

The stock price of BYD prediction using LSTM and ARIMA

Bianmengyan Ji^{1,a}, Xiyuan Tong^{1,b}, Xiaohan Wang^{1,c}

¹*School of Mathematics and Physics, Xi'an Jiaotong-Liverpool University, Ren Ai Street, Suzhou, China*

^a*Bianmengyan.Ji21@student.xjtlu.edu.cn*, ^b*Xiyuan.Tong21@student.xjtlu.edu.cn*,

^c*Xiaohan.Wang21@student.xjtlu.edu.cn*

Keywords: BYD stock price, LSTM model, ARIMA model

Abstract: The increasing attention towards low-carbon initiatives has led to a surge in interest for products and services that contribute to a sustainable future. As a result, cap-and-trade policies, green bonds, and low-carbon stocks have emerged as significant areas of investigation. This study aims to explore the predictability of low-carbon stock prices, using BYD (Build Your Dream), a prominent new energy vehicle brand, as a case study. To effectively analyze and forecast BYD stock closing prices, we have evaluated various models and determined that Long Short-Term Memory (LSTM) and Autoregressive Integrated Moving Average (ARIMA) models exhibit superior performance in comparison to other alternatives. Employing a range of evaluation metrics, such as Standard Deviation (STD), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE), we demonstrate that the selected models exhibit a satisfactory level of fit. In future research endeavors, we aspire to expand the scope of our investigation to encompass additional facets of low-carbon stocks and their potential impact on the broader financial landscape.

1. Introduction

In recent years, global governments have exhibited growing concern regarding ecological preservation and carbon emission reduction. Over 40 governments have implemented various measures, such as carbon pricing, direct carbon taxation, or carbon cap-and-trade policies [1].

Cap-and-trade is a regulatory program established by governments to restrict or cap the aggregate emissions of specific chemicals, primarily CO₂, resulting from industrial activities. Advocates of cap-and-trade argue that it offers a viable alternative to carbon taxation, as it aims to mitigate environmental harm without unduly burdening industries. For example, in 2005, the European Union (EU) introduced the world's first international cap-and-trade program to address carbon emissions, anticipating a 21% reduction in emissions by 2020 [2].

Green bonds have also gained traction as fixed-income instruments designated for raising funds for climate and environmental initiatives. These bonds encompass a broader category of instruments linked to environmentally beneficial projects. The World Bank, a prominent green bond issuer, has issued \$14.4 billion in green bonds between 2008 and 2020, supporting 111 global projects in areas such as renewable energy, clean transportation, and sustainable agriculture. The Rampur Hydropower

Project, for instance, provided low-carbon hydroelectric power to northern India's electrical grid, preventing 1.4 million tons of carbon emissions through green bond financing [3].

Moreover, addressing the challenges of transitioning to a low-carbon economy necessitates the utilization of financial instruments like green bonds and low-carbon stocks to expedite eco-friendly investment. Assessing the risk and hedging characteristics of these securities is crucial for developing climate-conscious portfolios and formulating financial incentives that encourage private and institutional investors to allocate resources towards low-carbon initiatives. Green bonds can offer protection against unfavorable price fluctuations in low-carbon stocks, enabling environmentally aware investors to hedge portfolio risks using green financial instruments [4]. Consequently, the low-carbon trend appears to be gaining irreversible momentum. In this article, we first introduce the advantages and disadvantages of several current stock forecasting models in the literature review, then clarify the current situation of BYD and the popularity and significance of stock forecasting. Next, we use the review as an introduction of the model we use, and analyzes the pictures produced by the model. Finally, we make a conclusion.

2. Literature Review

In this section, we review the relevant literature pertaining to our study, which encompasses three primary streams. First, we examine literature discussing the merits and limitations of XGBoost, LSTM, AR, MA, ARIMA, ARCH, GARCH, and Neural Networks. Second, we assess literature on BYD corporate status. Lastly, we explore literature emphasizing the significance of stock price analysis. By synthesizing these literature reviews, we propose employing LSTM and ARIMA models for predicting and analyzing BYD future stock prices.

In recent years, numerous scholars have concentrated on the evaluation of prediction models. Regarding LSTM, Sairam and K [5] and Qiao et al. [6] both found that this model exhibits superior accuracy. LSTM neural networks can extract information from extensive original data without relying on prior knowledge of predictors [6]. With respect to ARIMA, Goyal and Raj [7] demonstrated that this model can be applied to both univariate and multivariate time series data, yielding satisfactory results in terms of prediction accuracy and error rates. However, Shao [8] argued that XGBoost and LSTM models may struggle to achieve optimal prediction performance when data volume is limited and single-factor correlations are excessively high. Additionally, Selvin et al. [9] contended that both linear (AR, MA, ARIMA) and non-linear algorithms (ARCH, GARCH, Neural Networks), which are employed to forecast stock indices, merely fit data to specific models instead of identifying underlying dynamics present within the data. Consequently, these prediction models may fail to accurately capture dynamic changes.

The prediction of stock prices and returns has long been an active area of research. The stock market serves as a crucial conduit between fundraisers and investors in the capital market while facilitating flexible buying and selling of stocks in the secondary market. Furthermore, the stock market plays a vital role in resource allocation, enabling the free flow of capital [10]. Consequently, it is essential to analyze factors influencing stock prices and to price stocks accurately from both investor and financial market perspectives [10]. Investors and researchers aim to predict future stock price trends and returns [6], making stock market forecasting a highly sought-after topic [11]. Although specific evaluations are conducted, they do not always yield precise results, necessitating the development of strategies for more accurate predictions.

In this study, we will analyze and forecast BYD stock prices using short- and long-term models [6]. BYD is one of three major high-tech enterprises integrating IT, automotive, and new energy technology industries. The company has collaborated with international giants such as General Motors, Chrysler, and Daimler-Benz on electric vehicle projects, and researched new energy vehicles

with Intel in 2011 [12]. With a higher market value of debt, BYD holds a substantial advantage over competitors like Mercedes Benz and Tesla. The trade-off theory posits that a higher market value of debt can facilitate corporate income tax deductions and generate increased cash flows, potentially improving managerial efficiency and reducing unnecessary expenditures [13].

Despite a 27.7% decline in stock value in 2022, BYD stock has risen by 23.6% in 2023. The company has penetrated multiple major markets and plans further expansion in the coming months, while simultaneously broadening its model lineup and targeting upscale markets [14]. Amid increasing demand for low-carbon products and pressure to reduce carbon emissions, BYD and Tesla have successfully developed low-carbon vehicles by substituting gasoline and diesel with lithium-ion power batteries [1]. BYD operational performance and industry standing indicate strong profitability and explosive growth potential in the new energy sector, presenting a favorable outlook for the company's stock. A positive correlation between the company's share price and price-to-earnings (P/E) ratio aligns with expectations [15].

Based on the research discussed, our team has chosen to employ the LSTM and ARIMA models for our analysis.

3. Model

(1) LSTM Model

The Long Short-Term Memory (LSTM) model comprises memory cells that replace traditional artificial neurons in hidden networks. These cells enable networks to retain data structure over time, exhibiting a high predictive capacity and making them well-suited for stably capturing data structures across time periods [11]. In the LSTM architecture, LSTM cells replace hidden layers. These cells consist of various gates that regulate input flow. An LSTM cell contains an input gate (which includes the input), a cell state (which extends across the entire network and can add or remove information via gates), a forget gate (which determines the proportion of permitted information), and an output gate (which includes the LSTM-generated output). Additionally, the cell features a sigmoid layer (which generates values ranging from zero to one, indicating the extent to which each component should be allowed), a tanh layer (which produces a new vector to be added to the state), and a point-wise multiplication operation. The cell state is updated based on these outputs. Mathematically, the LSTM model can be represented by the following equations, where x_t denotes the input vector, h_t represents the output vector, c_t signifies the cell state vector, f_t indicates the forget gate vector, i_t corresponds to the input gate vector, o_t symbolizes the output gate vector, and W and b are the parameter matrix and vector, respectively [11].

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$c_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (3)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (4)$$

$$h_t = o_t * \tanh(c_t) \quad (5)$$

The LSTM model can predict future instances by using information from previous lags [9].

(2) ARIMA Model

The ARIMA model is rooted in statistical methods for analyzing and modifying time series data. It accommodates a wide range of autoregressive configurations in time series data, connecting persistent and lagged dependencies through a dynamic relationship. The model utilizes the residual errors of the moving average model to monitor lagged compliance [11]. ARIMA models are widely employed in economic and financial time series modeling. As a linear method, ARIMA is designed

to be a linear function of past observations for predicting the future value of a variable. The ARIMA model is a generalized autoregressive moving average (ARMA) model, combining autoregressive (AR) processes with moving average (MA) processes to create a composite time series model. In this context, AR refers to autoregression, which is a regression model that leverages the relationships between observations and multiple lagged observations (p); I denotes integration, which is used to adjust the time series by calculating differences at various time points (d); and MA signifies moving average, which accounts for the dependencies between observations and residual error terms when using a moving average model to examine lagged observations (q) [15].

4. Experimental Analysis

In the experiment, linear model (LSTM) and nonlinear model (ARIMA) are used to process the data. Since the data is time continuous, both are sequential storage structures. We used data on BYD stock prices in recent years obtained from the Yahoo Finance website. The data included the opening prices, highest prices, lowest prices, closing prices, adjusted closing prices, and the volume.



Figure 1: BYD Close Price History.

We collect BYD stock closing price data from 10th, January, 2000 to 3rd, April, 2023. The line chart demonstrates how BYD stock price varies during the last 200 weeks. There are total five lines in Figure 1 that show the various trends. The blue one is about connecting the dots of 200 weeks that one dot per week, which shows the most specific changing situation of BYD closing stock price. We can see there are lots of fluctuation pots in blue curve. The red one are more glabrous than the blue ones, corresponding to one dot per 10 weeks. The yellow one corresponds to one dot per 20 weeks. The green one corresponds to one dot per 30 weeks. The purple one corresponds to one dot per 40 weeks, which is a simplified smooth curve for BYD closing price history. The picture compares the five lines and their trends fluctuate but overall trends are up, especially in the duration of 125 and 75 weeks before. The results lay the foundation for the prediction of BYD stock.

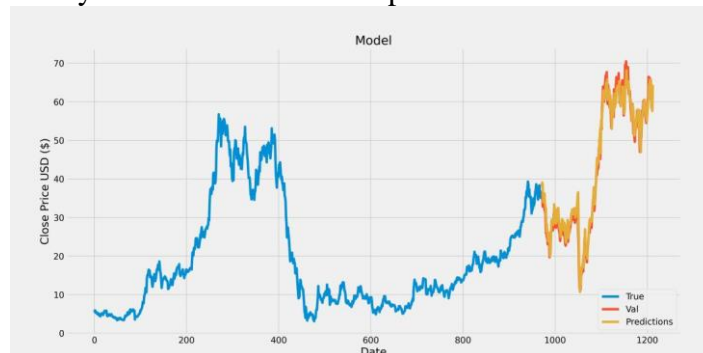


Figure 2: LSTM Model.

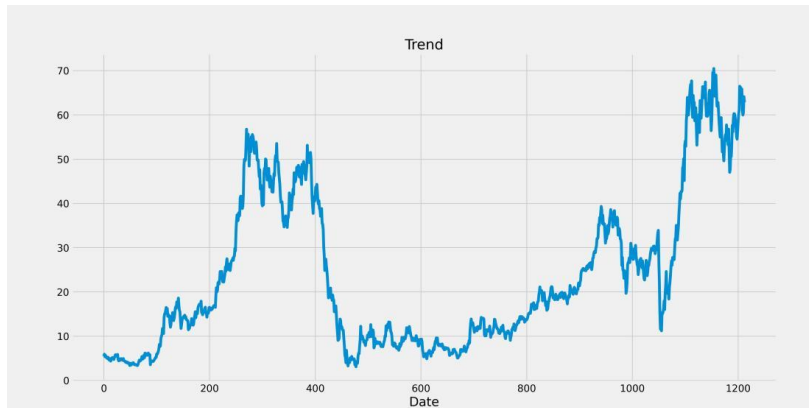


Figure 3: Trend.

LSTM (Long Short Term Memory) is a special recursive neural network. This kind of network is different from the general feedforward neural network, LSTM can use time series to analyze the input. According to the data of the stock prices of BYD from 10.1.2000 to 3.4.2023, we created Figure 2. Figure 2 shows LSTM model's output of close shock price of BYD. We choose 80 percent of BYD stock closing price data as training data and the remaining 20 percent as testing data. The True values of BYD stock are colored blue, and the Val prices are printed red, and the Predictions of its stock prices are colored yellow. From the chart, we can see that in the first four hundred weeks or so, the real price volatility of stocks rises, then falls sharply in the following one hundred weeks or so. Approximately, it rises steadily between 500 and 1000 weeks, and rises sharply around 1100 week. Around 1175 weeks, the BYD stock price reached the peak, which is over 70. Figure 3 shows the results for the LSTM model on the same data set and inputs. Compared with Figure 2, there is only one undulation line in Figure 3 to show the trade of BYD shock price trend distinctly.

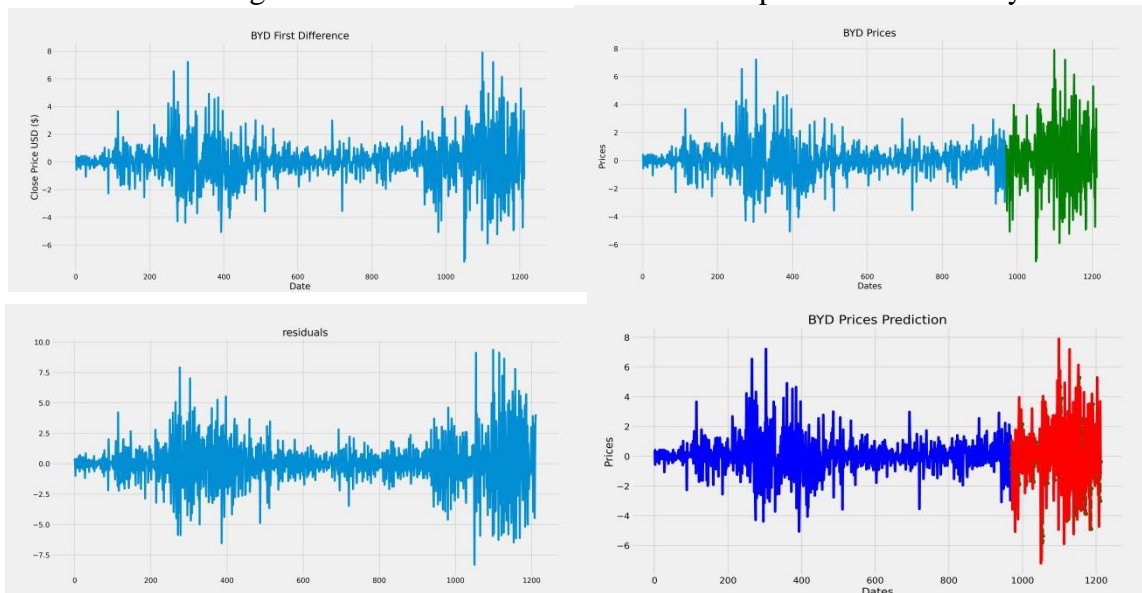


Figure 4: ARIME Model.

We use first-difference to measure the change rate or growth rate of BYD closing price to analyze the trend of BYD stock price. A positive value on the y-axis indicates that BYD stock price has grown during this period and the size of this value stands for how much the stock has grown. Conversely, a negative value shows the decreasing degree of BYD stock price. Figure 4 shows that in recent 200 weeks, the stock price of BYD has huge fluctuation, the maximum absolute value is

nearly 8 USD, indicating the risk was increasing for the investors. During 300 to 700 weeks before, the trend of BYD stock price is relatively smooth. On the ARIMA model we introduced before, we choose 80 percent of BYD stock price data as training data and the remaining 20 percent as testing data. From Figure 4, we can observe the goodness-of-fit between the actual price and the predicted price. In terms of the degree of coincidence, our ARIMA model has a high prediction accuracy.

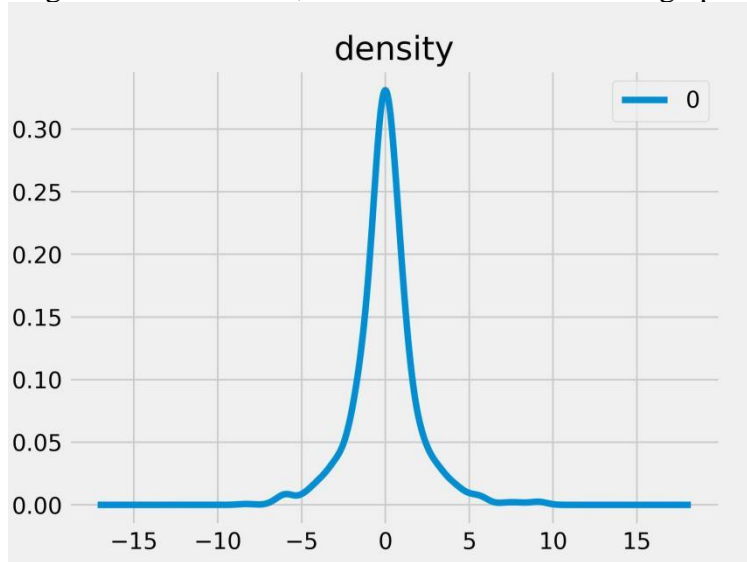


Figure 5: Density.

Then we calculate the STD(Standard Deviation), MSE (Mean Squared Error), RMSE (Root Mean Squared Error) and draw the density distribution curve (Figure 5) to examine the accuracy of our ARIMA model. The formulas of these three methods are below:

$$MSE = \frac{\sum_{i=1}^r (n_i - 1) s_i^2}{N - r} \quad (6)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^r (n_i - 1) s_i^2}{N - r}} \quad (7)$$

After calculating with code, we find that the mean is -0.000101, which demonstrates the are most highly frequent. Nearly 50% change rate is between -0.613203 and 0.627373. Meanwhile, our data shows STD= 1.541537, MSE= 9.317 and RMSE= 3.052, which are all quite small and indicate the model fits well and is stable.

5. Conclusion

In this study, we initially focus on the low-carbon society and examine BYD, a representative new energy vehicle company. We then review and synthesize the research of numerous scholars to gain an understanding of various prediction models, ultimately selecting the LSTM and ARIMA models for our analysis. Subsequently, we predict BYD stock price using both the LSTM and ARIMA models and discuss the accuracy of our prediction models by employing metrics such as Standard Deviation (STD), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). In predicting BYD stock price, both the LSTM and ARIMA models demonstrate high accuracy.

Our research offers several avenues for future exploration, including delving deeper into BYD history, examining the impact of the change rate on relevant indices, assessing BYD future development trends, and more. This would provide a more comprehensive understanding of BYD stock price dynamics and contribute to the broader literature on low-carbon economies and stock

price prediction.

References

- [1] Wang, H., Wang, L. and Ding, Q. (2022) 'Manufacturers' product line strategies and environmental impacts under carbon cap and trade policies', *Environmental Science and Pollution Research*, pp. 1–15. doi:10. 1007/s11356-021-17909-1.
- [2] Kenton, W. (2020) *Cap and Trade Basics: What It Is, How It Works, Pros & Cons*. Available at: <https://www.investopedia.com/terms/c/cap-and-trade.asp#toc-how-did-cap-and-trade-work-in-california> (Accessed: 14 April 2023)
- [3] Segal, T. (2022) *Green Bond: Types, How to Buy, and FAQs*. Available at: <https://www.investopedia.com/terms/g/green-bond.asp#toc-real-world-example-of-green-bonds> (Accessed: 14 April 2023)
- [4] Reboredo, J. C., Ugolini, A. and Ojea-Ferreiro, J. (2022) 'Do green bonds de-risk investment in low-carbon stocks?', *Economic Modelling*, 108. doi:10. 1016/j. econmod. 2022. 105765.
- [5] Sairam, P. V. and K, L. (2022) 'Automatic Stock Market Prediction using Novel Long Short Term Memory Algorithm compared with Logistic Regression for improved F1 score', 2022nd International Conference on Innovative Practices in Technology and Management (ICIPTM), *Innovative Practices in Technology and Management (ICIPTM)*, 2022 2nd International Conference on, 2, pp. 578–582. doi:10. 1109/ICIPTM54933. 2022. 9754116.
- [6] Qiao, R., Chen, W. and Qiao, Y (2022) 'Prediction of stock return by LSTM neural network', *Applied Artificial Intelligence*, 36(1), pp. 1–20. doi:10. 1080/08839514. 2022. 2151159.
- [7] Goyal, V. and Raj, R. R. (2022) 'Using LSTM, DNN, and ARIMA approaches to predict the trends in the Stock Market', 2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS), *Advanced Computing and Communication Systems (ICACCS)*, 2022 8th International Conference on, 1, pp. 918–922. doi:10. 1109/ICACCS54159. 2022. 9785150.
- [8] Shao, G. (2022) 'Stock price prediction based on multifactorial linear models and machine learning approaches', 2022 IEEE Conference on Telecommunications, Optics and Computer Science (TOCS), *Telecommunications, Optics and Computer Science (TOCS)*, 2022 IEEE Conference on, pp. 319–324. doi:10. 1109/TOCS56154. 2022. 10016086.
- [9] Selvin, S. 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), *Advances in Computing, Communications and Informatics (ICACCI)*, 2017 International Conference on, pp. 1643–1647. doi:10. 1109/ICACCI. 2017. 8126078.
- [10] Zhao, Y. (2022) 'A Comparative Analysis of Multiple Linear Regression Models and Neural Networks for Stock Price Prediction - Take BYD as an Example', *Proceedings of the 2022 2nd International Conference on Enterprise Management and Economic Development (ICEMED 2022)*, pp. 221-226. doi: 10. 2991/aebmr. k. 220603. 038
- [11] Gayatri, V. and Rao, P. P. (2022) 'Predictive Analytics for a Trustworthy Stock Exchange using LSTM and ARIMA Model', 2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS), *Advanced Computing and Communication Systems (ICACCS)*, 2022 8th International Conference on, 1, pp. 1419–1424. doi:10. 1109/ICACCS54159. 2022. 9785198.
- [12] Liu, J. and Meng, Z. (2017) 'Innovation Model Analysis of New Energy Vehicles: Taking Toyota, Tesla and BYD as an Example', *Procedia Engineering*, 174, pp. 965–972. doi:10. 1016/j. proeng. 2017. 01. 248.
- [13] Qiu, Y., Zhang, K, Zong, Y. (2022) 'The Market Analysis for BYD', *Proceedings of the 2022 2nd International Conference on Financial Management and Economic Transition (FMET 2022)*, pp. 246-256. doi: 10. 2991/978-94-6463-054-1_28
- [14] Carson, E. (2023) 'Tesla Stock Vs. BYD Stock: Racing Tesla Taps On Brakes; BYD Set To Boom After China New Year', *Investors Business Daily*, 13 February, p. N. PAG. Available at: <https://search-ebshost-com.ez.xjtlu.edu.cn/login.aspxdirect=true&db=bwh&AN=161846700&site=eds-live&scope=site> (Accessed: 13 April 2023).
- [15] Wijesinghe, G. and Rathnayaka, R. (2020) "Stock Market Price Forecasting using ARIMA vs ANN; A Case study from CSE, " 2020 2nd International Conference on Advancements in Computing (ICAC), Malabe, Sri Lanka, 2020, pp. 269-274, doi: 10. 1109/ICAC51239. 2020. 9357288.