forecasting Accuracy Using LSTM and Bi-LSTM Models [1]" explores the application of LSTM in stock market prediction by using historical price data, LSTM models are able to capture long-term dependencies in time series and perform well on multiple stocks. Compared with the above paper authors. Time series forecasting of stock market indices based on DLWR-LSTM model

Literature [2] " is a paper predicting stock prices based on LSTM. Experiments show that the LSTM model has advantages in the processing of nonlinear time series data. Tackling misinformation in mobile social networks a BERT-LSTM approach for enhancing digital literacy [3]" proposes a stock price prediction method combining LSTM and sentiment analysis. By introducing social media sentiment data,

ISSN 2616-5902 Vol. 7, Issue 4: 39-43, DOI: 10.25236/AJBM.2025.070405

the model can better capture the impact of market sentiment on stock prices. This is a new way of predicting stock models. By. Bao, W., Yue, J., & Rao, Y. "(A Deep Learning Approach for Stock Market Prediction Using LSTM) [4] "paper, this paper proposes a deep learning framework based on LSTM for predicting short-term price movements of stock markets. The model uses historical price data and trading volume as input features and proposes and emphasizes the advantages of LSTM in processing high-frequency data and nonlinear relationships.

## 3. Data and algorithms

### *3.1. Data*

The data set contains stock's code, name each dates' closing price, the lowest price, the highest price, opening price, last price, change amount, change percent, turnover rate, turnover, total volume, circulation market value, market value from $18^{th}$ 5, 2004 to $3^{rd}$ June, 2020 for 3906 rows × 16 columns which correspond the 3905 days. The data is used for studying and predicting the stock tendency over the following period after founding a model and training the model by Long Short-Term Memory Model (hereinafter referred to as LGTM). The attributes of data are explained in detail in Table 1.

*Table 1: Attributes description*

| **Attributes** | **Description** |
|---|---|
| Stock's code | A unique series of letters assigned to a publicly traded company's stock. |
| Closing price | The final price at which a stock is traded during the regular trading session on a given day. |
| Lowest price | The lowest price at which the stock traded during the trading day. |
| Highest price | The highest price at which the stock traded during the trading day. |
| Opening price | The price at which the stock first trades when the market opens for the day. It is influenced by after-hours trading and pre-market activity. |
| Last price | The most recent price at which stock was traded. |
| Change amount | The percentage changes in the stock's price are relative to the previous day's closing price. |
| Turnover rate | A measure of how frequently stock is traded relative to its total shares outstanding. |
| Turnover | The total value of shares traded during a specific period, usually a day. |
| Total volume | The total number of shares traded during a specific period, typically a day. |
| Circulation market value | The total market value of a company's shares is available for trading. |
| Market value | The total market value of a company's outstanding shares. |

### *3.2. Feature engineering*

Feature engineering, a pivotal step in neural network-based stock price forecasting, assists in creating useful datasets according to stock market tendency and allows the model to learn it efficiently. As the stock market is complicated and dynamic, feature engineering is decisive to the eventual consequence of prediction.

By carefully engineering features, the performance of a neural network in stock price forecasting can be improved and make more accurate predictions. First of all, when there are considerable proportion of missing values, it is feasible that to directly discarded those values in order to minimize the negative impact of final result. Otherwise, we tackled the issue by filling in the data. After that, Principal Component Analysis (PCA) can be used to reduce the dimensionality of the data while retaining most of the important information. PCA transforms the original features into a new set of uncorrelated variables called principal components, which can be used as input features for the neural network.

Ultimately, some fundamental features including open, high, low, close, last and turnover volume are selected as the features. The closing price in historical prices is the final result of a stock's trading on a given day and reflects the market's comprehensive evaluation of the stock. Trading volume reflects the trading activity of the market. High trading volume is often followed by evident fluctuations in prices, therefore it is also included in the feature set.

ISSN 2616-5902 Vol. 7, Issue 4: 39-43, DOI: 10.25236/AJBM.2025.070405

### *3.3. Algorithm*

The algorithm used in this paper is LTSM (Long Short-Term Memory), which mainly introduces "memory units" and "gating mechanisms" to better capture long-distance dependencies. The "memory unit" is the core of the LSTM, which acts like a conveyor belt and can transmit information throughout the sequence. Its function is to preserve long-term information, avoid information loss in the process of transmission, to promote stability and efficiency of information transmission.

"Gate mechanism" is the mechanism used by the LTSM algorithm to control the flow of information. It mainly completes this task through three "gates", which are respectively the forgetting gate, the input gate and the output gate. The forgetting gate is used to determine what information is discarded from the memory unit, the formula is

$$f_t = \sigma\left(w_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{1}$$

Among them:σ is the Sigmoid activation function with an output value between 0 and 1; $W_f$ is the weight matrix of the forgetting gate. $B_f$ is the offset of the forgetting gate.

The input gate is used to determine which new information is stored in the memory unit, and its working steps are mainly divided into two parts:

Enter the activation value of the gate:

$$i_t = \sigma(w_i \cdot [h_{t-1}, x_t]) \tag{2}$$

Candidate memory unit state

$$\tilde{c}_t = \tanh(\mathrm{W_C} \cdot [\mathrm{h_{t-1}}, \mathrm{x_t}] + \mathrm{b_c}) \tag{3}$$

The output gate is used to determine what information is output from the memory unit to the hidden state $h_t$, the formula is

$$o_t = \sigma(w_0 \cdot [h_{t-1}, x_t] + b_0) \tag{4}$$

$$h_t - o_t \cdot \tanh(C_t) \tag{5}$$

These "gates" are controlled by the sigmoid function to control the flow of information, and in the process of work, after forgetting gate and the input gate, there is also a step to update the memory unit $C_t$. This step needs to update the memory unit by combining the results of the forgetting gate and the input gate, the formula is

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \tag{6}$$

### *3.4. Evaluation method*

More specifically, the opening price of the first (3905-300=3605) days is used as the training set, and the last 300 days is used as the test set, and the opening price is normalized so that the data fed into the neural network is distributed between 0 and 1. We then define an LSTM layer with X memory; and a certain amount of dropout. While reducing the accuracy of predictions, dropout allows the model to make more accurate predictions on data it has never been trained on. Similarly, without dropout, the model cannot make predictions on untrained data. Test machine input model to make predictions. The predicted data and the real data were restored respectively. Then the comparison curve between the real data and the predicted data is presented. Finally, output MSE, RMSE and MAE values to select the best model.

### *3.5. Result evaluation*

Through the above method, we draw the comparison curve between the predicted data and the actual data, which is shown in Figure 1. Through the output of MSE, RMSE and MAE three values and pictures, we can conclude that the change trend of the predicted value and the actual data is basically consistent within 300 days. The calculated values of MSE, RMSE and MAE provide concrete methods for measuring the accuracy of the model. In addition, we also tried two other models, namely the GRU and CNN+LSTM models. After 10 tests of untuned parameters (the parameters of the three models are the same), LSTM has the smallest error. The predictions were also more accurate, as shown in Figures 2 and 3 of the other two models. Besides, we draw table which shows the data helping people compare the degree of accuracy more precisely and clearly, the table is shown above in Table 2.

ISSN 2616-5902 Vol. 7, Issue 4: 39-43, DOI: 10.25236/AJBM.2025.070405

*Table 2: The data of each model*

| The data / Stocks' name | MSE | RMSE | MAE |
|---|---|---|---|
| LSTM | 0.292821 | 0.541129 | 0.271060 |
| GRU | 0.309831 | 0.556625 | 0.306597 |
| CNN+LSTM | 7.258976 | 2.694249 | 1.785906 |

In summary, the LSTM model gives reasonable prediction accuracy and stands out among many models, but there is still room for further improvement to better predict the actual stock price fluctuations, such as changing parameters, or conducting more rounds of simulation to achieve the best prediction effect, but due to dropout, errors are inevitable. One can test further at longer intervals to achieve better predictions.

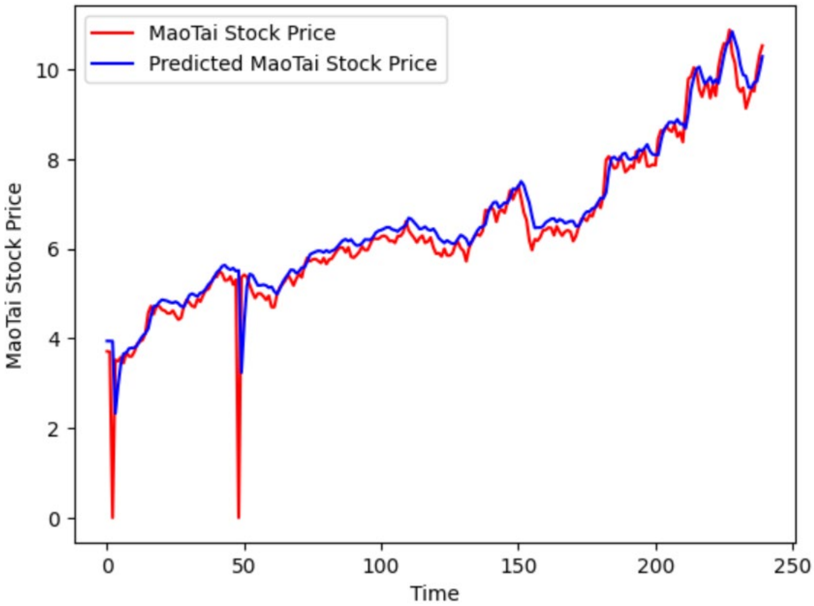


*Figure 1: MaoTai Stock Price Prediction by LSTM model*

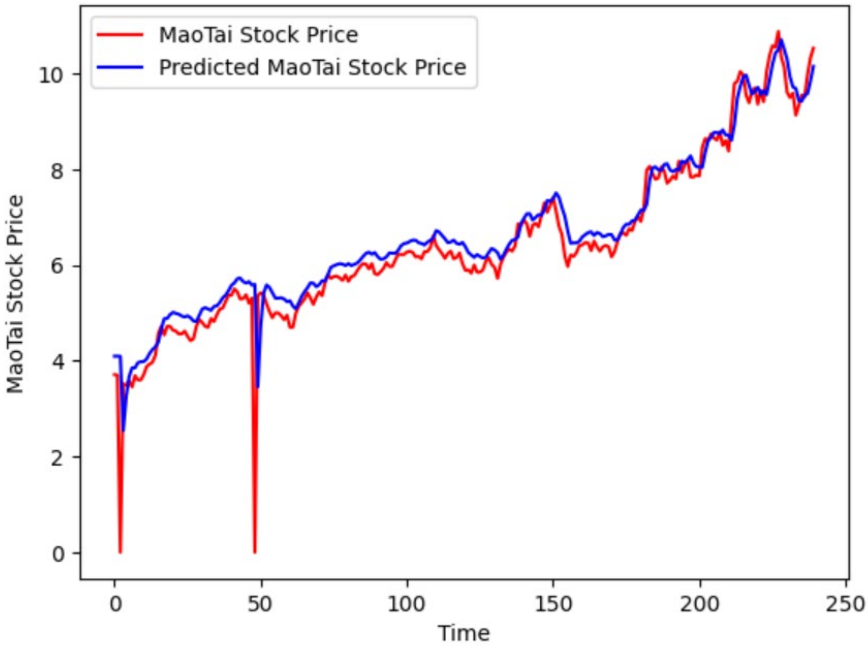


*Figure 2: Maotai Stock Price Prediction by GRU model.*

ISSN 2616-5902 Vol. 7, Issue 4: 39-43, DOI: 10.25236/AJBM.2025.070405

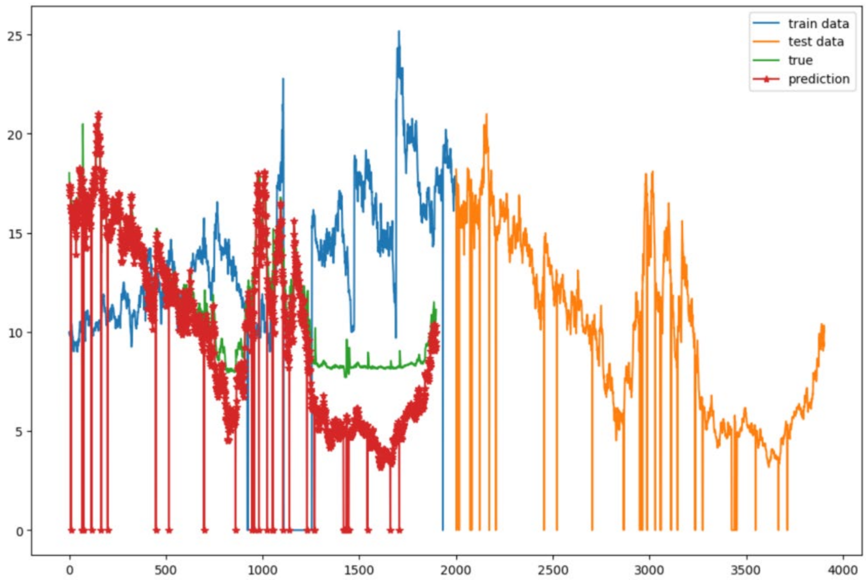


*Figure 3: Maotai Stock Price Prediction by CNN+LSTM model*

## 4. Conclusion

This study mainly describes how to establish a stock price prediction model based on LSTM to predict the trend and price of Maotai stock in the future period. It also examines how well the LSTM model captures the nonlinearity and long-term dependencies of stock prices. At the same time, three different models were tested under the condition of uniform parameters, and the model with the highest prediction accuracy was selected through the output data and pictures. The experimental results show that the LSTM model can predict the stock price well and can capture the long-term correlation and nonlinear characteristics in the time series data. In the future, the prediction accuracy and computational efficiency of LSTM model will be further improved.

## References

*[1] Wang, H. (2025). Enhancing Stock Price Forecasting Accuracy Using LSTM and Bi-LSTM Models. In ITM Web of Conferences (Vol. 70, p. 04008). EDP Sciences.*
*[2] Yao, D., & Yan, K. (2024). Time series forecasting of stock market indices based on DLWR-LSTM model. Finance Research Letters, 68, 105821.*
*[3] Cao, C., Li, Z., Li, L., & Luo, F. (2025). Research on the evolution of college online public opinion risk based on improved Grey Wolf Optimizer combined with LSTM model. PloS one, 20(1), e0311749.*
*[4] Bao, W., Yue, J., & Rao, Y. (2017). A deep learning framework for financial time series using stacked autoencoders and long-short term memory. PloS one, 12(7), e0180944.*