

## A Machine Learning Based on the Impacts of COVID-19 on Indian Stock Market

Yuankai Zhu<sup>1,\*</sup>, Kailong Miao<sup>2,a</sup>, Chenyu Wang<sup>3,b</sup>, Di Wang<sup>4,c</sup>, Yifan Wei<sup>5,d</sup> and Ningxin Zheng<sup>6,e</sup>

<sup>1</sup>Wenzhou-Kean University, Wenzhou, Zhejiang, China

<sup>2</sup>Hohai University, Nanjing, Jiangsu, China

<sup>3</sup>Pennsylvania State University (University Park), University Park, PA, United States

<sup>4</sup>Chongqing University, Chongqing, China

<sup>5</sup>International Department of Beijing No.35 High School, Beijing, China

<sup>6</sup>University of Toronto, Toronto, Canada

\*Corresponding Author: 1005505432@qq.com, <sup>a</sup>373779363@qq.com,  
<sup>b</sup>18001916596@163.com, <sup>c</sup>1134914871@qq.com, <sup>d</sup>1720947927@qq.com,  
<sup>e</sup>ningxin.zheng@mail.utoronto.ca

These authors contributed equally to this work

### Abstract

**The main purpose of this study is to explore the impact of COVID-19 on the Indian stock market regarding to the transportation. It uses different types of regression models and machine learning to analyze and predict the impact of COVID-19 on the Indian stock market regarding to the transportation. According to the data analysis of the visualization graph, after the arrival of the epidemic, the number of people in the stock market increased substantially, but the stock market price decreased significantly, and then gradually recovered, but still did not return to the earliest price.**

### Keywords

**Stock Market; Machine Learning; COVID-19.**

### 1. Introduction

As the COVID-19 spread to a global pandemic, equities plummeted and market volatility exacerbated increased around the world. From the perspective of measuring the impacts of COVID-19 on financial market, the most visible expressional form is the price trend on the stock market. Between February 20 and March 19, the S&P 500 index fell by 28% (from 3,373 to 2,409), the FTSE 250 index fell by 41.3% (from 21,866 to 12,830), and the Nikkei fell by 29% (from 23,479 to 16,552). Global stock markets lost \$6 trillion in value over six days from 23 to 28 February. According to S&P Dow Jones Indices [1]. All these fierce volatility on stock market shows the necessity of research in this topic.

There are so many literature articles focused on the impact of COVID-19 on stock market, and many methods were used to predict the stock trends affected by the pandemic. While standing at the point of view in analysis and forecasting, three mainstream methods are fundamental analysis, technical analysis, and evolutionary analysis. The latter two methods are mainly applied to the time and space judgments of specific investment operations, as an important supplement to improve the effectiveness and reliability of securities investment analysis. Recent years have seen the wide use of time series forecasting (TSF) for predicting the future

price stock, modeling, and analyzing of the stock market. Most time series forecasting techniques typically use linear methods, such as AR, MA, ARMA, ARIMA, etc. or nonlinear models, such as ARCH, GARCH, etc. The authors of [1] investigated how the coronavirus outbreak led to spillovers into the major sectors of the global economy, and the effect of social distancing policies on the level of economic activities and stock index prices. Giridhar Maji et al., [2] proposed an association rule mining with Apriori studying index values of 6 industrial sectors from NSE, employing many different techniques for knowledge discovery and prediction such as classification, clustering, sequential pattern mining, association rule mining and analysis. Researcher from [3] has presented efficient stock forecasting model RNN-LBL by combining recurrent neural network (RNN) with log bilinear (LBL) for modeling both long and short-term behavior pattern of stock market, respectively, along with the assessment methods in terms of Root Mean Squared Relative Error (RMSRE) and Direction Prediction Accuracy (DPA) to evaluate the performance of proposed RNN-LBL model. Milind Yadav et al., [4] used the Support Vector Regression (SVR) model to solve the four different types of COVID-19 related problems. And compared the performance results with the three different well-known regression methods such as Simple Linear Regression, Polynomial Regression, the Support Vector Regression method based on coronavirus analysis given promising results.

Murtaza Roondiwala et al., [5], they presented a recurrent neural network (RNN) and Long Short-Term Memory (LSTM) approach to predict the stock returns of NIFTY 50. And this work establishes the fact that using recurrent neural networks (RNN) for processing sequential data along with LSTM for introducing the memory cell makes the model suit to grasp the structure of data dynamically over time with high prediction capacity.

Corona Virus Disease 2019 known as COVID-19, caused by SARS-CoV-2 initially appeared in Chinese south Central city Wuhan. Since December 2019, novel coronavirus infection cases of acute respiratory tract infections have been identified in some hospitals in Wuhan, Hubei, and have been found in many cases of unknown pneumonia with the exposure history of Southern China seafood market. The novel coronavirus pneumonia was named "COVID-19" in February 11, 2020 by WHO director general Tan de Se in Geneva, Switzerland. Within a short span of a month, WHO declared the outbreak as a public health emergency of international concern [1]. With its exponential spread, more than 200 countries has been affected, and US (United States) has the most confirmed cases (more than 6.8 million up as of September 19, 2020) followed by India, which has more than 5 million cases with extremely rapid growth rate. Global death toll has reached to 951 thousand and number of confirmed cases are 30 million as of 19th September, 2020 [2]. Till now, over 400 vaccines and treatments are under study as researchers rush to produce effective medicines for the disease, according to reports. While health organizations say no medicines are yet available to prevent or cure the coronavirus, some home remedies may alleviate mild symptoms. In China, 25 new coronal vaccines have entered the clinical trial or preclinical stage, of which 11 vaccines have entered the clinical stage and 4 vaccines have entered the clinical phase III. This deadly virus that has forced various countries to adopt social distancing, travel restrictions, border shutdowns and many other preventive measures [3].

As a result of these restrictions, the labor force in all sectors has been reduced, which in turn has resulted in many people losing their jobs. Lockdown has also tremendously affected the economies of various countries. The outbreak of this virus has escalated the concerns of all agencies worldwide, it is nearly impossible for any individual or enterprise to survive while not be affected by the this disease. The stock market has been volatile recently, which could affect a lot of people's investments in pensions or personal savings accounts (ISA). Since the outbreak of the epidemic on December 31 last year, the FTSE index, the Dow Jones Industrial Average and the Nikkei index have all fallen sharply. Among them, the Dow and the FTSE both recorded the biggest one-day decline since 1987. Many Investors worry that the spread of the new

coronavirus could destroy economic growth and that government action may not be enough to stop the stock market from falling. To this end, many countries' central banks have taken measures to reduce interest rates in response. In theory, a central bank cut in interest rates should boost the economy by reducing borrowing costs and encouraging consumption. So far, the US Federal Reserve and the Bank of England and other major global central banks have cut interest rates substantially. Withdrawal of money from the markets by the foreign investors is continuous [4]. The figure I below shows central bank interest rates change for the top five worst affected countries [5]. Panic selling by the investors has made decision making more difficult. The uncertainty in prediction of economy and health is rising due to continuous increase in number of cases and deaths. The figure I below shows the number of confirmed cases for the top five worst affected countries. This study attempts to measure impact of COVID-19 on stock market outcomes. The paper first summarizes the existing literature on impact of pandemics on stock markets, then explores and analyzes the current available data and attempts to correlate the effect Indian stock market, in particular, the transportation sector. It uses different types of regression models and machine learning to analyze and predict the impact of COVID-19 on Indian stock market regarding to the transportation.

## 2. LSTM for Close Price Prediction

In 1997, long short-term memory (LSTM) neural network was proposed by Hochreiter and Schmidhuber to alleviate the problem that traditional RNN had less ability to improve its prediction ability due to the disappearance of the gradient. LSTM is one of time recursive neural networks. Its basic idea is based on generating the path through time, so that the derivative neither disappears nor explodes. This important feature makes the LSTM suitable for processing and predicting important events with long intervals and delays in time series. The most important structure of the LSTM model is composed by three gates: forget gate  $F$  determines which information is filtered out by the cell; Input gate  $I$  determines which values pass through the input gate to update the state of memory; The output gate  $O(t)$  determines which parts of the input and cellular memory are output. We can use the following formula to express the calculation principle of LSTM:

$$I_i^{(t)} = \sigma \left( \sum_j W_{i,j}^I x_j^{(t)} + \sum_j U_{i,j}^I h_j^{(t-1)} + b_i^I \right)$$

Where  $x_j$  is the input value at the current time  $t$ ,  $h_j$  is the input value of the current hidden layer, and  $b$ ,  $W$  and  $U$  are the bias amount, input weight and output weight in LSTM cells.

$$F_i^{(t)} = \sigma \left( \sum_j W_{i,j}^F x_j^{(t)} + \sum_j U_{i,j}^F h_j^{(t-1)} + b_i^F \right)$$

Then we update state value at current time  $\tilde{C}_i^{(t)}$  and state of cell in LSTM  $S_i^{(t)}$ :

$$\tilde{C}_i^{(t)} = \tan h \left( \sum_j W_{i,j}^c x_j^{(t)} + \sum_j U_{i,j}^c h_j^{(t-1)} + b_i^c \right)$$

$$S_i^{(t)} = I_i^{(t)} \tilde{C}_i^{(t)} + F_i^{(t)} S_i^{(t-1)}$$

Then the output gate  $O_i^{(t)}$  read above value and calculate result  $h_i^{(t)}$

$$O_i^{(t)} = \sigma \left( \sum_j W_{i,j}^O x_j^{(t)} + \sum_j U_{i,j}^O h_j^{(t-1)} + b_i^O \right)$$

$$h_i^{(t)} = \tan h \left( S_i^{(t)} \right) O_i^{(t)}$$

Where  $b$ ,  $W$  and  $U$  are the output bias, input weight and recycle weight. With the above calculation, the LSTM can effectively utilize the input to enable its long-term memory function.

### 3. Data Processing

We choose close price of one stock which can most represent the stock trend in each sector, and then divide each stock data to two parts, training set and testing set. The training set includes 66.67% of whole data set which from around

January 1 2020. The validation set includes last 33.33% of whole data end in July 29 2020. The close price is already pre-processed using the interpolation we mentioned in previous section so that we can ensure there are no null value in data set. We choose Minimum-maximum scaler to normalize the dataset to improve the training results and transform the dataset to matrix whose time step =1 and each row includes close price from time t-3 to t, where 3 is the wide of time window we chosen. Then we divide the matrix into two sub-matrixes. One is training/validation feature matrix which includes close price from time t-3 to t-1 and another is the training/validation label matrix which only includes close price of time t.

### 4. Model Building

We build a sequence LSTM model which starts with a LSTM layer with units 3 and ends with a dense layer with unit 1. The model is compile by the Adam optimizer and the loss is calculated by the mean square error. The model diagram is showed in Figure 4.

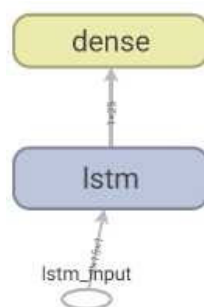


Figure 1. LSTM Model Diagram

### 5. Data Visualization

We import three packages-pandas, sklearn and numpy, for the preparation of the data visualization firstly. At the beginning of every section, we import plotly.graph\_objects additionally.

Covid-19 Daily Data of The Day Before Each Stock Exchange Date

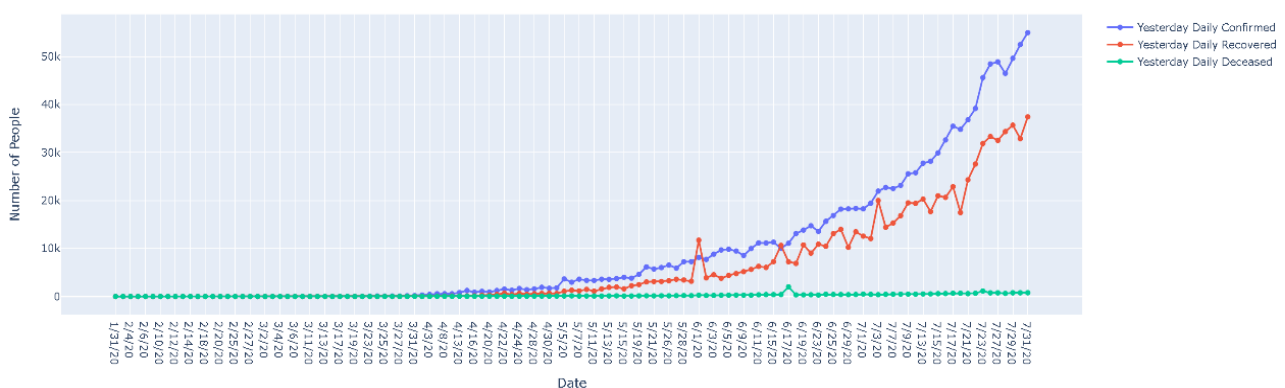
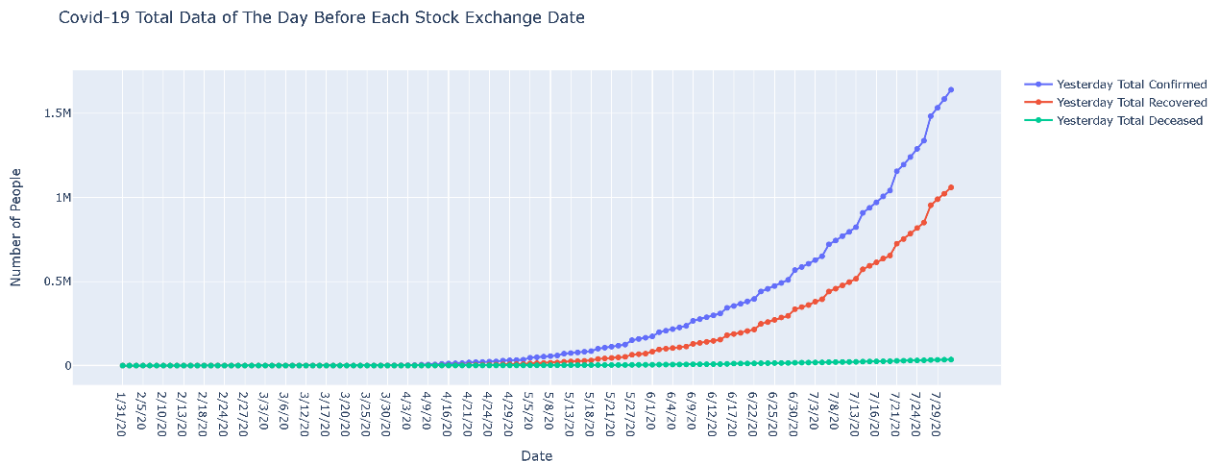
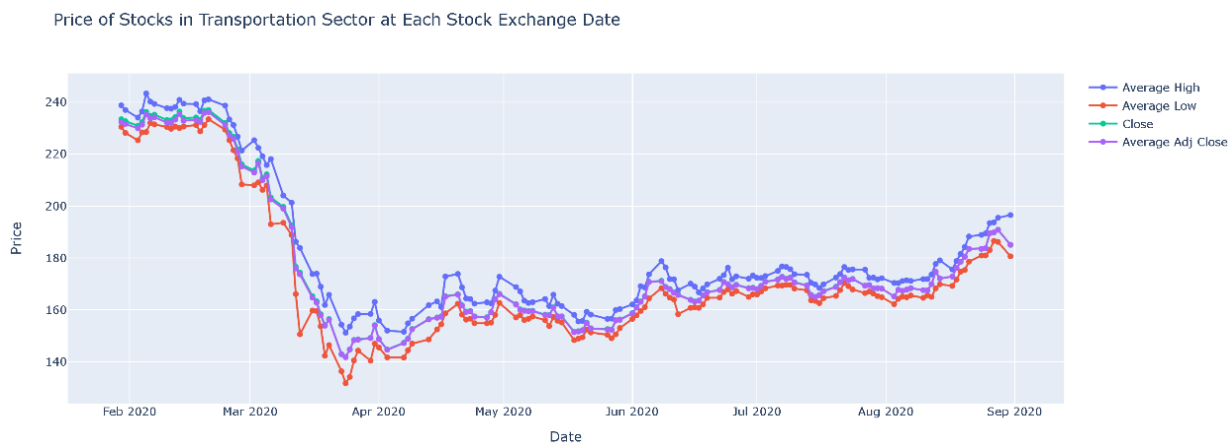


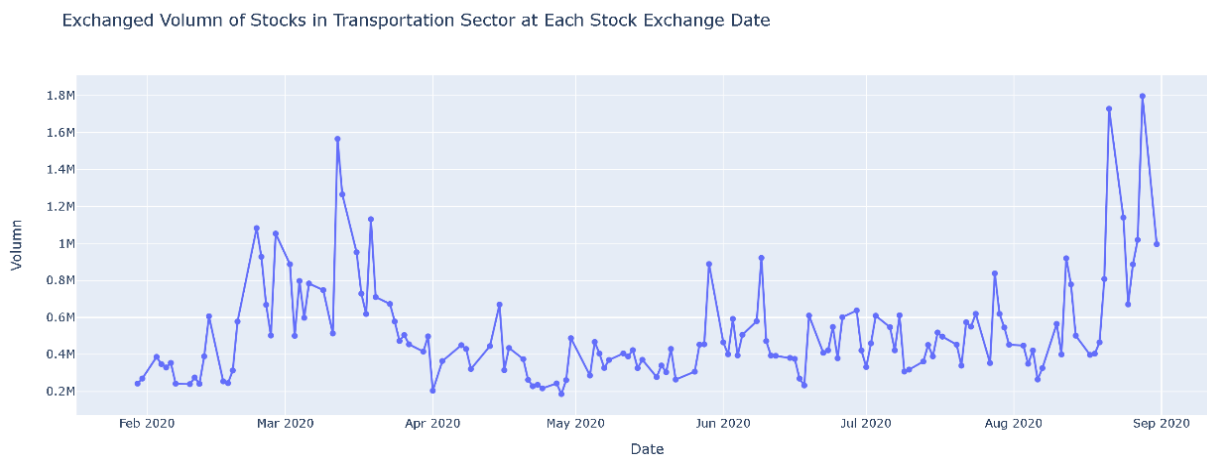
Figure 2. Shows covid-19 daily data of the day before each stock exchange date. It includes three curves. They represent the change of the number of people of yesterday daily confirmed, yesterday daily recovered and yesterday daily deceased in the wake of date respectively



**Figure 3.** Shows covid-19 total data of the day before each stock exchange date and it is composed of three curves. They represent the change of number of people of yesterday total confirmed, yesterday total recovered and yesterday total deceased over time



**Figure 4.** Shows price of stocks in transportation sector at each stock exchange date. It consist of four curves. They represent the stock price of average high, average low, close and average adj close on different dates



**Figure 5.** Shows exchanged volumn of stocks in transportation sector at each stock exchange date

The curve represents the change of exchanged volume of stocks in transportation sector at each stock exchange date from Feb 2020 to Sep 2020.  
 (Model used in this paper)

## 6. Model Prediction and Evaluation

When we make prediction using the model we trained, it is important to make inverse transformation to the dataset, which can secure the accuracy of the mean square error we calculated. In average, the root square error of training prediction and validation prediction is 2.88 and 2.70. The average epoch loss diagram is showed in Figure 4, and the prediction results is showed in Figure 5.

## 7. Conclusion

The world has completely changed over the past 16 months and everything in all sectors of the economy whether banking, transport, health, tourism and the likes have witnessed drastic changes like never seen before. This paper specifically explains the influence of novel coronavirus on the Indian stock market using data collected over the said period. Data is used to observe the relationship between the number of people, the stock price and the market value on the date, and then they are represented by a scatter matrix to judge their dimensions, the distribution of data and their relationship. As already pointed out, at the beginning of the pandemic restrictions that led to people working from home saw a number of new entrants into the stock market greatly due to the availability of online markets and internet connections. However, with time, prices in the stock market rapidly fell pushing out a large number of the new entrants eventually leading to a slight increase in the stocks considering that a number of restrictions have systematically been taken down across the globe. However still, stock markets are yet to get back to their original prices. All these have been proven by the long term short term memory neutral network (LSTM) use in this study's methodology. Basically, the system processes data, builds a model with the same data and gives a visual output in terms of graphs and histograms. Finally, through various mathematical models, we are able to learn and predict the influence of novel Coronavirus on the transport sector of the stock market. Looking at current developments in the global economy, investors are currently worried about the impact of covid-19 on the stock markets considering that no matter how much the government tries, there has been a rapid increase in the number of cases and deaths witnessed in recent months. That's why governments in recent times have been working on interest rate reduction in order to sustain the economy by cutting down cost of loans while also boosting consumption. However still, the economic effectiveness of such initiatives in most countries shows little response as most investors withdraw their funds and undertake panic sales in stock markets. This means that the future of the Transport sector prices in the Indian stock exchange will remain on the low until covid-19 cases and deaths decline and all sectors of the economy are opened up.

## References

- [1] Peterson Ozili, Thankom Arun, United Kingdom, Spillover of COVID-19: impact on the Global Economy.
- [2] Giridhar Maji, Soumya Sen, Amitrajit Sarkar, Share Market Sectoral Indices Movement Forecast with Lagged Correlation and Association Rule Mining.
- [3] Uma Prashant Gurav, S. Kotrappa, RNNLBL: A Recurrent Neural Network and Log Bilinear based Efficient Stock Forecasting Model.
- [4] Milind Yadav, Murukessan Perumal, Dr. M Srinivas, Analysis on novel coronavirus (COVID-19) using machine learning methods.
- [5] Murtaza Roondiwala, Harshal Patel, Shraddha Varma, Predicting Stock Prices Using LSTM.