Predicting the US stock market during the COVID-19 pandemic

Ling Yao, Mengyu Ma, Yujia Yang, Qi Li, Haoquan Lin, Jiahong Hu

1 INTRODUCTION

The Coronavirus disease 2019 (COVID-19) has triggered an unprecedented impact on the global economy and financial markets. From the US equity market standpoint, natural gas, food, healthcare, and software industries show high positive returns, while the oil, real estate, entertainment, and hospitality stocks show a significant fall in response to the impact of COVID-19 [14].

On the other hand, there have been numerous researches and implementations applying various machine learning algorithms or statistical models to equity price prediction. The performance varies, while a key issue is that these algorithm is unable to automatically take in new information or market sentiment when market event happens. Therefore, we are interested in exploring the area combining the information during COVID-19 pandemic and stock price prediction, to reveal the unique patterns of stock market during pandemic.

2 PROBLEM DEFINITION

A question that interests us is, how does each equity sector perform during the pandemic? What could be the key takeaway from the COVID-19 crisis that can help investor's decision making process in the next health care crisis? To answer the question, we collected pandemic related information from CDC, and various control variables from policy rate to high frequency economic indicator, and daily SP 500 equity prices from January 2018 to March 2022, to construct models that are able to predict stock prices in different equity sectors.

3 SURVEY

As can be seen from the human history of flighting against pandemics including Black Death, smallpox and Spanish flu, the development of human civilization is always accompanied by the outbreak of viruses [15]. As such, the outbreak of next pandemic is only a matter of time. In order to help ease the panic of investors in the future pandemic, a tool that can predict the stock price is particularly important. The COVID-19 has triggered an unprecedented impact on the global economy [11]. Herein, our study will focus on predicting the stocks in different sectors in response to the effect of COVID-19 using ensemble machine learning algorithms and comparing the results of the prediction considering and not considering COVID-19 effect.

Since COVID-19 was first reported in Wuhan, China in December 2019, it becomes increasingly prevailing around the world [13]. The COVID-19 has triggered an unprecedented impact on the global economy [11]. Undoubtedly, the US economy cannot stand alone. Chen's study shows that during the pandemics, any intervention involving a stay-home order can cause a significant economic loss [5]. The COVID-19 outbreak has created uncertainty in financial markets and beyond. From an economic and financial standpoint, it has led to a short-term slowdown in trade and investment flows [17]. In the stock market, natural gas, food, healthcare and software industries show high positive returns, while the oil, real estate, entertainment and hospitality stocks show a significant fall in response to the impact of COVID-19 [14]. Kartal's group combines pandemic indicators with additional independent financial market variables and shows that stock market is negatively affected by pandemic [12]. Gurav et al discussed stock market-related technical indicators, mathematical models, the most commonly used algorithms in the data science industry and analysis of various machine learning algorithms [8]. Weng et al used a neural network regression ensemble, a support vector regression ensemble, a boosted regression tree and a random forest regression to form the ensemble machine learning algorithm for the stock prediction. Historical stock prices, several well-known technical indicators, counts and sentiment scores of published news articles for a given stock, trends in Google searches for the given stock ticker and number of unique visitors for pertinent Wikipedia pages are used for data preparation for the prediction and concluded that the online sources improve the prediction performance of machine learning based methods [16]. Eleftheriou's group finds the evidence of COVID-19 impacts the global stock market by analyzing the relationship of Coronavirus Government Response Tracker index and stock markets across 45 countries using a spatial econometric approach [7]. On top of that, fiscal and monetary stimulus plans bring positive abnormal return to equities in some certain regions and/or sectors [9]. Alparslan and Kim used Long Short-Term Memory Networks to predict the extreme volatility of meme stocks [3].

Davis and his colleague track the stock prices to various economic activities from mid-February 2020 to June 2021 and document the pattern of the relationship [6]. Hong used various models to examine the relationship between Covid-19 and the stock market performance [10]. Almehmadi introduced a new way to predict stock price by using social media as the increasing case of Covid-19 changed customer behaviors. [2].

Machine learning also plays an important role in the study of the stock market during the pandemics. Katral et al. uses random forest and support vector machine to find out the factors that affect stock market indices during pandemic in emerging countries [12]. Baker and his colleague use regression method to analyze how stock market and different sectors react to pandemics [4].Alassafi et al. uses deep learning models include recurrent neural network(RNN) and long short-term memory(LSTM) networks to predict the COVID-19 cases [1].

4 PROPOSED METHODS

Since stock's market is a dynamical system which is highly noisy and sensitive, to overcome the challenge and to get higher accuracy on time series forecasting, we apply recurrent neural networks to build the machine learning of stock's market.

4.1 Intuition

A traditional RNNs shown as Figure 1, are constructed by input layer x_t, hidden layer and output layer o_t. For each timestep t, the activation for hidden layer s_t and output can be expressed as following:

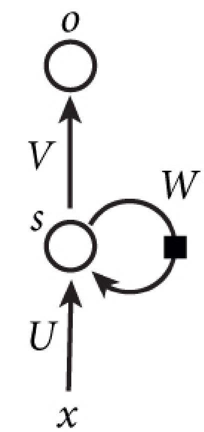

Figure 1: A demonstration for the idea of RNNs

$$o_t = g(Vs_t) \tag{1}$$

$$s_t = f(Ux_t + Ws_{t-1}) \tag{2}$$

Here, U and V are the corresponding weight matrix of input layer and output layer.W is the weight of s_{t-1}. g and f are the activation function. Based on equation 2, the activation s_t depends on the input value on current time t and the activation from previous time $t-1$. Hence, the output is sequential dependence.

4.2 Algorithms

By using the iteration of equation 1 and equation 2, the current output value can be expressed as

$$o_t = Vf(Ux_t + Wf(Ux_{t-1} + Wf(Uxt - 2 + ...)) \tag{3}$$

The output 0_t can be affected from any time's input values. Moreover, in order to avoid gradient Vanishing/exploding , we introduces more gates Γ into the algorithms to form Gated Recurrent Unit (GRU) and Long Short-Term Memory units (LSTM), such that the algorithms can decide to drop the previous information and erase a cell.

The activation functions in our algorithms are the sigmoid and Tanh which are defined as

$$f_{sigmoid} = \frac{1}{1 + e^{-x}} \tag{4}$$

and

$$f_{Tanh} = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{5}$$

4.3 User Interface

The analysis results will be saved as csv files and Angular will be used to build web-based User Interface. Www.netlify.com will be used to deploy the website. At homepage, the user is able to view the composition of each sector of SP500 as a pie chart. The user is able to click the sector on the pie chart, so that they're able to go to the sector page to view the analysis result of the selected sector. The analysis result of SP500 will also show on the home page. A multi-axis timing line chart of the sector's relative return to index and the COVID-19 cases will also be shown on the first page. The user is able to selected the sectors on the line chart so that only the sectors that they're interested will be shown on the line chart. On the pages of each sector, a table of the general information of the sector will be provided as well as the line charts of the analysis result.

5 EXPERIMENTS AND EVALUATION

The data is composed of three parts. The first part is the stock's prices for all companies listed on S&P500 from 2018 January to 2022 March, with each company's respective information, such as sector, market capitalization and weight of S&P500 index. The reason of selecting such time period is that we want to observe the potential pattern change before and after COVID. In the horizon picked, the ratio of pre-pandemic (2018 January to 2020 February) and post-pandemic (2020 February to 2022 March) are roughly 1:1. The second part of data are control and explanatory variables, including the high frequency indicators of macro economy and general condition of the financial market, including 10 Year U.S. Treasury Yield, Effective Fed Fund Rate, Federal Reserve Balances, 5-Year Breakeven Inflation Rate and WTI (West Texas Intermediate) Oil Price Front Month. The third part is COVID related data sourced from Centers for Disease Control and Prevention (CDC), including total case, new cases, death count, ICU Patient count and vaccination rate starting from 2020 January to 2022 March.

In the data pre-processing step, the stock's price are averaged into 11 sectors by taking the weight of each company. Each sector is analyzed under LSTM, RNN and GRU learning algorithms by using Keras packages in R language as detailed in Section 3, while first 70% of data is used for model train and the rest is used for test and predicts.

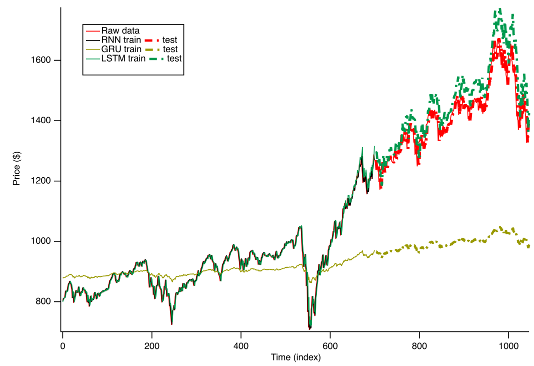

Figure 2: The model result for the consumer discretionary sector

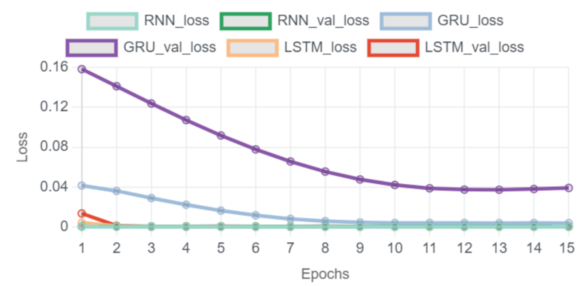

Figure 3: The model result for the consumer discretionary sector

As an example, Figure 2 is the train and predict results for the consumer discretionary sector, and Figure 3 is the model loss of consumer discretionary sector by epochs. As we can see from the plot, RNN and LSTM outperform GRU for the consumer discretionary sector.

6 CONCLUSIONS AND DISCUSSIONS

We compared three machine learning models, GRU, LSTM, and RNN, in predicting the price of different sectors in SP500, including Information Technology, Health Care, Financials, Consumer Discretionary, Communication Services, Industrials, Consumer Staples, Energy, Utilities, Real Estate, and Materials. Among these three models, LSTM model has a more precise performance during testing. Conversely, with abnormal loss and epoch, RNN model is malfunctional for the prediction. This abnormality can be explained by the nature of RNN algorithm. Under the LSTM model, all sectors and S&P500 plummet in March 2020, during which the number of confirmed COVID-19 cases in the US dramatically increased and restrictions on work and travel, e.g., social distancing, were implemented across the whole country. Information Technology is predicted to perform the best, with about 25% increase in price during testing. Although this project only evaluates the stocks in SP500, it provides insights on prediction of stock price using machine learning algorithms and elucidates LSTM model is a plausible method for the prediction.

The impact of a global pandemic like COVID to equity market, and more broadly financial market, is an extremely complicated topic. We think the future work can continue on below aspects:

(1) Regional market differentiation among countries with various anti-epidemic measures. For example, it would be interesting to have the same experiment on stock sectors in other regions such as Europe (Stoxx 50) and Hong Kong (Hang Seng Index), and observe if the models could reach a similar conclusion.
(2) Instead of doing analysis on sector level, one could explore to break down equities further into secondary industries. In this way, we could see more detailed impacts for companies with different operations. For example, "Consumer Durable Apparel" and "Hotels, Restaurants Leisure" are both under Consumer Discretionary Sector, while the impact of COVID greatly diverges.

In the end, we shall clarify that due to multiple variables affecting the stock price, it is a fact that the stock market can never be predicted. The model in this project can only estimate the future trend of the market based on the given database. Thus, this model should not be used for individual financial investment in reality.

All team members have contributed a similar amount of effort. For the detailed breakdown of the project tasks and task allocation, please refer to the Plan of activities in Proposal and Progress Report.

REFERENCES

[1] M Alassafi, M Jarrah, and R Alotaibi. 2022. Time series predicting of COVID-19 based on deep learning. *Neurocomputing* 468 (2022), 335–344.
[2] A Almehmadi. 2020. Covid-19 Pandemic Data Predict the Stock Market. (2020).
[3] Y Alparslan and E Kim. 2021. Extreme volatility prediction in stock market: When gamestop meets long short-term memory networks. *arXiv* 2103 (2021), 01121.
[4] S Baker, N Bloom, S Davis, K Kost, M Sammon, and T Viratyosin. 2020. The unprecedented stock market impact of COVID-19. (2020).
[5] J Chen, A Vullikanti, J Santos, S Venkatramanan, S Hoops, H Mortveit, B Lewis, W You, S Eubank, M Marathe, C Barrett, and A Marathe. 2021. Epidemiological and economic impact of covid-19 in the US. Scientific Reports. (2021).
[6] S Davis, D Liu, and X Sheng. 2021. Stock Price and Economic Activity in the Time of Coronavirus. (2021).
[7] K Eleftheriou and P Patsoulis. 2020. COVID-19 lockdown intensity and stock market returns: A spatial econometrics approach. (2020).
[8] U Gurav and N Sidnal. 2018. Predict Stock Market Behavior: Role of Machine Learning Algorithms. *Intelligent Computing and Information and Communication* (2018), 383–394.
[9] A Harjoto, F Rossi, and K Paglia. 2021. COVID-19: Stock market reactions to the shock and the stimulus. *Applied Economics Letters* 28(10) (2021), 795–801.
[10] H Hong, Z Bian, and C Lee. 2021. Covid-19 and Instability of Stock Market Performance: evidence from the U.S. (2021).
[11] T Ibn-Mohammed, B Mustapha, J Godsell, Z Adamu, A Babatunde, D Akintade, A Acquaye, H Fujii, M Ndiaye, A Yamoah, and L Koh. 2021. A critical analysis of the impacts of COVID-19 on the global economy and ecosystems and opportunities for circular economy strategies. *Resources* 164 (2021), 105169.
[12] T Kartal, O Depren, and K Depren. 2020. The determinants of Main Stock Exchange index changes in emerging countries: Evidence from Turkey in covid-19 pandemic age. *SSRN Electronic Journal* (2020).
[13] A Latinne, B Hu, J Olival, G Zhu, L Zhang, H Li, A Chmura, E Field, C Zambrana-Torrelio, H Epstein, B Li, W Zhang, F Wang, L Shi, and P Daszak. 2020. Origin and cross-species transmission of bat coronaviruses in China. *Nature Communications* 11 (2020), 4235.
[14] Mieszko Mazur, Man Dang, and Miguel Vega. 2021. COVID-19 and the march 2020 stock market crash. Evidence from S&P1500. *Finance research letters* 38 (2021), 101690.
[15] J Piret and G Boivin. 2021. Pandemics Throughout History. *Frontiers in Microbiology* 11 (2021).
[16] B Weng, L Lu, X Wang, M Megahed, and W Martinez. 2018. Predicting short-term stock prices using ensemble methods and online data sources. *Expert Systems with Applications* 112 (2018), 258–273.
[17] A Yong and E Laing. 2021. Stock market reaction to COVID-19: Evidence from US Firms' International exposure. *International Review of Financial Analysis* 76 (2021), 101656.