2021 International Conference on Emerging Smart Computing and Informatics (ESCI)
AISSMS Institute of Information Technology, Pune, India. Mar 5-7, 2021

Improving Traditional Stock Market Prediction Algorithms using Covid-19 Analysis

Rajni Jindal
HOD-Computer Engineering Department
Delhi Technological University
New Delhi, India
rajnijindal@dce.ac.in

Nikhil Bansal
Computer Engineering Department
Delhi Technological University
New Delhi, India
nikhilbansal19.nb@gmail.com

Nitin Chawla
Computer Engineering Department
Delhi Technological University
New Delhi, India
nitinchawla.emails@gmail.com

Sanskriti Singhal
Computer Engineering Department
Delhi Technological University
New Delhi, India
sanskriti0512@gmail.com

2021 International Conference on Emerging Smart Computing and Informatics (ESCI) | 978-1-7281-8519-4/20/$31.00 ©2021 IEEE | DOI: 10.1109/ESCI50559.2021.9396887

***Abstract*—The stock market is an organized body where public companies offer their stocks through initial public offerings and traders buy/sell these stocks so as to obtain profits. It is dynamic and volatile in nature which makes the task of stock market trend prediction a complex problem. In recent times, the COVID-19 pandemic has made this task even harder. With the rising number of COVID-19 cases across the globe, the market has never been more volatile. This has resulted in the poor performance of various traditional trend prediction algorithms because these algorithms do not account for the impact of the pandemic on the stock market trends. The proposed work aims to enhance the stock market prediction ability of various common prediction models by taking into account the factors related to COVID-19. The forecasting techniques analysed are Decision Tree Regressor, Random Forest Regressor and Support Vector Regressor (SVR). Currently the most affected countries by COVID-19 are the United States of America, India and Russia. Therefore we have analysed the prediction performance of various approaches discussed in this paper on S&P 500 Index, Nifty50 Index and RTS Index using Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). Results obtained showcase that all the techniques used perform better when the COVID-19 features were included.**

Keywords*—COVID-19, Stock Market Trend, Decision Tree Regressor, Random Forest Regressor, Support Vector Regressor (SVR), S&P 500 Index, Nifty50 Index, RTS Index.

I. INTRODUCTION

The daily stock market movement is extremely volatile and dynamic. It is influenced by many hidden factors and complex relationships which leads to a high degree of nonlinearity. Some theories on the nature of the stock market state that the stock market movement is completely random. One such theory is the Random walk theory [1] according to which stock market movement is independent of its past movement and hence cannot be predicted. While other assumptions like the Mean Reversion [2] argue that over a period of time the stock price is likely to move towards the average price. Another theory which is a development on the previous one called the Moving Average Reversion (MAR) argues that the average price is about the same as the mean of the prices in a previous time frame. Also various technical analysis show that past data can be analysed to make educated guesses that are quite close to the actual value.

COVID-19 outbreak is caused by SARS-CoV-2 and its first case was reported in China in the Wuhan region. WHO (World Health Organization) acknowledged it as a public health emergency internationally. COVID-19 has adversely affected more than 200 countries across the globe.

The lockdown in various countries has affected many lives, jobs and businesses. Businesses witnessed reduction in revenue generated which created cash flow challenges for the companies, thereby also affecting the economy. Some economists have declared COVID-19 as a Black-Swan Event for the stock market. It has had a catastrophic impact on the stock market. Numerous stocks have witnessed severe shifts. Some of them experienced drastic falls while the prices of others shot up. The investors are facing a hard time in investing their money as the prices are so unpredictable during the pandemic.

Various traditional prediction techniques have performed poorly and given unreliable results during the time of the pandemic since they do not take into account the various dimensions of the impact that the pandemic has had on the world. In this paper, we have compared the performance of such models by considering various COVID-19 related factors that have affected the stock market prediction.

II. RELATED WORK

Huang, Nakamori, and Wang, in their research paper [3] have proposed the widely used machine learning technique called Support Vector Machine to forecast the stock prices. In the paper they took into account the relationship that exists between various macroeconomic variables and the stock index. They have also highlighted the factor of how the US economy influences the Japan stock market flow therefore the S&P 500 Index is chosen as another input to be fed into the model. The authors have compared the performance of LDA, QDA and Elman Backpropagation Neural Networks (EBNN) and the proposed SVM. It was observed that SVM successfully forecasted the weekly stock market movement flow with the least amount of error.

Risul Islam Rasel et al. [4] used Artificial Neural Network (ANN) after pre-processing the data with the help of windowing techniques as an approach for predicting stock market trend to analyse financial instability. The paper proposed 3 efficient models predicting 1 day ahead, 5 days ahead and 10 days ahead. It was observed that the 1 day ahead model performed the best among the three models. They have also compared the performance of their approach with some other widely used algorithms - Support Vector Machine (SVM) and K-Nearest Neighbours (KNN). It was observed that ANN could predict the stock price more accurately than SVM and KNN.

Hyejung Chung et al. [5] proposed a “hybrid approach” combining the Genetic Algorithm with the LSTM neural network. This research suggests a method to decide the LSTM model’s time window and configuration using Genetic Algorithm for stock market financial prediction. In all the error metrics, Genetic Algorithm improved LSTM Model showed enhanced results because of an improved training procedure on the LSTM network.

Kang Zhang et al. [6] presented a novel approach of applying generative adversarial networks to the problem of stock price prediction. The paper considered building the generator as an LSTM network and the discriminator as a Multilayer Perceptron (MLP). Both the generator and the discriminator indulge in a zero sum game. The generator tries to learn the distribution of the dataset and predict values as close as possible to the real data while the discriminator performs the task of distinguishing between real data and the data predicted by the generator. In this way both the generator and the discriminator improve over time resulting in a better learning process.The results showed that the proposed GAN architecture performed better than the classical methods for stock market trend prediction.

Anshul Mittal et al. [7] proposed a technique wherein the twitter dataset is used to perform sentiment analysis and the tweets are classified into 4 common feelings- Calm, Happy, Alert and Kind. These predicted moods from the sentiment analysis and the processed Dow Jones Industrial Average (DJIA) values are then given as an input to their Self Organizing Fuzzy Neural Networks (SOFNN) model to learn and predict future DJIA values.

It was observed that none of the approaches above could account for the impact of a pandemic on stock market trends.

III. RESEARCH METHODOLOGY

The following section elaborates upon the methodology and techniques adopted in our approach towards improving the traditional stock market prediction algorithms for the prediction task.

A. *Our Approach*

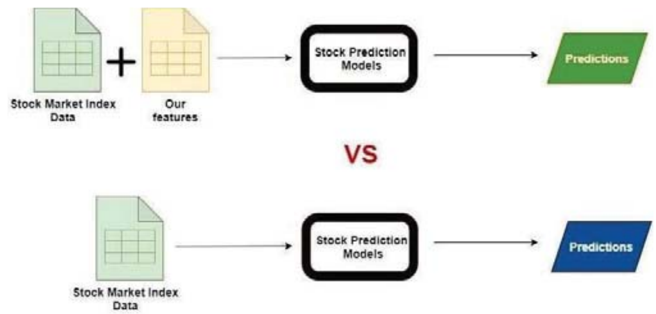

Fig. 1. Difference in Approach

With the rising number of COVID-19 cases across the globe, the market has never been more volatile. This has resulted in the poor performance of various traditional trend prediction algorithms because these algorithms do not account for the impact of the pandemic on the stock market trends. In our approach we have included various features to account for the pandemic so that our models are better able to learn the trend and give more accurate results. For the purpose of validation, we have tested our technique with stock market data of 3 different countries (The USA, India and Russia) while making use of 3 different traditional stock market prediction algorithms for the purpose of making predictions. The machine learning algorithms used include Decision Tree Regressor, Random Forest Regressor and Support Vector Regressor.

Before feeding the data into the model we have combined the standard stock market index attributes with the Covid-19 features. We then shifted the closing value column of our dataset by 1 field since we want to be able to predict the closing value of the index for the next day given the data of the current day. The dataset has been divided into a split of 80%- 20% i.e. 80% of the data is being used for the training purpose and 20% of the data is being used for testing purposes. We have then analyzed and compared the results obtained by using our features to those obtained by using the traditional features only.

B. *Dataset Description*

COVID-19 has adversely affected more than 200 countries across the globe. Currently the most affected countries by COVID-19 are the United States of America, India and Russia .Therefore we have analysed the prediction for S&P 500 Index, Nifty50 Index and RTS Index.

C. *Data Collection*

Historic data for each trading day of S&P 500 Index, Nifty50 Index and RTS Index : S&P 500 index is evaluated by taking a weighted average of 500 large companies registered in the United States. RTS Index is evaluated by taking a weighted average of 50 stocks in Russia listed on the

Moscow Exchange. The data for the above was obtained from the Yahoo Finance website [8]. Nifty50 is evaluated by taking a weighted average of the largest 50 establishments in NSE India. The data for this was obtained from the website of National Stock Exchange (NSE) [9].

The statistical data for the COVID-19 was obtained from the website [9] of the World Health Organisation (WHO). The timeframe of the datasets is the year 2020 i.e. 3th January 2020 – 30th December 2020.

D. Data Preprocessing

Before feeding the data to various models to be analysed and compared we have performed some pre-processing. It was found that in the dataset of the stock market indexes that the values corresponding to dates on weekends and holidays were missing as the stock market is not active on those days. Therefore, in order to clean the data, we omitted the rows corresponding to those dates in both our datasets. The duplicate entries and other noises were carefully handled. For scaling of our data we have used the MinMaxScaler module of the scikit learn framework which adjusts the value of our input features. We have adjusted the value of each of our features between 0 and 1. Scaling the features brings the feature vectors to the same scale so that the model is not negatively impacted by large variations in the data and enables models to learn better. The following features have been chosen Date, Opening Price (OP) for the trading day, High Price (HP) for the trading day, Low Price (LP) for the trading day, Closing Price (CP) for the trading day, Shares Traded, Turnover, New Cases, Cumulative Cases, New deaths, Cumulative Deaths.

IV. RESULTS

We have used Root Mean Squared Error (RMSE) to evaluate our results. Another metric that is used to compare the results is the MAPE (Mean Absolute Percentage Error) since we are concerned about the variance of predicted value from the actual value i.e. relative error.

$$RMSE=\sqrt{\frac{\sum_{i=1}^{N}(Predicted_i - Actual_i)^2}{N}} \tag{1}$$

$$\mathrm{M} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \tag{2}$$

The best prediction results in the below table are in bold. The lower the value of RMSE and MAPE the better the model performs on the forecasting task. We observe that the RMSE values and MAPE values are less when we include the COVID-19 features in our dataset.

TABLE I. COMPARISON OF RESULTS

Country	Machine Learning Model Used	Root Mean Squared Error		Mean Absolute Percentage Error	
		With Covid Features	Without Covid Features	With Covid Features	Without Covid Features
India (Nifty50 Index)	Decision Tree Regressor	**212.672**	270.014	**1.569**	1.942
	Random Forest Regressor	**177.436**	190.426	**1.285**	1.355
	Support Vector Regressor	**128.883**	132.451	**0.902**	0.906
Russia (RTS Index)	Decision Tree Regressor	**57.517**	66.200	**1.574**	1.600
	Random Forest Regressor	**45.490**	50.762	**1.202**	1.265
	Support Vector Regressor	**33.924**	53.764	**0.922**	1.521
United States of America (S&P 500 Index)	Decision Tree Regressor	**53.409**	59.381	**1.273**	1.338
	Random Forest Regressor	**49.642**	53.272	**1.169**	1.281
	Support Vector Regressor	**45.572**	58.700	**1.134**	1.419

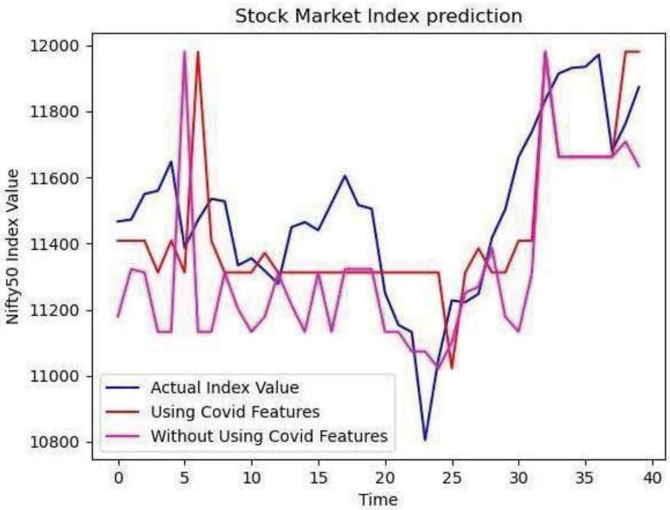

Fig. 2. Decision Tree Regressor (India)

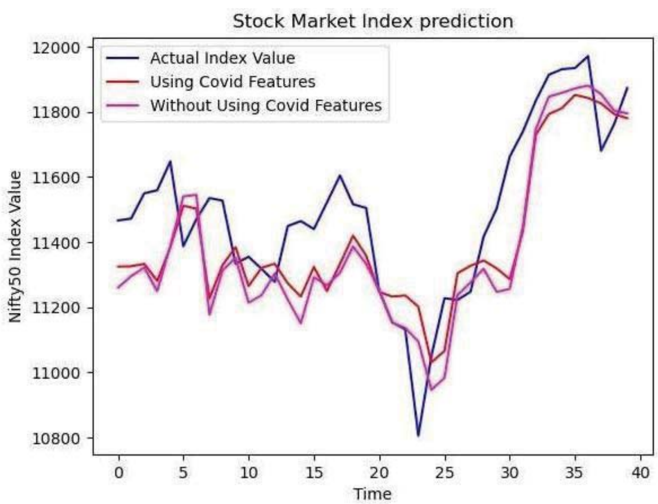

Fig. 3. Random Forest Regressor (India)

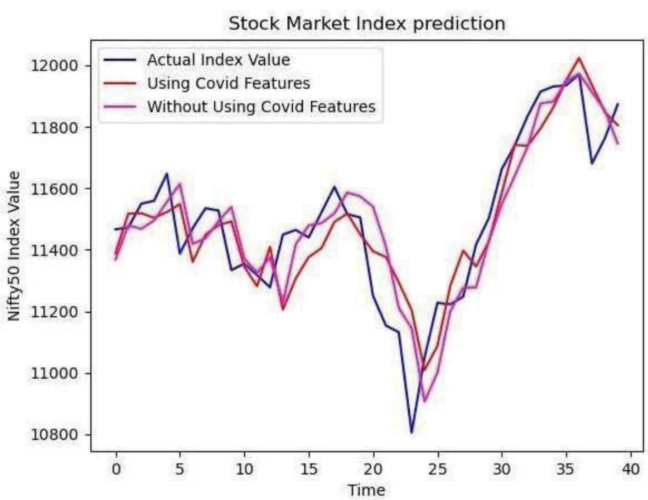

Fig. 4. Support Vector Regressor (India)

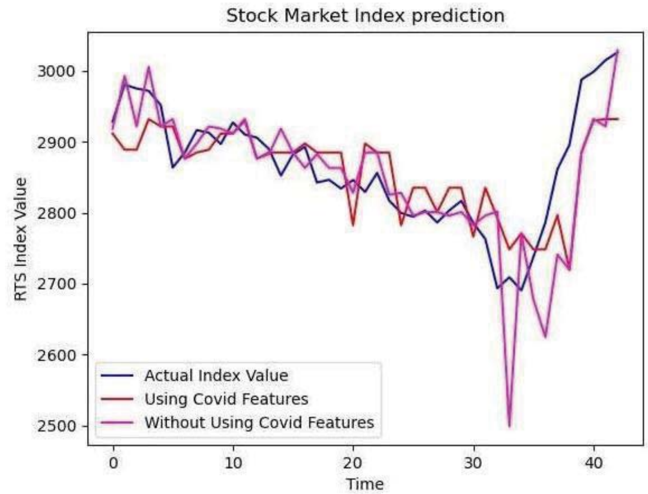

Fig. 5. Decision Tree Regressor (Russia)

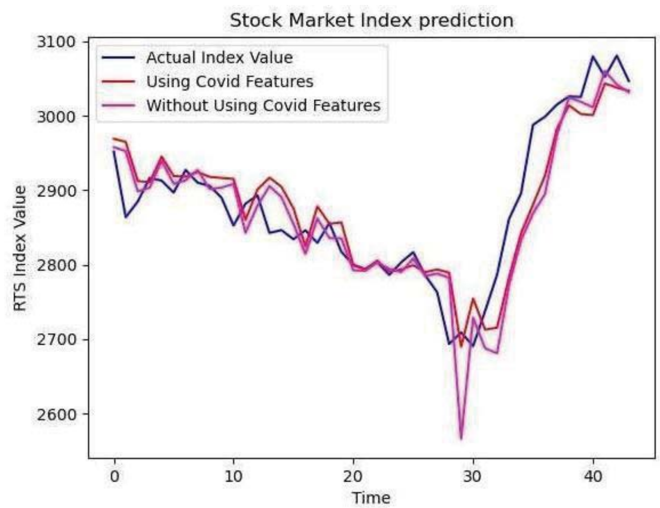

Fig. 6. Random Forest Regressor (Russia)

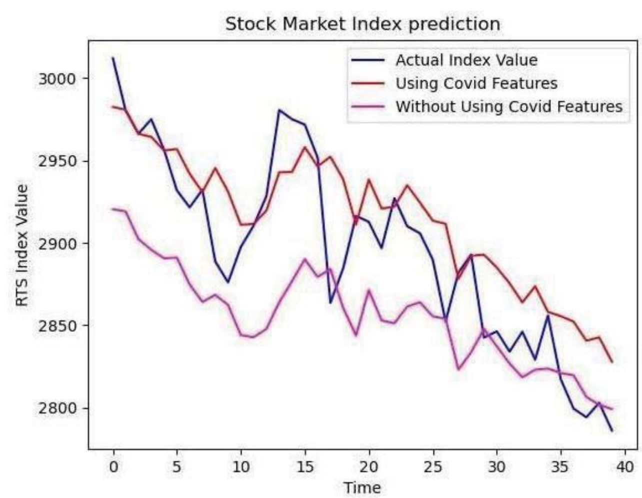

Fig. 7. Support Vector Regressor (Russia)

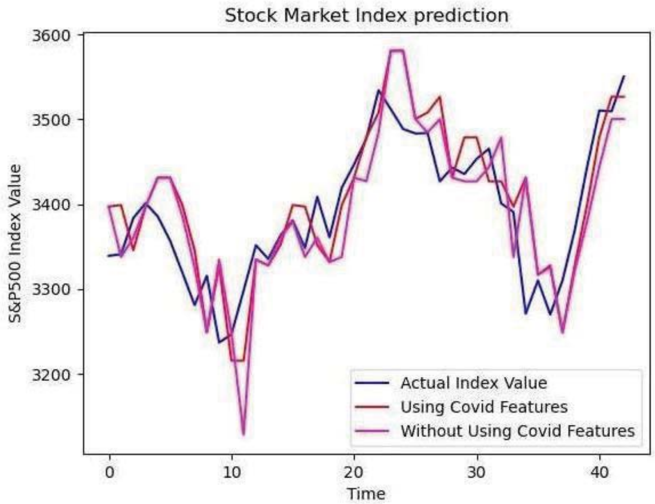

Fig. 8. Decision Tree Regressor (USA)

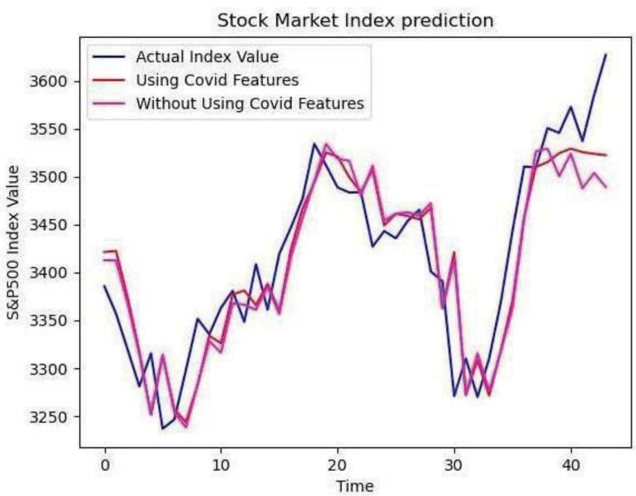

Fig. 9. Random Forest Regressor (USA)

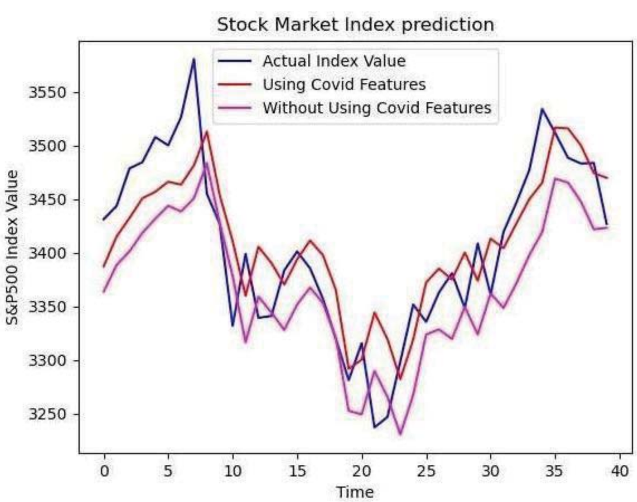

Fig. 10. Support Vector Regressor (USA)

V. CONCLUSION

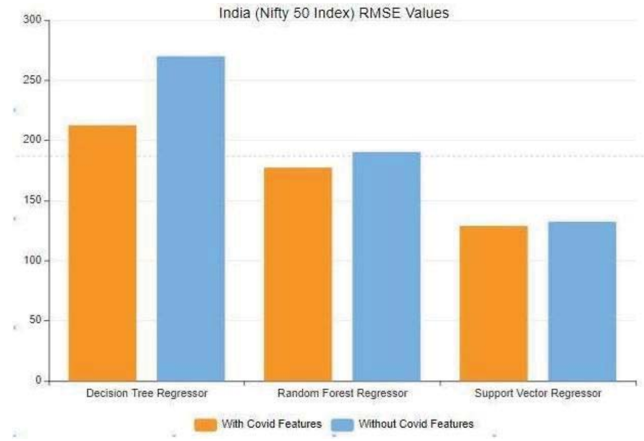

Fig. 11. RMSE Comparison Nifty50 Index (India)

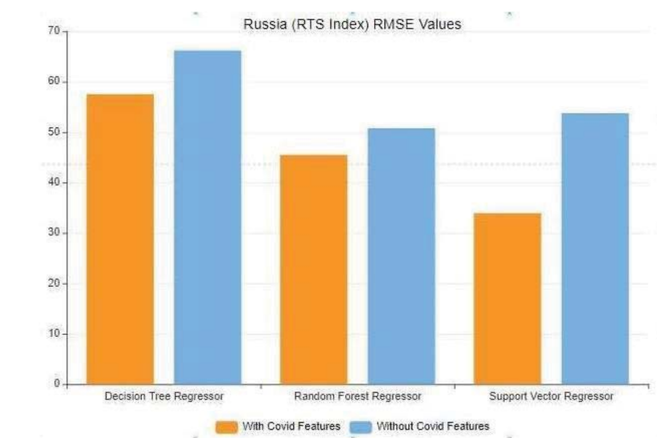

Fig. 12. RMSE Comparison RTS Index (Russia)

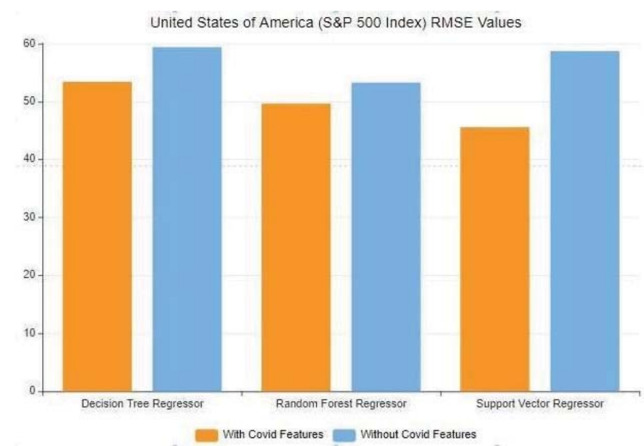

Fig. 13. RMSE Comparison S&P 500 Index (USA)

The following conclusions can be drawn from the above results :-

- There is a correlation between stockmarket trends and Covid-19 cases.

- Making use of the number of cases andthe number of deaths in a country due to COVID-19 improves the prediction ability of the models.

VI. FUTURE SCOPE

As more data is being made available by the day there is a scope for predicting stock market trends during times of pandemic using neural network based approaches which are proven to be more robust in prediction tasks involving time series data like the LSTM (Long Short Term Memory) and GRU (Gated Recurrent Units) techniques.

REFERENCES

[1] Van Horne, J. C., & Parker, G. G. (1967). The random-walk theory: an empirical test. Financial Analysts Journal, 23(6), 87-92.

[2] Poterba, J. M., & Summers, L. H. (1988). Mean reversion in stock prices: Evidence and implications. Journal of financial economics, 22(1), 27-59.

[3] Huang, Wei, Yoshiteru Nakamori, and Shou-Yang Wang. "Forecasting stock market movement direction with support vector machine." Computers & operations research 32.10 (2005): 2513-2522.

[4] Rasel, R. I., Sultana, N., & Hasan, N. (2016). Financial instability analysis using ANN and feature selection technique:Application to stock market price

[5] Chung, H., & Shin, K. S. (2018). Genetic algorithm-optimized long short-term memory network for stock market prediction. Sustainability, 10(10), 3765.

[6] Zhang, K., Zhong, G., Dong, J., Wang, S., & Wang, Y. (2019). Stock market prediction based on generative adversarial networks. Procedia computer science, 147, 400-406.

[7] Mittal, Anshul, and Arpit Goel. "Stock prediction using twitter sentiment analysis." Stanford University, CS229 (2011 http://cs229.stanford.edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis.pdf) 15 (2012).

[8] "Data for stock Market Index"[Online]. Available: https://in.finance.yahoo.com/ [Accessed Dec 20,2020].

[9] "Data for Nifty50 Index" [Online]

[10] Available: https://www.nseindia.com/ [Accessed Dec 20,2020]

[11] "Data for Covid" [Online]

[12] Available: https://www.who.int/data/collections/ [Accessed Dec 20, 2020]