Proceedings of the 5th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/22/20231213

Prediction and investigation of stock price related to China's new energy vehicles after the opening of the pandemic based on the LSTM model

Yanghan Peng

Chongqing Nankai Middle School, Chongqing, 400030, China

631402150230@mails.cqjtu.edu.cn

Abstract. This research examines the repercussions of the COVID-19 pandemic on China's new energy vehicle market through the utilization of machine learning models for stock price prediction. Specifically, the Long Short-Term Memory (LSTM) and Artificial Neural Network (ANN) are employed to forecast stock prices and price changes for BYD, Changan Automobile, and Guangzhou Automobile Group. While the LSTM model successfully captures the patterns in the stock price data, it exhibits a lag of one day in its predicted outcomes, indicating its reliance on the previous day's price. However, both models encounter challenges in accurately predicting stock price changes, displaying notable disparities from actual values. The classification task of forecasting whether prices will rise or fall also yields unsatisfactory accuracy scores, highlighting the models' limitations in comprehending the dynamics of the stock market. This study reveals that understanding the impact of the pandemic on the NEV market holds significant importance for informed decision-making and effective navigation of China's automotive industry in the aftermath of the pandemic. Furthermore, further modification for the model is also required to enhance the precision and dependability of stock price forecasts.

Keywords: stock price prediction, machine learning, LSTM

1. Introduction

The COVID-19 pandemic has had a profound impact on various industries and markets globally, and the automotive sector is no exception [1, 2]. As pandemic control measures gradually ease, the new energy car market has emerged as a focal point of attention. Particularly, China, as one of the largest automotive markets, has attracted considerable investors' attention in the field of new energy vehicles. However, there are numerous factors that influence stock prices, such as economic indicators, company performance, industry outlook, political factors, market sentiment and so on. These factors interact with each other, making the fluctuation of stock prices complex and difficult to predict. Also, making rational and informed investment decisions is crucial. Despite human comprehension and technical analysis, discrepancies between outcomes and expectations can arise. Machine learning techniques offer the potential to enhance accuracy in predicting stock trends, thus facilitating informed recommendations for investment choices.

In recent years, deep learning has experienced rapid development, leading to the emergence of diverse algorithms and models that have found extensive applications across various domains, including natural language processing, computer vision, and finance [3-5]. Stock price prediction, in particular, has received significant attention, with researchers successfully utilizing models such as time series analysis, Sentiment Analysis (SA), Support Vector Machines (SVM) and Random Forest to forecast the price trends of certain stocks. For instance, the survey from Obthong et al. provides an overview of various machine learning techniques and algorithms used for stock price prediction, including the utilization of time series analysis and sentiment analysis [6]. Furthermore, Nikou et al. compared various machine learning algorithms and demonstrated that the deep learning method outperformed other methods [7], while the support vector regression method ranked second in performance, followed by the neural network and random forest methods in stock price prediction. However, despite the relatively accurate outcomes exhibited by these models in stock price prediction, there persist doubts regarding the dependability of such forecasts. The fluctuation of stock prices is influenced by a myriad of intricate elements, encompassing macroeconomic policies, financial states of companies, and market sentiment, among others. While deep learning models offer advantages in processing extensive datasets and discerning complex patterns, the capability of these models to effectively encapsulate and comprehend these intricate factors, thereby enabling precise predictions, remains an open question.

This paper uses stock price data from Changan Automobile, BYD, and Guangzhou Automobile Group and employs a Long Short-Term Memory Network model (LSTM) and artificial neural network for stock price prediction. The significance of this analysis lies in its contribution to the understanding of the actual effects of the pandemic on the new energy car market in China, providing real-time information and insights for relevant stakeholders. By employing machine learning models and data mining techniques, the aim is to discern whether the epidemic has yielded positive or negative consequences for the new energy vehicle stock market. Additionally, a more comprehensive examination was conducted on the forecasted stock prices, incorporating return rates for prediction and classification purposes. The experimental findings reveal that the model exhibits a delay effect, merely replicating the previous day's stock price without assimilating meaningful information. This observation is further substantiated by the results obtained from return rate prediction and classification, thus underscoring the unreliability of the predictions. Consequently, further refinements are warranted to enhance the precision and dependability of stock price forecasts.

2. Methodology

2.1. Dataset description and preprocessing

In this study, the opening price, closing price, and price change data for BYD, Changan Automobile, and Guangzhou Automobile Group from Investing were collected, spanning from January 5, 2015, to June 2, 2023 on each trading day. The stock price trends can be observed from in Figure 1.

DOI: 10.54254/2755-2721/22/20231213

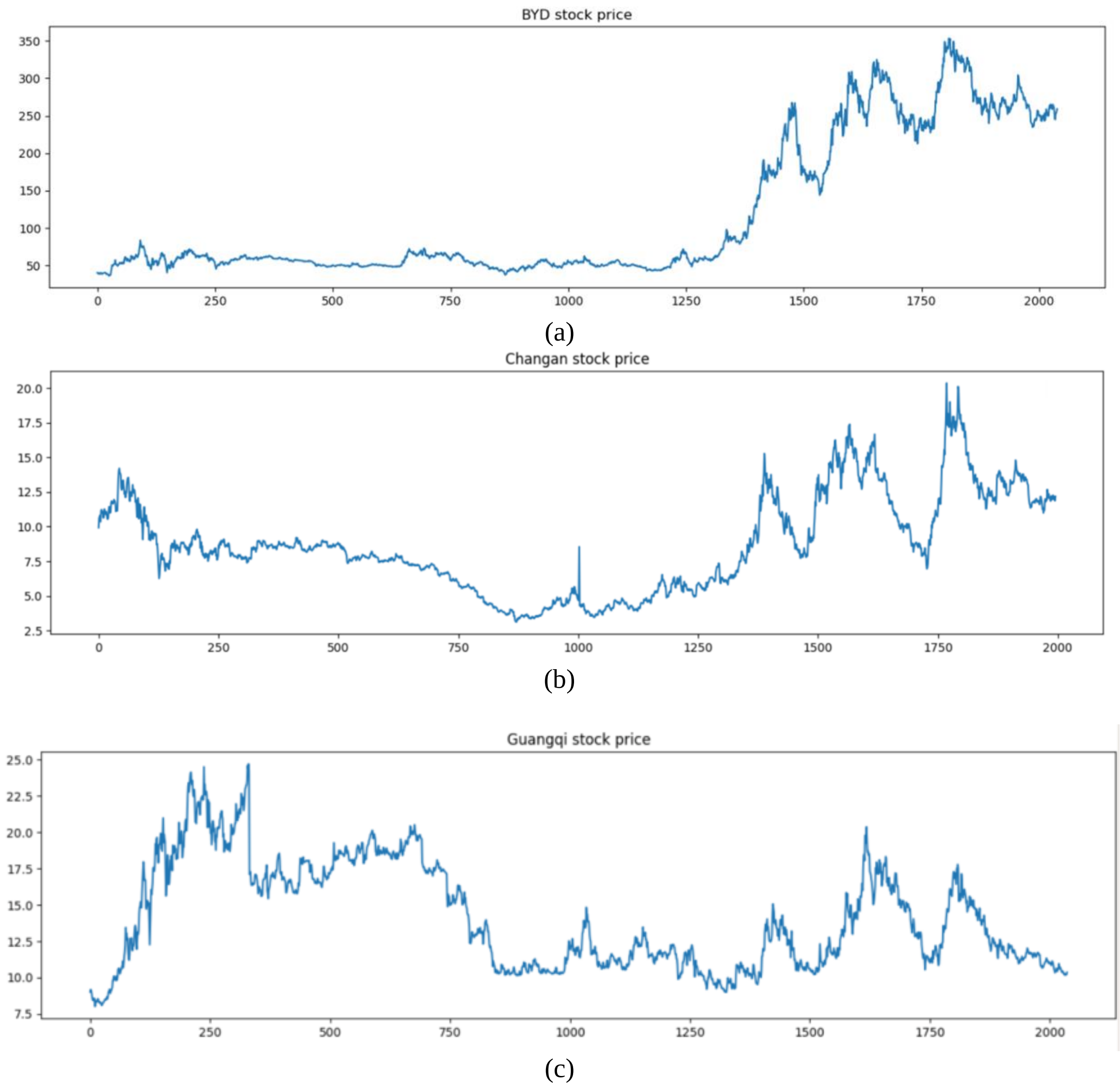


Figure 1. The stock price of three collected stocks related to (a) BYD (b) Changan Automobile, and (c) Guangzhou Automobile Group.

The dataset contains a significant number of data points, allowing for a comprehensive analysis of the stock price trends and fluctuations over time. The first 1500 days will serve as the training set, while the prices or returns for the following 500 days will be used as the test set. The min-max normalization was applied to the dataset for scaling the stock prices between 0 and 1, effectively removing scale differences across different stocks and speeding up the convergence of the model.

2.2. Long short-term memory network

In the finance industry, LSTM has gained significant attention due to its ability to handle time-series data and predict stock prices by capturing long-term dependencies. LSTM is a type of recurrent neural network that can learn order dependence in sequence prediction problems [8]. In the context of sequence prediction problems, LSTM stands out as a type of recurrent neural network that excels at learning the inherent order dependence within data. Its distinctive cell state allows for the storage of past information, while selectively retaining or discarding information based on its relevance to the current prediction. This characteristic is especially advantageous when dealing with sequential data, as it enables LSTM to retain crucial historical information over extended periods. Consequently, LSTM proves to be well-suited for predicting stock prices, given their nature as time-series data heavily influenced by past trends.

The model in this study consists of 3 LSTM layers, with the first layer having 32 neurons, the second layer having 5 neurons, and the final layer having only 1 neuron corresponding to the output of price or price change. The TensorFlow framework was employed to implement the model and trained it on a high-performance 3090 Ti graphics card. To train the model, this study employed the Adam optimizer and Mean Absolute Error (MAE) as the loss function. To improve model performance, this study conducted 20 training iterations, where the complete training dataset underwent forward and backward propagation through the LSTM layers, followed by parameter updates through the optimizer and loss function to minimize the error between predicted and actual values.

2.3. Artificial neural network

An ANN is a computational model that mimics the biological neural system and is used to solve complex nonlinear problems [9, 10]. The ANN model consists of multiple neurons connected in a hierarchical structure. It includes an input layer, hidden layers, and an output layer. The input layer receives external data as the model's input, the hidden layers perform information processing and transformation between the input data and output results, and the output layer provides the final prediction. The hidden layers play a crucial role in learning and extracting key features from the input data to gain a deeper understanding and representation of the problem. Each neuron calculates the weighted sum of input signals using weights and biases and applies a nonlinear transformation through an activation function.

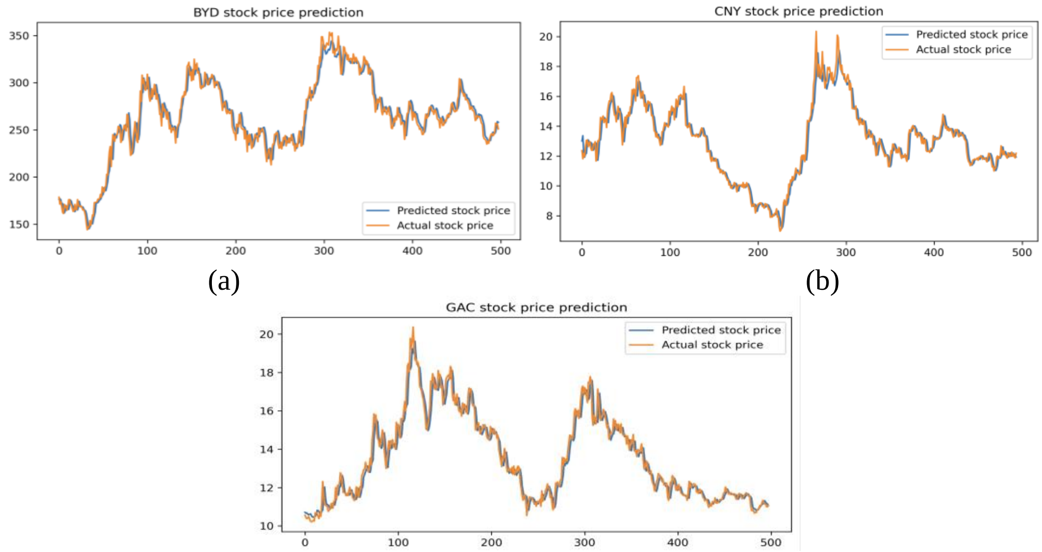
The architecture of the constructed ANN model encompasses four layers in this study. With a capacity of 50 neurons, the input layer can accommodate 50 distinct feature variables. To mitigate the risk of overfitting, a dropout layer with a probability of 0.2 is incorporated into the model. Following the input layer is the hidden layer, which comprises 25 neurons responsible for extracting and transforming nonlinear features from the input data. The output layer comprises a solitary neuron that produces a binary classification outcome indicating either an increase or decrease. During the model's training phase, the Adam optimizer is employed to iteratively adjust the weights and biases, minimizing the designated loss function. For this particular binary classification problem, the model employs binary cross-entropy as the chosen loss function. While the specific number of training iterations is not specified, it can be tailored based on specific requirements and the dataset size to attain desired model performance and generalization. By constructing and training this ANN model, it demonstrates the capacity to classify and predict whether stocks will rise, or fall based on the provided input data.

3. Results and discussion

3.1. Closing price prediction

This study evaluates the predictive performance of LSTM using metrics such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination (R-squared). According to the results shown in Figure 2 and Table 1, for the stock BYD, the model has an RMSE of 0.04206, MAE of 0.0327, and R-squared of 0.959 Similarly, for the stock CNY, the RMSE is 0.0401, MAE is 0.0275, and R-squared is 0.9444. For the stock GAC, the RMSE is 0.0467, MAE is 0.0331, and R-squared is 0.9538. The low values of these metrics indicate that the predictive model accurately fits the stock price data.

DOI: 10.54254/2755-2721/22/20231213



(a) (b) (c)

Figure 2. The comparison between the predicted and actual stock price.

Table 1. The predicted performance of different stocks in terms of the stock price.

Stock name	RMSE	MAE	R square
BYD	0.0421	0.0327	0.9590
CNY	0.0401	0.0275	0.9444
GAC	0.0467	0.0331	0.9538

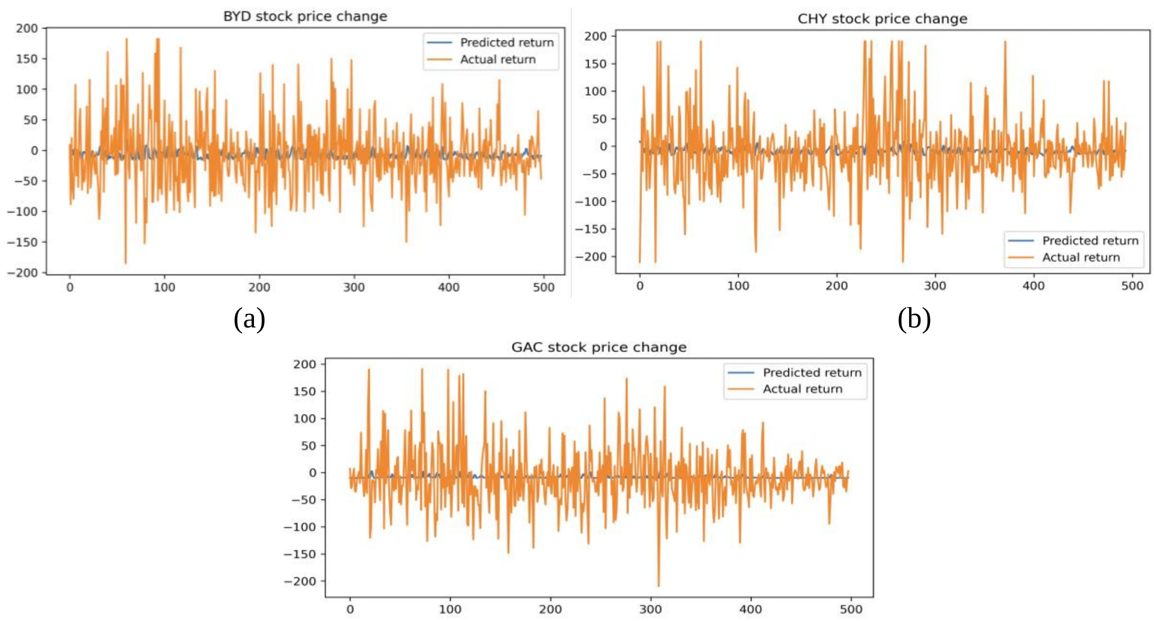
However, it can be also observed an important phenomenon, namely a one-day lag in the predicted results compared to the actual stock prices. This suggests that the model directly uses the previous day's stock price as an input during the prediction calculation, resulting in a time gap between the predicted and actual results. Therefore, further research is required to validate the reliability of the model on highly volatile stocks and determine whether adjustments to the model's input features or training approach are necessary.

3.2. Stock return prediction

Based on the Figure 3 and Table 2, including metrics called RMSE, MAE, and R-squared, it can be observed that predicting the fluctuations in stock return is challenging. The large values of RMSE and MAE indicate significant differences between the predicted and actual values. Additionally, the negative R-squared value suggests that the predictive model cannot explain a substantial portion of the variance in stock price movements.

This, in part, supports the notion that machine learning models may appear superficially accurate in predicting stock prices, but in reality, they may not have learned any meaningful information. If AI models were genuinely capable of accurately predicting stock price trends, they should also be able to forecast returns effectively. However, the results of the predictions are not satisfactory. This is reasonable because stock price movements are influenced by various factors, such as policies, which may be challenging for AI to capture.

DOI: 10.54254/2755-2721/22/20231213



(a) (b)

(c)

Figure 3. The comparison between the predicted and actual stock return.

Table 2. The predicted performance of different stocks in terms of the stock return.

Stock name	RMSE	MAE	R square
BYD	2.9119	2.1724	-79.1898
CNY	3.2034	2.3230	-173.4064
GAC	2.7131	1.9863	-493.7424

3.3. Stock price change prediction

Previous results have indicated the inherent difficulties encountered by AI models in accurately forecasting precise values of stock price fluctuations. To delve deeper into their predictive potential, this study adopts a classification approach to alleviate the complexity of prediction, with a specific focus on determining whether stock prices would increase or decrease. Regrettably, even within this classification framework, the performance of AI models in accurately predicting movements in the stock market has proven to be unsatisfactory.

Table 3. The classification score of the various stocks.

Stock name	Accuracy score
BYD	0.486
CNY	0.490
GAC	0.480

From Table 3, the experimental results indicate that the accuracy of AI models' predictions is even lower than random guessing, suggesting that AI models are unable to provide reliable forecasts for

DOI: 10.54254/2755-2721/22/20231213

highly volatile stocks. This difficulty may stem from the complexity and uncertainty of the stock market, making it challenging for models to capture the key factors involved.

4. Conclusion

In this study, the impact of the COVID-19 pandemic on China's new energy vehicle market was investigated using machine learning models for stock price prediction. The LSTM and ANN models were employed to forecast stock prices and price changes for BYD, Changan Automobile, and Guangzhou Automobile Group. The LSTM model demonstrated accurate fitting of the stock price data, indicating its capability to capture long-term dependencies. However, a one-day lag was observed in the predicted results, suggesting reliance on the previous day's price. Both models performed poorly in predicting stock price changes, with significant differences between predicted and actual values. When treated as a classification task to predict stock price rise or fall, the models' accuracy scores were lower than random guessing, highlighting their limited reliability for highly volatile stocks. To enhance the accuracy and reliability of the predictions, future research should focus on refining input features, addressing the lag issue, and incorporating additional data such as macroeconomic indicators and market sentiment. The complex and uncertain nature of the stock market indicates the need for a combination of machine learning techniques, expert analysis, and human judgment to provide more reliable predictions for China's NEV market.

References

[1] Klein C Høj J Machlica G 2021 The impacts of the COVID-19 crisis on the automotive sector in Central and Eastern European Countries.

[2] Černá I Éltető A Folfas P et al 2022 GVCs in Central Europe: A perspective of the automotive sector after COVID-19.

[3] Behrad F Abadeh M S 2022 An overview of deep learning methods for multimodal medical data mining Expert Systems with Applications 117006.

[4] Yu Q Wang J Jin Z et al 2022 Pose-guided matching based on deep learning for assessing quality of action on rehabilitation training Biomedical Signal Processing and Control vol 72 103323.

[5] Chen L Pelger M Zhu J 2023 Deep learning in asset pricing Management Science.

[6] Obthong M et al 2020 A survey on machine learning for stock price prediction: algorithms and techniques In 2nd International Conference on Finance, Economics, Management and IT Business Vienna House Diplomat Prague, Prague, Czech Republic pp 63-71.

[7] Nikou M et al 2019 Stock price prediction using deep learning algorithm and its comparison with machine learning algorithms. Intelligent Systems in Accounting, Finance and Management, 26(1) pp 22-42.

[8] Teo B G 2021 Stock Prices Prediction Using Long Short-Term Memory (LSTM) Model in Python (Medium) Retrieved from https://medium.com/the-handbook-of-coding-in-finance/stock-prices-prediction-using-long-short-term-memory-lstm-model-in-python-734dd1ed6827.

[9] Krogh A 2008 What are artificial neural networks? Nature biotechnology vol 26(2) pp 195-197.

[10] Zou J Han Y So S S 2009 Overview of artificial neural networks Artificial neural networks: methods and applications pp 14-22.