Stock forecasting based on LSTM model in the context of COVID-19 epidemic

Luoxi Lian[1, *]

[1]Department of Computer Science, Shanghai University, Shanghai, China

*Corresponding author: Bubble_llx@shu.edu.cn

Abstract. The COVID-19 pandemic has exploded at the beginning of 2020, which lead to great impacts on a wide range of areas. In addition to the change of mankind's daily behaviors, the global economy is also hit severely by the pandemic. In this context, it is of great significance for the stock market and the investors to conduct in-depth research on quantitative investment considering COVID-19 factor. This study selected 5 representative stocks in the market as the research objects and construct the LSTM model to forecast the stock price as well as analyze the correlation between stock price and COVID-19 pandemic. According to the analysis, it can be obviously seen that the pandemic has had a big impact on the stock market and the price of different stocks fluctuated during the whole period. Moreover, the model was successfully built and output desirable predictions about the stock. These results shed light on guiding further exploration of stock market forecasting and the investment during the pandemic time.

Keywords: LSTM model; COVID-19 pandemic; Stock price forecasting.

1. Introduction

The ups and downs of the stock market are usually related to the country's economy; hence the circumstance of the stock market can be described as the weathervane of the country's economic development. If one can predict the stocks' fluctuation trend with models, the possible economic risks can be avoided in advance.

At present investment market, there are many methods to predict the stock price through different types of models including Fuzzy Neural Networks [1, 2]. Quantile Regression Neural Network is also used in previous study [3], which builds a four-layer neural network and predict eight financial time series. Multi-layer perceptron Neural Network is also called Feedforward Neural Network and Artificial Neural Network, which is set up to the forecasting model based on various basic factors of the combination of price series [4]. The decision tree model is also commonly used in stock forecasting. Scholars classifies and make predictions of the data sets through data mining [5]. As the decision tree has the ability to study itself, it can conduct supervised learning to enhance the accuracy.

In 2019, the stock price development trend of most industries in China is showing an increasing trend, such as high-tech industry, manufacturing industry, construction industry, etc. Companies in all areas had different levels of development and were expected to have a broad space for development. Nevertheless, the sudden outbreak of COVID-19 pandemic at the beginning of 2020 affected not only Chinese economy but also the global economy. The impact of COVID-19 on stocks have been discussed and the solutions of the situation have been given [6]. The outbreak has undoubtedly led to a serious impact on food and tourism industry. On this occasion, consumers have reduced their consumption in these industries because of their own safety considerations. It can be obviously seen in other studies that the recurring pandemic have given these industries no respite [7, 8]. They have been completely hit but are still actively looking for measures to rescue the industry market.

On the other hand, not all the industries have suffered only negative effects. Since the outbreak began, demand for medical products (e.g., masks) has increased dramatically, while people and countries have also pinned their hopes on many medical research companies to develop vaccines and tests for virus. Different from other industries, under such hopes, many medical-related industries have developed rapidly. Besides, they have been standing in the front line of fighting against the epidemic. There's no doubt that their stock price went up. Li and Zhang referred to the volatility of

the US stock medical industry after the outbreak, which was of positive significance for seeking investment opportunities, controlling risks and pricing financial products in the future [9].

With the rapid development of China's market economy, citizens gradually have the consciousness of financial management and are keen to focus on stock trading. Investors also combine computer technology with stock investment for quantitative analysis with the increasing maturity of computer technology, which has become a very mainstream investment method at present. In order to form a more rational investment environment, it is necessary to provide investors with more reasonable investment strategies during the pandemic when the economic situation is less optimistic [10].

In this paper, three stocks of medical related companies and two stocks of catering industry were selected as data sets to construct a LSTM model to predict the stock trend using python language. After the prediction, the volatility trend of each stock price was analyzed under the pandemic circumstance. The reminder sections of the paper are fabricated as follows. The Sec. 2 will introduce the data origination, relevant background and models utilized in this study. Afterwards, the Sec. 3 will demonstrate the empirical evaluation of the models and present the explanation as well. Eventually, a brief summary will be given in Sec. 4.

2. Data & Method

2.1 COVID-19

COVID-19 is widespread, spreading across the country in a short period of time. The epidemic has covered all economically developed areas and strong cities in China. The further implementation of extensive measures (e.g., "city closures" and "road closures") brought economic activity to a standstill. Nowadays, the tertiary industry has accounted for the mainstream(54%) and the impact of a major epidemic outbreak on the tertiary industry is much greater than that of the secondary industry. It is witnessed that the tourism, the film, the catering, the exhibition and other industries are almost completely eliminated.

Meanwhile, China is highly dependent on the international economy at present. As the world's second largest economy, the proportion of export-oriented economy in China is considerably high. The outbreak of COVID-19 caused by the international "shutdown" and "flow restriction" has severely affected the export-oriented economy of China. More importantly, anti-globalization is on the rise now and the trade war between China and the United Stated continues to escalate, gradually expanding to the war of science, technology, finance, public opinion and other all-round games. The environment has deteriorated significantly, and the Chinese economy itself is in a downward transition from high speed to medium-high speed. The arrival of COVID-10 has also dealt a further blow to the economy. In a word, the global economic situation is not optimistic under the impact of the pandemic.

2.2 Stock Forecasting

As a matter of fact, the performance of the stock market will be affected by various aspects, including the country's macroeconomic policies, the company's activities, the introduction of laws and regulations, and most importantly, the attention and confidence of shareholders. Changes in stock prices are usually abrupt with nonlinear characteristic due to a combination of factors. Therefore when selecting the prediction model, all aspects should be considered. It is not enough to simply use the linear or nonlinear model. Only by using the model with good combination of the two, can the modeling result with higher accuracy be obtained. In many experiments, investors have been assumed to be rational in investment forecasting, but this is often not the case in the real world. The psychological factors of investors are prone to great fluctuations and become an unstable factor affecting stock prices. Herd mentality is common among investors. Information including publicity reports or hints of stock market price fluctuations will affect investors' mentality, thus making unpredictable stock selling behaviors, which to a large extent increases the risk of stock price

fluctuations. Especially with the impact of COVID-19 epidemic, the daily growth of pandemic will be linked to the stock market.

2.3 Stock Price Prediction Model

2.3.1 Recurrent Neural Networks

The human brain is a highly complex nonlinear dynamic learning system, which can handle many complex tasks with high efficiency. Inspired by the structure of human brain, scholars have successfully constructed simulated mathematical models of artificial neural networks, which can learn relevant knowledge and content according to the external environment of the neural network and store the learned knowledge. Recurrent Neural Networks are a kind of Neural Networks with short-term memory function. Different from ordinary feed-forward Neural Networks, RNN can receive information from other neurons as well as its own, forming a loop. The running process of RNN model is shown in Figure 1.

The procedures can be given as follows:

$$h_{(t)} = \sigma\left(W_{hx}x_{(t)} + W_{hh}h_{(t-1)} + b_n\right), \quad (1)$$

$$\hat{y}^{(t)} = \sigma\left(W_{yh}h_{(t)} + b_y\right), \quad (2)$$

where W_{hx} is the weight matrix between input layer and hidden layer, W_{hh} is the weight matrix between the hidden layer and itself at adjacent time steps, W_{yh} is the weight matrix between hidden layer and output layer. The vectors b_h and b_y are the bias parameters. The two equations above specify the necessary calculations for RNN at every step. The advantage of RNN is that it can combine time series in neural networks and retain the temporal nature of data. However, RNN is only effective in short-term memory. With the increase of sequence length and model length, the possibility of gradient disappearance and explosion will arise when calculating the gradient. In this case, it is necessary to use more long-term memory neural network model to solve such problems.

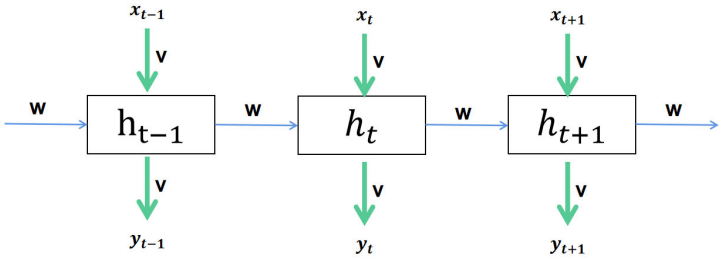

Fig. 1 A sketch of the RNN model implemented in this study.

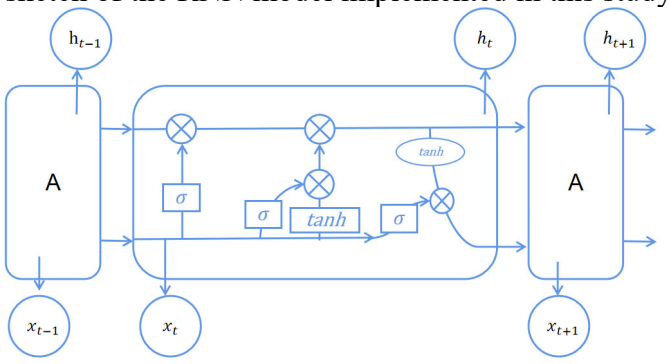

Fig. 2 A sketch of the LSTM cell.

2.3.2 LSTM

LSTM is an improved RNN which contain gates to fulfill the function of learning remote dependencies than the ordinary RNN. The model of LSTM cell is showed in Fig. 2. The information transfer process of the whole model can be described as the following steps: Input data x_t and hidden layer output h_{t-1} pass through the forgetting gate, and expression of state f_t is as follows:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right), \quad (3)$$

Here, σ is the sigmoid function, W_f is the weight vector of the forgetting gate, and b_f is the bias value of the forgetting gate. Input data x_t and hidden layer output h_{t-1} pass through the input gate, and the internal state and vector expression to be input are obtained as follows:

$$\hat{C}_t = tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (4)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \quad (5)$$

Here, $\hat{C}_t$ represents the value of the candidate state, W_c is the weight of the internal state after change, and b_c is the bias value of the internal state. i_t, W_i and b_i are the output vector, the updated weight and the offset value respectively. Input data x_t and hidden state h_{t-1} pass through the output gate, and the output result o_t is calculated. The information of o_t in the internal state C_t screening at time *t* is used as the final output h_t, and the expression is as follows:

$$C_t = f_t C_{t-1} + i_t \hat{C}_t, \quad (6)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \quad (7)$$

$$h_t = o_t tanh(C_t) \quad (8)$$

where o_t is the output vector, W_o is the weight matrix, b_o is the bias value, h_t is the output value of external state at time *t*.

2.4 Data Sources and Data Preprocessing

The daily stock line data in this paper comes from the Tushare open source library in Python and the data set includes 9 indicators (e.g., open price, high price and amount) as listed in Table 1. The stock from March 2, 2020 to June 30, 2022 are selected. The epidemic heat index is derived from Baidu index data analysis platform. Based on the behavior data of Internet users, the website obtains the daily search heat since the outbreak of the epidemic in 2020 through crawlers. This model mainly uses the historical stock trading data of the previous day and the epidemic heat index to predict the highest price of stocks on the next day.

Table 1 Data Indicators

Name	Description	Indicator	Name	Description	Indicator
Open	Opening Price	x_1	Amount	Turnover	x_6
Close	Closing Price	x_2	Change	Rise and Fall	x_7
Low	Minimum Price	x_3	Data	Epidemic Heat Index	x_8
High	Maximum Price	x_4	T_high	Maximum Price the Next Day	x_9
Volume	Number of Transactions	x_5			

In this model, the training set and the test set will be divided by a ratio of 9:1 before input. First of all, as the value range of the 9 input data of each data sample is inconsistent, it is necessary to normalize the data set. Data normalization can unify all data measurement units, which has a great impact on improving the accuracy of the model. After checking the data set of this experiment, it was found that there were no abnormalities or omissions. Thus, only normalization process was needed. The formula for data normalization is as follows:

$$x_i^{'} = \frac{x_i - \bar{x}}{\max(x) - \min(x)} \quad (9)$$

Here, x_i is the i^{th} variable, $\bar{x}$ is the average value of x_i, respectively.

In this paper, three medical stocks and two catering stocks are selected for analysis. Among them, the medical companies were all involved in the research and development or production of COVID-19 related antigen detection products during the COVID-19 pandemic. The selected catering enterprises are all representative enterprises focusing on the development of offline stores. The process of stock prediction in this experiment is given in Fig. 3.

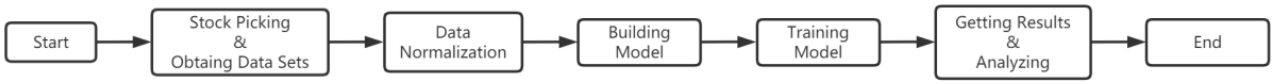

Fig. 3 The procedure of the paper

3. Experimental Analysis

The output result of the model is the daily highest price of the stock. Meanwhile, according to the result, one can calculate the return rate of each stock under the model prediction. The calculation formula is as follows (taking $P_0 = 100\%$ in the initial period of investment as an example):

$$pre_{c_i} = \frac{P_{i+1}}{P_i} \tag{10}$$

$$\mathrm{P_i} = \mathrm{P_{i-1}} * \mathrm{true_{c_i}} \tag{11}$$

$$\mathrm{n} = \frac{\mathrm{P_j} - \mathrm{P_i}}{\mathrm{P_i}} * 100\% \tag{12}$$

where $\mathrm{P_i}$ is the predicted daily high price, pre_c_i is the rise and fall calculated by the model, $\mathrm{true\_c_i}$ is the real rise and fall, n is the calculated return rate within the period. According to the daily quotation of the stock, the maximum draw-down rate can be calculated:

$$\mathrm{MaxDrawdown} = \frac{\mathrm{Max}(\mathrm{P_i} - \mathrm{P_j})}{\mathrm{P_i}} \tag{13}$$

Here, P is the net value of a day, i is a day, j is a day after i, $\mathrm{P_i}$ is the net value of the product on the day i, and $\mathrm{P_j}$ is the net value of a day after day i.

The prediction results of the model are given in Figs. 4-8, where the true value and predicted value from the model are marer in the blue and red color, respectively. The left panels of each figure show the full period curve and each right counterpart shows the trend over the last two months. Nearly all the predictions have a high accuracy rate with the exception of the Quanjude Roast Duck whose figures are less accurate.

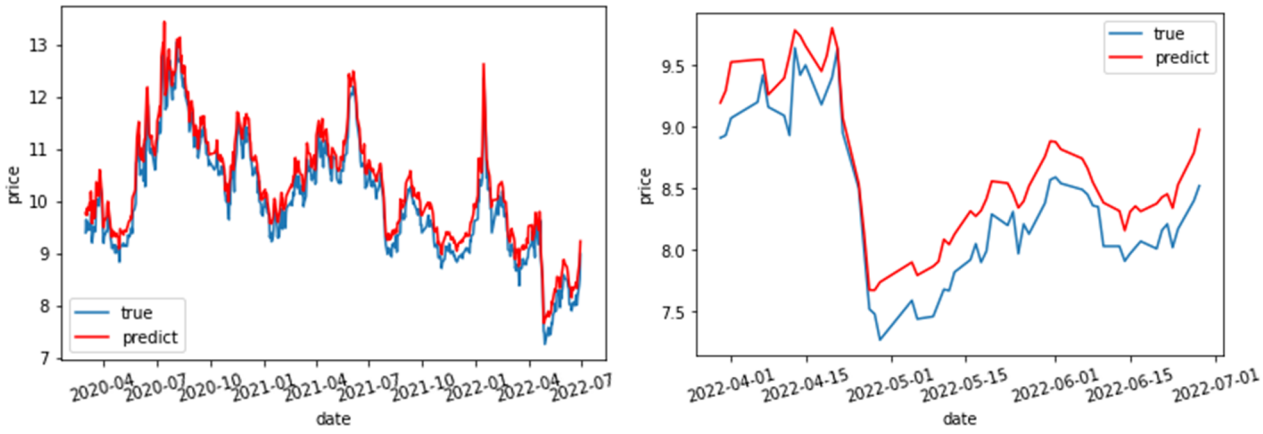

Fig. 4 The forecasting and real price of China Quanjude Roast Duck during the chosen period.

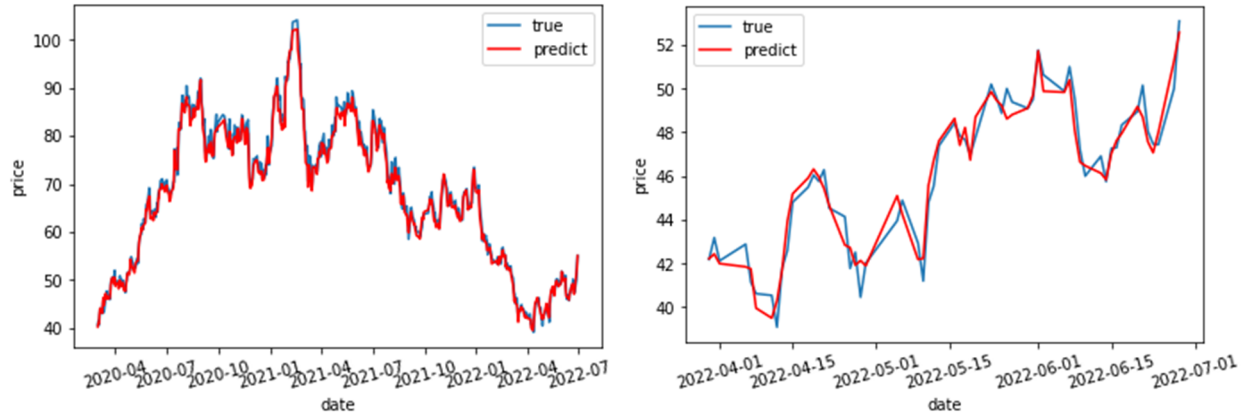

Fig. 5 The forecasting and real price of Juewei Duck Neck during the chosen period.

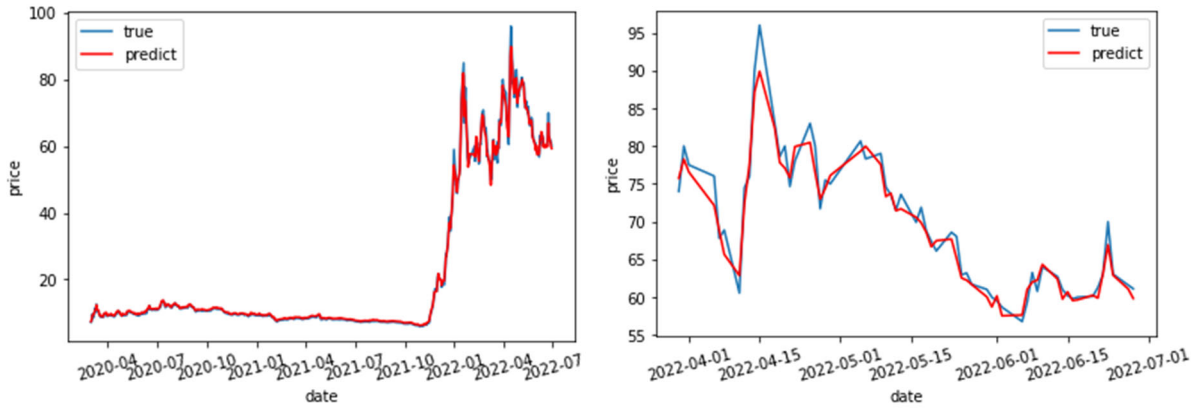

Fig. 6 The forecasting and real price of Andon Medical during the chosen period.

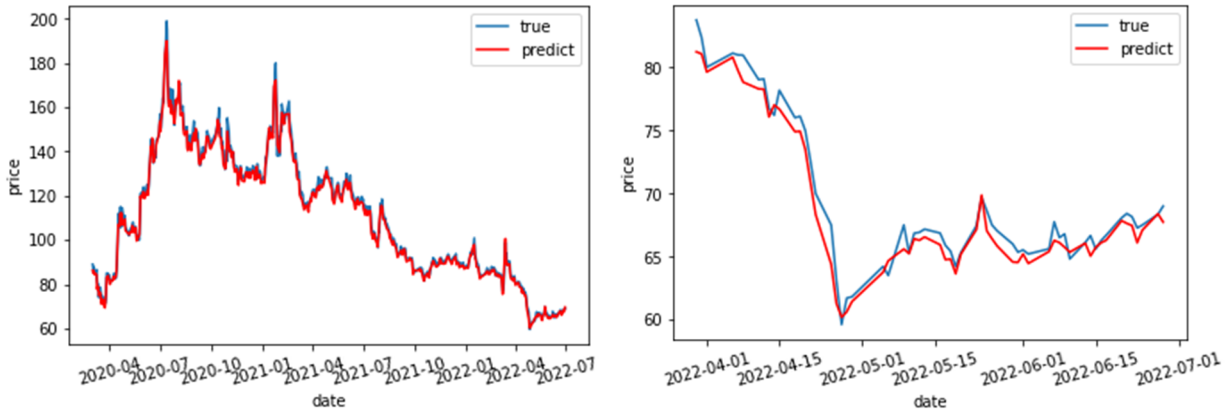

Fig. 7 The forecasting and real price of BGI gene during the chosen period.

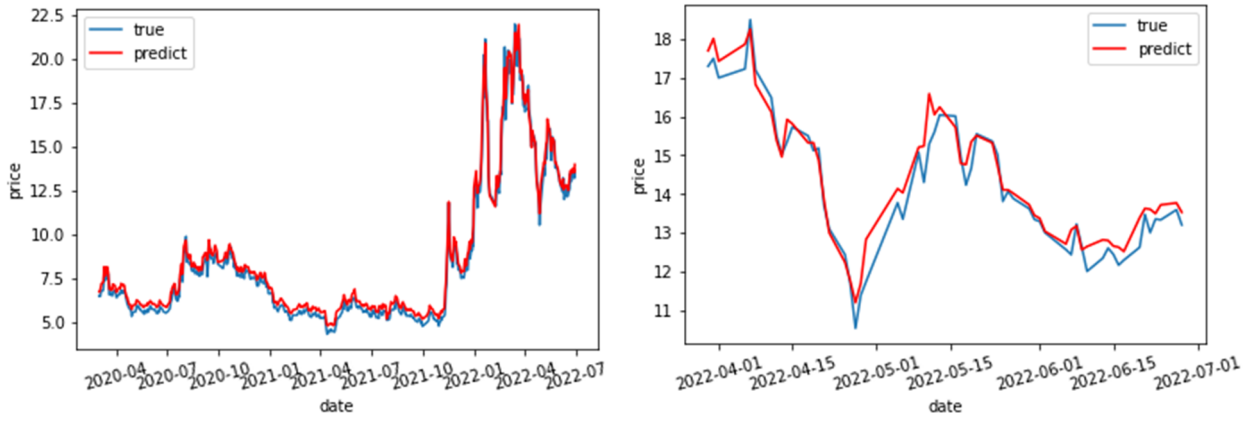

Fig. 8 The forecasting and real price of Hybio Pharmaceutical during the chosen period.

Taking the rise and fall as the boundary to distinguish, if the forecast result (no matter in the case of rise or fall) is greater than 0%, then one needs to buy the stock; otherwise, if it is less than 0, then the trader ought to sell the stock. The forecast rate of return can be calculated and obatined as listed in Table. 2. According to the results presented in the table, it clearly illustrates that Andon Medical and Hybio Pharmaceutical both are getting high rates of return and other stocks have also done well.

Table 2 Return Rates and Maximum Draw-down Rates

Stock	Return Rate	Maximum Draw-down Rate
China Quanjude Roast Duck	54.09%	24.53%
Juewei Duck Neck	82.52%	28.92%
Andon Medical	173.09%	29.99%
BGI gene	39.36%	15.53%
Hybio Pharmaceutical	116.54%	30.60%

4. Conclusion

In conclusion, this paper investigates the stock market and make some analysis based on LSTM models. According to the forecast charts from the model, one can see the general development trend of enterprise stocks. The medical industries and the catering industries have all been greatly affected by the epidemic. In particular, in March 2022, the COVID-19 epidemic has made new changes that has brought a great impact on the real industry, which can be clearly seen in the stock fluctuations. Based on LSTM model, the stock price prediction effect is good, the degree of fit is high, and can obtain an objective return rate during the period. Nevertheless, the reason for the current high returns is to use the previous day's stock data each time to predict the next day's price. In the future, the data of the previous month can be utilized to predict the returns of the next month and improve the accuracy. Overall, these results offer a guideline for stock price forecasting.

References

[1] Li R and Xiong Z. Forecasting stock market with fuzzy neural networks, 2005 International Conference on Machine Learning and Cybernetics, 2005, pp. 3475-3479 Vol. 6.

[2] Gholamreza J, et al. Application of Fuzzy-neural networks in multi-ahead forecast of stock price. African Journal of Business Management, 2010, 4(6): 903-914.

[3] Pradeepkumar D, Ravi V. Forecasting financial time series volatility using particle swarm optimization trained quantile regression neural network. Applied Soft Computing, 2017, 58: 35-52.

[4] Samarawickrama A. J. P. and Fernando T. G. I. A recurrent neural network approach in predicting daily stock prices an application to the Sri Lankan stock market. 2017 IEEE International Conference on Industrial and Information Systems (ICIIS), 2017, pp. 1-6.

[5] Yu Wang, et al. Stock forecasting based on Cart Decision tree and Boosting. Journal of Harbin University of Science and Technology, 2019,24(06):98-103.

[6] Zeren F, Hizarci A. The impact of COVID-19 coronavirus on stock markets: evidence from selected countries. Muhasebe ve Finans İncelemeleri Dergisi, 2020, 3(1): 78-84.

[7] Nakat Z, Bou-Mitri C. COVID-19 and the food industry: Readiness assessment. Food control, 2021, 121: 107661.

[8] Yeh S S. Tourism recovery strategy against COVID-19 pandemic. Tourism Recreation Research, 2021, 46(2): 188-194.

[9] Shuai Li, Qiang Zhang. An empirical study on the impact of COVID-19 on the returns of the US stock healthcare industry. China's prices, 2021(09):83-86.

[10] Singh A. COVID-19 and safer investment bets. Finance research letters, 2020, 36: 101729.