Vol. 4, Issue II, Apr-Jun, 2021 http://www.bharatpublication.com/journal-detail.php?jID=25/IJTSE

AN IN DEPTH STUDY OF THE CORRELATION BETWEEN COVID 19 ANALYSIS AND STOCK MARKET TO ENHANCE THE EFFICACY OF STOCK MARKET PREDICTION ALGORITHMS

Anika

Birla Institute of Management Technology

ABSTRACT

The stock market serves as a structured platform where public companies issue their stocks via initial public offerings, and traders buy and sell these stocks to realize profits. Its inherent dynamism and volatility pose significant challenges to predicting market trends accurately. The emergence of the COVID-19 pandemic has exacerbated these challenges. With the escalating global spread of COVID-19 cases, market volatility has reached unprecedented levels, rendering traditional trend prediction algorithms less effective due to their failure to incorporate the pandemic's impact on market dynamics. This study proposes an approach to augment the predictive capability of standard market prediction models by integrating COVID-19-related factors. The analysis encompasses Decision Tree Regressor, Random Forest Regressor, and Support Vector Regressor (SVR) forecasting techniques. Given the substantial impact of COVID-19 on countries like the United States, India, and Russia, the prediction performance evaluation focuses on significant indices such as the S&P 500 Index, Nifty50 Index, and RTS Index. Assessment metrics include Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). Results demonstrate improved performance across all techniques when incorporating COVID-19 features.

INTRODUCTION

The stock market, a complex entity, is a dynamic system characterized by extreme volatility and constant dynamism. Influenced by numerous hidden factors and intricate relationships, it operates within a realm of high nonlinearity. Different theories attempt to explain its nature, with some suggesting that stock market movements are entirely random, as proposed by the Random Walk Theory. This theory posits that stock market movement is independent of its past, rendering it unpredictable. Conversely, theories like Mean Reversion argue that stock prices tend to gravitate towards an average price over time. Building upon this, Moving Average Reversion (MAR) contends that the average price aligns closely with the mean of past prices within a specific timeframe. Moreover, various technical analysts suggest that past data can be leveraged to make informed predictions that closely approximate actual values.

The COVID-19 pandemic, an unprecedented global public health crisis triggered by the SARS-CoV-2 virus, first emerged in the Wuhan region of China. Its adverse effects have been felt across more than 200 countries, precipitating lockdown measures and disrupting lives, jobs, and businesses. This led to a significant reduction in revenue for many companies, resulting in cash flow challenges and cascading effects on the economy. Economists have labelled COVID-19 as a Black Swan Event for the stock market, denoting its unexpected and catastrophic impact. Stocks

have experienced profound fluctuations, with some plummeting while others surging. Investors have grappled with the unpredictability of market prices during this unprecedented period.

Traditional prediction techniques have faltered amidst the pandemic, delivering unreliable results due to their failure to account for the multifaceted impact of COVID-19 on the global landscape. This paper, recognizing the need for new approaches in stock market prediction, aims to evaluate the performance of such models by incorporating various COVID-19-related factors that have influenced stock market predictions.

RESEARCH METHODOLOGY

This section provides a detailed explanation of the methodology and techniques employed in our endeavour to enhance traditional stock market prediction algorithms for the prediction task.

A. Our Approach

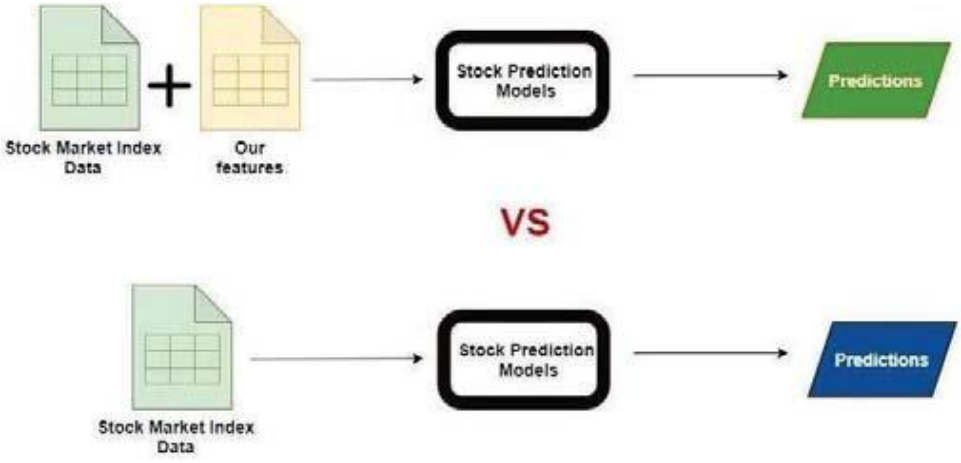

Fig. 1. Difference in Approach

With the escalating number of COVID-19 cases worldwide, market volatility has reached unprecedented levels. Consequently, traditional trend prediction algorithms have exhibited subpar performance due to their inability to account for the pandemic's impact on market trends. In our approach, we have integrated various features to capture the pandemic's effects, thereby enhancing the accuracy of our models in predicting trends. To validate our technique, we conducted tests using stock market data from three countries—the USA, India, and Russia—employing three traditional stock market prediction algorithms: Decision Tree Regressor, Random Forest Regressor, and Support Vector Regressor.

Before model training, we combined standard stock market index attributes with COVID-19 features. Additionally, we shifted the closing value column of our dataset by one field, enabling us to predict the index's closing value for the next day based on the current day's data. The dataset was split into an 80%-20% ratio for training and testing. Subsequently, we analyzed and compared the results obtained using our augmented features against those obtained using traditional features alone.

Dataset Description:

COVID-19 has profoundly impacted over 200 countries globally, with the United States, India, and Russia most affected. Hence, our analysis focuses on predicting trends for the S&P 500 Index, Nifty50 Index, and RTS Index.

Data Collection:

Historical data for each trading day of the S&P 500 Index, Nifty50 Index, and RTS Index were obtained. The S&P 500 index comprises a weighted average of 500 large US companies, while the RTS Index reflects a weighted average of 50 Russian stocks listed on the Moscow Exchange. Data for these indices were sourced from Yahoo Finance. The Nifty50 Index represents a weighted average of the top 50 establishments in India and was obtained from the National Stock Exchange (NSE) website. Statistical COVID-19 data was collected from the World Health Organization (WHO) website, covering the timeframe from January 3, 2020, to December 30, 2020.

Data Preprocessing:

Before model analysis, preprocessing steps were undertaken. Missing values corresponding to weekends and holidays in the stock market index datasets were removed. Duplicate entries and other noises were also addressed. Data scaling was performed using the MinMaxScaler module from the sci-kit-learn framework, adjusting input feature values between 0 and 1. This scaling ensures consistency among feature vectors, mitigating the impact of data variations on model performance. Selected features include Date, Opening Price (OP), High Price (HP), Low Price (LP), Closing Price (CP), Shares Traded, Turnover, New Cases, Cumulative Cases, New Deaths, and Cumulative Deaths.

RESULTS

Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) were employed to evaluate and compare the accuracy of our results, focusing on the variance between predicted and actual values.

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

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

The table below highlights the best prediction results, with bold indicating the lowest values for RMSE and MAPE, indicating the model's better forecasting performance. Incorporating COVID-19 features into our dataset leads to lower RMSE and MAPE values.

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

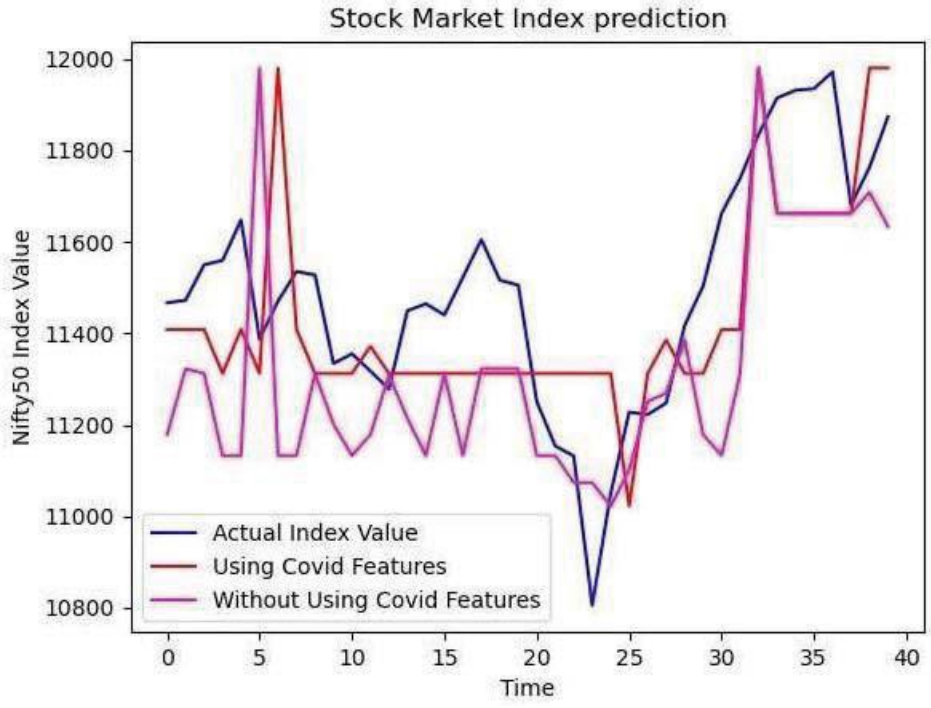

Fig. 2. Decision Tree Regressor (India)

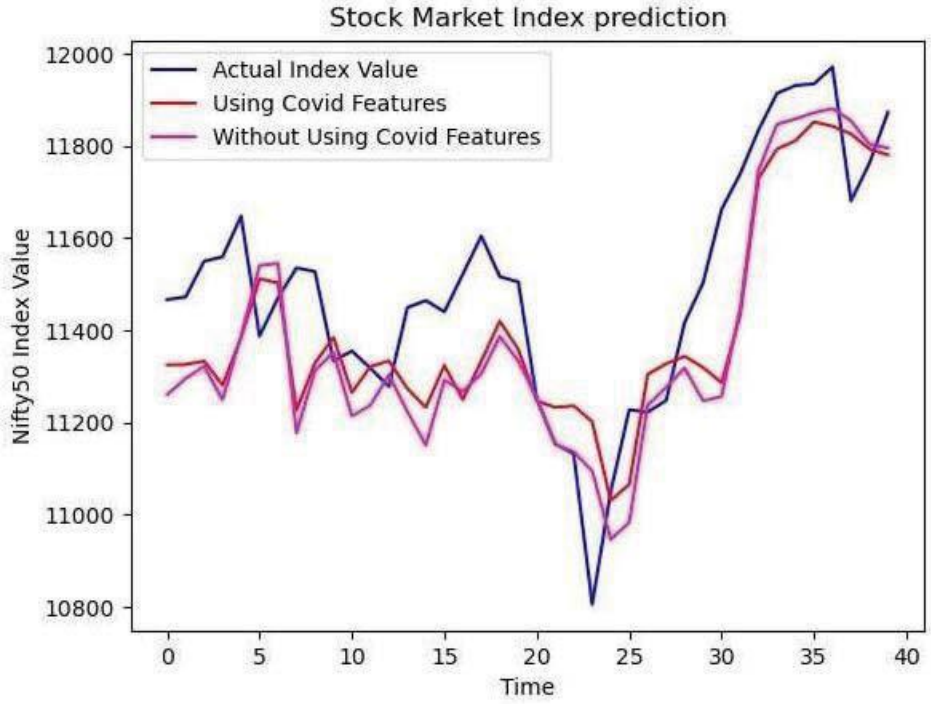

Fig. 3. Random Forest Regressor (India)

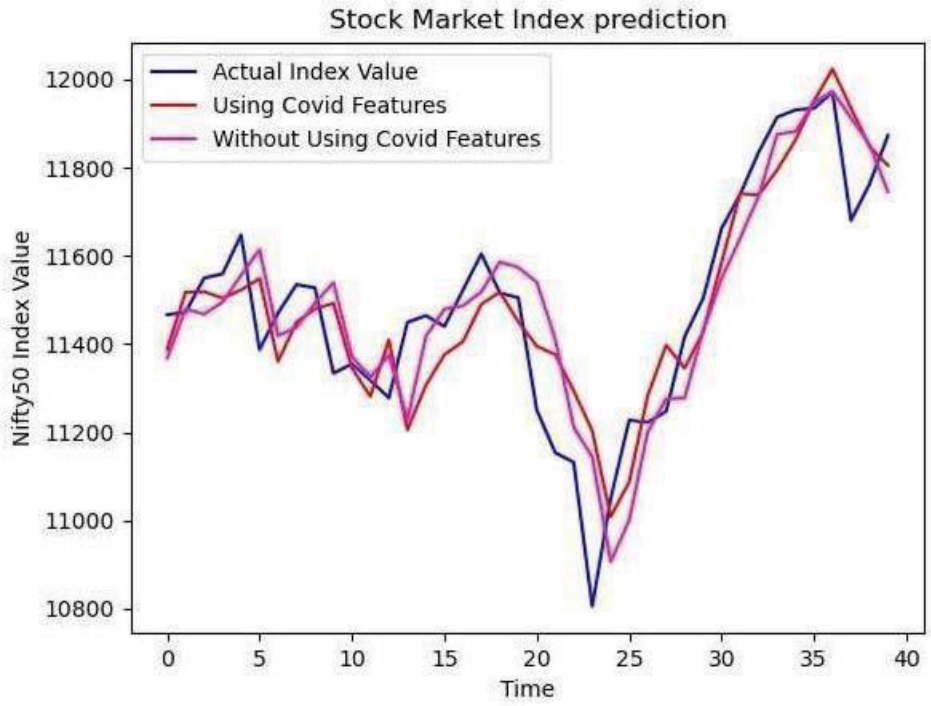

Fig. 4. Support Vector Regressor (India)

CONCLUSION

From the above results, the following conclusions can be drawn:

1. There exists a correlation between stock market trends and COVID-19 cases.

2. Utilizing the number of COVID-19 cases and deaths in a country enhances the prediction capability of the models.

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-Stock Market PredictionUsingTwitterSentimentAnalysis.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]