Proceedings of the 2nd International Conference on Financial Technology and Business Analysis
DOI: 10.54254/2754-1169/46/20230357

The Prediction of Tesla Stock Price During COVID-19 Utilizing Machine Learning Methods

Lijia Li[1,a,*]

[1]*Zhengzhou Foreign Language School, Fengyang Street, Zhengzhou, China*
a. Lijia.Li@calhoun.edu.cn
**corresponding author*

Abstract: Stock forecasting is an act of predicting the future trend of the stock price based on the fluctuation of the stock price over a period of time. Due to the economic lockdown caused by the COVID-19 pandemic, it is challenging to provide accurate forecasts of price movements to ensure the reliability of investments. This study used Long Short-Term Memory (LSTM), Gate Recurrent Unit (GRU), Random Forest and K-Nearest Neighbor (KNN) algorithms to predict the stock price of Tesla in the special period of the epidemic. In the training data set, the fluctuation situation during the epidemic period was added as the basis for prediction, and as expected, the stock trend of Tesla in the special period was more accurately predicted. According to the results, LSTM and GRU models have high accuracy. They can provide investors with more reliable information about stock prices, and their R-squared values reach nearly 0.95, while the rest two models had poor performances.

Keywords: stock price prediction, machine learning, deep learning

1. Introduction

Stock prediction entails the activities of proficient securities analysts who possess an intricate comprehension of the stock market, aiming to prognosticate the forthcoming trajectory of the stock market as well as the magnitude of ascents and descents. This practice holds the potential to aid investors in mitigating asset losses attributable to price fluctuations and other pertinent factors, thereby diminishing risks and enhancing returns. Moreover, it extends support to nations and organizations in formulating policy modifications. However, during periods of a pandemic, the prediction of stock prices encounters heightened challenges, particularly in the case of enterprises like Tesla that predominantly focus on electric vehicle and energy development. The sales volume of such enterprises is susceptible to being influenced, consequently exerting an impact on the corresponding stock prices, which requires more attention.

The result of stock prediction is affected by many factors. The dynamics within the stock market are intricately linked to various factors encompassing macroeconomic progress of a nation, the formulation and implementation of legislative measures, the operational efficacy of corporations, and the level of confidence exhibited by shareholders. These interrelated elements collectively shape and influence the fluctuations and trends observed within the stock market. Therefore, the so-called forecast is difficult to predict accurately. The forecast of securities analysts can only be used as a general reference for the operation of the stock market.

DOI: 10.54254/2754-1169/46/20230357

For machine learning, many algorithms have been applied to predict results in various fields, such as computer, medicine, natural source processing and finance. One particular investigation conducted by Nelson et al. (2017) employed the Long Short Term Memory (LSTM) model to forecast the future fluctuations in stock prices [1]. The findings exhibited the superior performance of the LSTM model when compared to conventional machine learning models and non-time series models. Similarly, Bao et al. (2017) proposed an integrated model and demonstrated it outperformed other comparable models [2]. Furthermore, Vo et al. (2019) observed that the Bi-LSTM model, which read the data in both forward and backward directions, exhibited enhanced prediction accuracy [3]. Lastly, Peng et al. (2019) focused primarily on the preprocessing methodology of data, incorporating techniques such as interpolation, wavelet de-noising, and data normalization, while exploring parameter configuration of the LSTM model [4].

However, these people did not take into account the impact of the special period when making the forecast, which is lack of challenge. In recent years, as the coronavirus epidemic has intensified, the share prices of many companies have been severely affected. The sudden limit and fall led to many investors failed to make timely adjustments to the program, resulting in huge losses.

Therefore, in order to reduce the impact on investment during the special period since the outbreak of COVID-19, this study used the Long Short Term Memory (LSTM) model, Gate Recurrent Unit (GRU) model, Random Forest algorithm and K-Nearest Neighbor (KNN) algorithm to predict the closing price of Tesla stock during the special period of COVID-19 based on the stock trend from 2010 to 2021 (including one year of COVID-19). The final result shows that this work successfully predicted the closing price.

2. Methods

2.1. Dataset Description and Preprocessing

This study employed the Tesla stock dataset provided by a dataset on Kaggle [5]. The dataset provides detailed information related to features such as Date, Open, High, Low, Close, Adj Close, Volumn etc. for the study. The original dataset consisted of 2,957 rows and 7 columns, spanning from January 29, 2010 to March 24, 2022.

Prior to conducting further analysis, data preprocessing procedures were implemented. Firstly, the dataset was examined for the presence of Null or NA values, which were subsequently eliminated. Subsequently, the closing price variable was identified as the target for prediction and underwent preprocessing, including normalization to a range of 0 to 1. Lastly, the dataset was partitioned into training and testing subsets using a specific ratio of 90:10, respectively, to facilitate model training and evaluation.

2.2. Machine Learning Models

2.2.1. LSTM Model

LSTM represents a refined iteration of the conventional Recurrent Neural Network (RNN). Distinguished from its predecessor, LSTM exhibits enhanced capacity in capturing meaningful correlations within extended sequences while mitigating the issues of vanishing or exploding gradients. Analogous to RNN, LSTM operates with two outputs and three inputs to facilitate the progression to the subsequent time step.

2.2.2. GRU Model

GRU shown in Figure 1 is a commonly used gated recurrent neural network to better capture the dependency relationships with large intervals in sequential data. It is a kind of recurrent neural networks. Similar to LSTM, GRU has been introduced as an alternative approach to address challenges related to long-term memory retention and gradient propagation in backpropagation. The input/output configuration of GRU remains consistent with that of the standard RNN. Input $\mathbf{x}^t$ at time t, and the hidden layer state $\mathbf{h}^{t-1}$ at time t-1. The hidden layer state contains information about the previous node. Output: Hidden node output $\mathbf{y}^t$ at time t, and hidden state $\mathbf{h}^t$ passed to the next node.

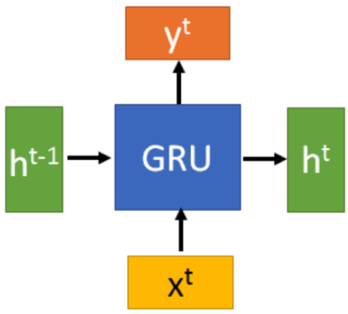


Figure 1: Input/output structure of GRU [6].

2.2.3. Random Forest

In the context of a Random Forest shown in Figure 2, multiple classification trees are employed. To classify a given input sample, the sample is subjected to classification by each individual tree within the forest. Each decision tree functions as a separate classifier, resulting in N classification outcomes for an input sample when N trees are utilized. Subsequently, the random forest combines these classification results through a voting mechanism, aggregating the votes from all trees. The category that receives the highest number of votes is then assigned as the final output or classification label for the input sample.

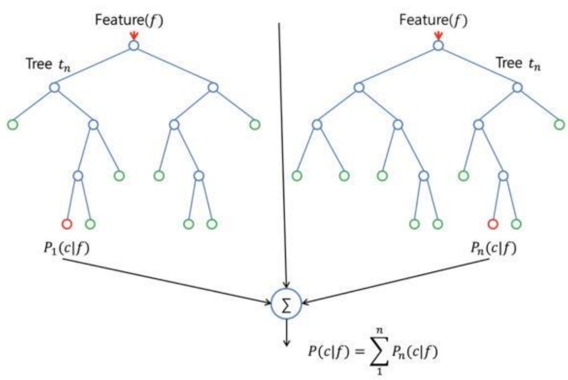


Figure 2: Introduction of Random Forest algorithm [7].

2.2.4. Decision Tree

A decision tree shown in Figure 3 is a hierarchical structure wherein internal nodes correspond to attribute tests, branches represent the outcomes of the tests, and leaf nodes symbolize distinct categories [8]. Prominent decision tree algorithms encompass C4.5, ID3, and CART. The ID3

DOI: 10.54254/2754-1169/46/20230357

algorithm employs information gain, the C4.5 algorithm utilizes information gain rate, while the CART algorithm relies on Gini coefficients for attribute selection during the construction of decision trees.

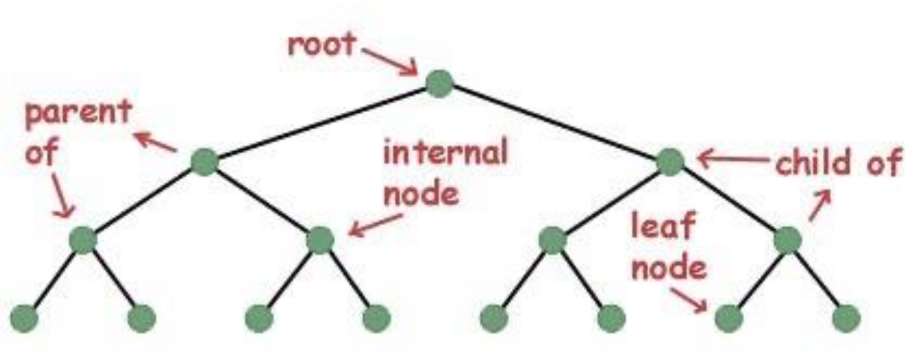


Figure 3: Decision Tree [7].

2.2.5. KNN

The K-Nearest Neighbor (KNN) regression technique demonstrates a close association with the KNN classifier [9, 10]. In the context of KNN regression, when provided with a designated value for K and a prediction point x0, the algorithm initially identifies K training observations that exhibit the closest proximity to x0, denoted as N0. Subsequently, it estimates the target value f(x0) by calculating the average of all training responses present within N0.

3. Results and Discussion

Upon comparing the evaluation outcomes of various models shown in Table 1, Table 2 and Figure 4, it is evident that both the GRU and LSTM algorithms exhibit commendable performance. Notably, the LSTM algorithm stands out as the most superior, as indicated by its smallest Root Mean Square Error (RMSE) value. This aligns with the findings from the existing literature reviewed in this study, affirming the success of the model fitting process. It is important to acknowledge that fluctuations in the stock market do not always adhere to a regular pattern or consistent cycle. The presence and duration of trends vary depending on the specific companies and sectors. Consequently, analyzing these dynamic trends and cycles can yield significant benefits for investors. In the case of analyzing high-volatility stocks such as Tesla, employing an integrated algorithmic approach like LSTM proves advantageous, given its robust resistance to interference and overfitting.

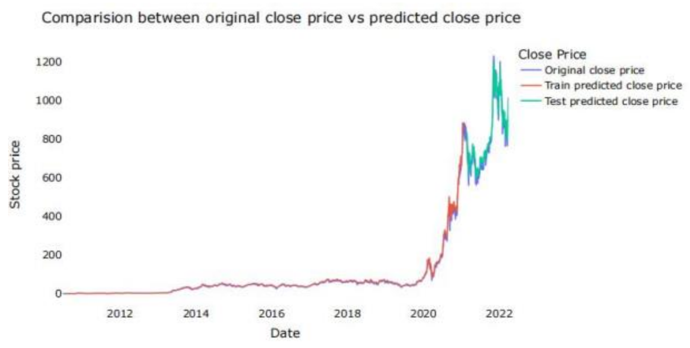
Table 1: The performance of various models evaluated by various indicators in training dataset.

	MSE	RMSE	MAE	R^2
LSTM	49.96565487212373	7.068638827392706	3.024335190810788	0.9955866953389068
GRU	34.18435586937707	5.846738908945488	2.191413057417963	0.9969806064289379
Random Forest	4.9311255947384085	2.2206137878384906	0.7429567787405437	0.9995644496279016
KNN	66.13872797714329	8.132572039468405	2.5815946407463444	0.9941581801097706

DOI: 10.54254/2754-1169/46/20230357

Table 2: The performance of various models evaluated by various indicators in testing dataset.

	MSE	RMSE	MAE	R^2
LSTM	1469.3775506554175	38.33246079572009	30.796759247460937	0.9478842683692899
GRU	1531.802771523471	39.13825202437471	27.715774746707595	0.9456701770649167
Random Forest	16207.943591806286	127.31042216490481	98.02034167439282	0.4251383259942344
KNN	21457.113836062654	146.4824693813654	146.4824693813654	0.23896129639993402

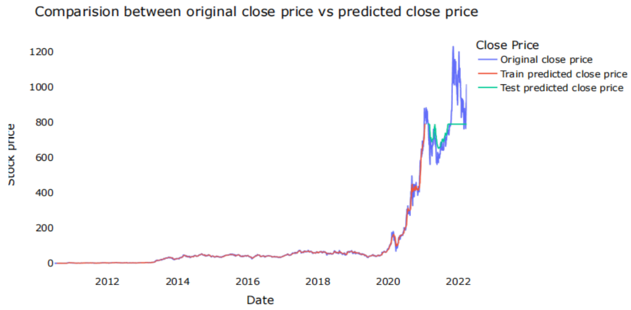


(a) GRU

(b) LSTM

(c) RF

DOI: 10.54254/2754-1169/46/20230357



(d) KNN

Figure 4: The performance of different models (Photo/Picture credit: Original).

4. Conclusion

This study used the Tesla stock dataset on Kaggle to predict the trend of Tesla stock in the special period and obtain ideal returns through investment. LSTM, GRU, Random Forest and KNN models were used to fit the dataset and predicted the trend on this dataset. After carrying out the experiments, it was found that GRU and LSTM models perform well, while Random Forests and K-Nearest Neighbor models perform poorly. In addition, LSTM model had the highest accuracy. Therefore, future prediction work should focus on optimizing the LSTM model, or just try to learn and use more model methods to further compare and predict stock closing prices.

References

[1] *Nelson, D.M.Q., Pereira, A.C.M., and de Oliveira, R.A. (2017) Stock market' s price movement prediction with LSTM neural networks, in 2017 International Joint Conference on Neural Networks, pp. 1419-1426.*

[2] *Bao, W., Yue, J. and Rao, Y.L. (2017) A deep learning framework for financial time series using stacked autoencoders and long-short term memory, Plos One, Vol. 12 No. 7, p. e0180944.*

[3] *Vo, N.N.Y., He, X., Liu, S. and Xu, G. (2019) Deep learning for decision making and the optimization of socially responsible investments and portfolio, Decision Support Systems, Vol. 124, UNSP 113097.*

[4] *Peng, Y., Liu, Y. and Zhang, R. (2019) Modeling and analysis of stock price forecast based on LSTM, Computer Engineering and Application, Vol. 55, pp. 209-212. (in Chinese).*

[5] *Kaggle (2022) TESLA Stock Data https://www.kaggle.com/datasets/varpit94/tesla-stock-data-updated-till-28jun2021?resource=download*

[6] *Cheng, C. (2020), A GRU that everyone can understand. Retrieved from https://zhuanlan.zhihu.com/p/32481747.*

[7] *CSDN (2018), An introduction to Random Forest, Retrieved from http://t.csdn.cn/OOqGk.*

[8] *Myles, A. J., Feudale, R. N., Liu, Y., et al. (2004) An introduction to decision tree modeling. Journal of Chemometrics: A Journal of the Chemometrics Society,18(6): 275-285.*

[9] *Qiu, Y., Chen, P., Lin, Z., et al. (2020) Clustering Analysis for Silent Telecom Customers Based on K-means++, 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC). IEEE, 2020, 1: 1023-1027.*

[10] Myles, A. J., Feudale, R. N., Liu, Y., et al. (2004) *An introduction to decision tree modeling. Journal of Chemometrics: A Journal of the Chemometrics Society, 18(6): 275-285.*