



COVID-19 Vaccination Effect on Stock Market and Death Rate in India

Jyotirmayee Behera¹ · Ajit Kumar Pasayat² · Harekrushna Behera¹

Accepted: 18 February 2022 / Published online: 11 March 2022

© The Author(s), under exclusive licence to Springer Japan KK, part of Springer Nature 2022

Abstract

The COVID-19 epidemic has brought attention to the vulnerability of new illnesses, and immunization remains a viable option for resuming normal life. This paper examines the influence of COVID-19 vaccination on the death rate and the performance of stock market in India. For this study, COVID-19 vaccination and death rate data is gathered from the Ministry of Health and Family Welfare (MoHFW) portal, and the data for the stock index is taken from the Bombay Stock Exchange (BSE), India. In order to achieve a precise representation of feature significance and distribution, EDA (Exploratory Data Analysis) is utilized in this study. The impact of COVID-19 immunization on the mortality rate and stock market index is investigated using both statistical analysis and Machine Learning Regression-based models. The models are remarkably accurate in reproducing actual result. The empirical study suggests that vaccination has a strong positive impact on the stock market and reducing the death rate. Furthermore, the policies recommended by government and monetary authorities coupled with COVID-19 vaccine supported the stock market recovery in pandemic.

Keywords COVID-19 · Vaccination · Stock market · Exploratory data analysis · Statistical data analysis · Machine learning

✉ Ajit Kumar Pasayat
akpasayat@iitkgp.ac.in
Jyotirmayee Behera
jyotirmayeebehera452@gmail.com
Harekrushna Behera
harekrum@srmist.edu.in

¹ Department of Mathematics, SRM Institute of Science and Technology, Kattankulathur, Chengalpattu, Tamil Nadu 603203, India

² Indian Institute of Technology, Kharagpur, Kharagpur, West Bengal 721302, India

1 Introduction

The Coronavirus disease was first found in Wuhan, China, in December 2019, and since then, it has spread throughout the world (Pasayat et al., 2020). As a result, on 11 March 2020, World Health Organization (WHO) declared coronavirus disease as a pandemic. After the Spanish Flu, the Coronavirus pandemic is represented as the most devastating health crisis in the world. Thus for the communicable COVID-19, several countries have imposed restrictions on their economic agents' ability to move from one country to another or even inside the same country. It is having a rising influence on the global economy, and due to stress on supply chains, decrease in government income, labor movement restrictions, stock market volatility, and reduced consumer activity the global health crisis has sadly turned into a global economic disaster (Du et al., 2021).

The death rate is the most significant element in determining whether a highly contagious disease becomes a public health problem and poses a danger of spreading to other people, resulting in a pandemic. The COVID-19 has sparked worldwide concern and fear since numerous fatalities have been recorded (Wang et al., 2020a; Wu et al., 2020). Recent research has produced a number of county-level studies on the COVID-19 death rate. The researcher Wang et al. (2020b) used the Patient Information Based Algorithm (PIBA) to estimate the death rate based on patient data in Wuhan and later in other Chinese cities in which he has demonstrated that PIBA can be used to forecast future fatalities and estimate the death rate of a new infectious illness in real-time. Possible explanatory variables for mortality rates at the state and county levels, including geography, demographics, socioeconomic, health conditions, hospital features, and politics, have been examined by Feinhandler et al. (2020).

This pandemic wave creates volatility and chaos, which affect the confidence of investors in both Indian markets and the world. As a consequence, a lot of researchers are interested in exploring the impact of the pandemic on financial markets. Al-Awadhi et al. (2020), Topcu and Gulal (2020), and Ashraf (2020) have shown that the growth rate of the confirmed covid cases and death rate influence the price of the stock negatively. In addition to equities, the influence of the other assets classes like cryptocurrency (Conlon & McGee, 2020; Demir et al., 2020), gold (Gharib et al., 2021); Mensi et al., 2020), real estate (Milcheva, 2021; Ling et al., 2020), oil price (Sharif et al., 2020; Wu et al., 2021) and bonds (Falato et al., 2020) have been explored by the research. Furthermore, by examining the potential internal and external determinants, Khattak et al. (2021) explored the potential predictors of European financial market during COVID-19 crisis. In this literature, they have used Least Absolute Shrinkage and Selection Operator (LASSO) Machine Learning technique for both selections of the model and regularization in the analysis.

Moreover, a lot of researchers studied the effect of COVID-19 and its consequences on the Indian economy. As India is the country with the second largest population in the world, the pandemic hit the stock market as well as the Indian economy drastically. Pegu (2020) examined the future probability of GDP in the

Indian economy as well as analyzed different sectors which are affected due to the COVID-19 pandemic. To study the significant effect of COVID-19 cases on twelve sectors, an Ordinary Least Square (OLS) regression model was developed by Yadav et al. (2021). The study found that at the initial stage of the pandemic, the stock market responded negatively, but after some days, the indices are in their recovery mode. Barbate et al. (2021) have shown the long-term and short-term effects of lockdown by taking a decision tree approach for making the projections on COVID-19. Using a multivariate econometric model, Baber and Rao (2021) examined the impact of lockdown and social distancing on the Indian economy. In this study, the authors compare the economy for the period before and after lockdown and analyzed the economic activity, stock market, and exchange rate in India. In addition to that, Alam et al. (2020) used the Market Model Event study methodology to check the Indian Stock market before and after the lockdown period in India. As a result, a positive average return has been found in the lockdown period, which proves that until the improved situation in India, the stock market performance was affected positively.

On the other hand, the arrival of the first COVID-19 vaccination in early 2021 came with a hope of stepping closer towards immunity, and due to this, the unexpected and uncontrolled growth of COVID-19 death rate has been decreased. Song and Yi (2021) predicted the impact of different vaccination strategies on mortality, productivity loss, and healthcare burden in the UK using a Susceptible-Exposed-Infectious-Removed (SEIR) model. In addition to that, using official information of Tennessee, United States, Roghani (2021) demonstrated how age-specific immunization regimens lower daily cases, hospitalizations, and mortality rates. From the study they found that, vaccination of older age groups is the most effective approach to avoid greater transmissions, lower hospitalization and mortality rates.

Furthermore, it brought a normal and stable condition to financial markets. Despite that, the impact of vaccination on this exciting field has not been fully explored. An asset pricing perspective is developed by Acharya et al. (2020), which examines the value of a cure by creating a novel "vaccine progress indicator". They estimated the level of exposure within each industry and studied the vaccine progress impact in the cross-section of the industries. Moreover, they found that the return of the stock market can be possible by the decrease in the expected vaccine deployment time in a year. The reactions of the stock market to COVID-19 vaccine human clinical trial has been analyzed by Chan et al. (2021). As a result, it is concluded that the investor's sentiment, as well as the return of the stock market of 49 countries, rises after the first day of the clinical trial. In other words, after the consecutive phases of the clinical trial, the equity market started to react positively. Rouatbi et al. (2021) explored the impact of vaccination on the volatility of the international equity market. In which, they have shown that the impact of vaccination in the developed market is relatively stronger than emerging one.

Since India has the second-largest population, the sixth-largest economy in the world by nominal GDP and the third-largest by purchasing power parity, the pandemic poses a particular threat to the country. The country has never seen such a widespread health epidemic; further, because of asymptomatic COVID-19 transmission during the incubation phase, vaccination is the only way to develop herd

immunity. In this regard, pharmaceutical companies around the world have developed COVID-19 vaccines, which have received regulatory approval for emergency use. Countries have begun major immunization campaigns. It is seen that vaccination in India has a great impact on the mortality rate and the stock market. To the best of the authors' knowledge, there are only a few studies on the impact of the COVID-19 vaccination on the Indian stock market and death rate.

The article has two objectives, one is to investigate the influence of COVID-19 immunization on the Indian stock market, and the other is to predict the mortality rate with response to the vaccination process. For this, data for COVID-19 vaccination, death rate, and share market index are collected for the same time period. This study implements both statistical data analysis and Machine Learning regression techniques. Moreover, EDA is employed to provide a statistical picture of the data distribution and to investigate the appropriate features. The major objective is to show that the introduction of the vaccination results in a stable stock market with a lower mortality rate.

Regression is a supervised learning technique that aids in the discovery of variable correlations. Also, it allows us to forecast a continuous output variable using one or more predictor variables. Prediction, forecasting, time series modelling, and determining the causal-effect relationship between variables are all common applications of regression (Dasgupta et al., 2011). On a target-predictor graph, regression depicts a line or curve that passes through all of the datapoints in such a way that the vertical distance between the datapoints and the regression line is as small as possible. Many Regression Techniques analyses are available, depending on a variety of factors such as the number of predictors, the type of target variable, and the shape of the regression line. However, in this work linear regression, support vector regression, random forest regression, and KNN regression techniques have been used for prediction process.

- **Linear Regression** Linear regression is one of the most easy and simple algorithms for calculating regression and displaying the relationship between continuous variables. It is called linear regression because it depicts the linear relationship between the independent variable (X -axis) and the dependent variable (Y -axis). Simple linear regression is defined as linear regression with only one input variable (x). When there are multiple input variables, the linear regression is referred to as multiple linear regression. The mathematical equation of linear regression is

$$Y = aX + b$$

where X is independent variables (predictor variables), Y is dependent variables (target variables), and a and b are constants (Noi et al., 2017). Linear regression assumes that the dependent and independent variables have a linear relationship, which is not always the case in real-world scenarios. This is when you should use non-linear regression techniques. Furthermore, if the independent variables are related to one another, or if there is multicollinearity, Linear Regression will fail. To avoid this, only keep one of the correlated independent variables.

- **Support Vector Regression** The Support Vector Regression algorithm is a continuous-variable regression algorithm. In SVR, we always try to find a hyperplane with the largest possible margin, so that the maximum number of datapoints are covered. SVR's main goal is to consider as many datapoints as possible within the boundary lines, and the hyperplane (best-fit line) must contain as many datapoints as possible (Smola & Schölkopf, 2004). SVR outperforms the linear regression model in high-dimensional space while being more robust to outliers. In support vector regression analysis, it is also very simple to accommodate new data points. However, training takes a long time and is not suitable for large datasets. If the number of samples is less than the number of features, SVR will perform poorly. The predictions do not have any probabilistic explanations.
- **Random Forest Regression** Random Forest Regression is a supervised learning algorithm for regression that employs an ensemble learning method. The ensemble learning method combines predictions from multiple machine learning algorithms to produce a more accurate prediction than a single model. During training, a Random Forest builds several decision trees and outputs the mean of the classes as the prediction of all the trees (Shataee et al., 2012). Random Forest Regression model is both effective and precise. It usually works well on a wide range of issues, including those involving non-linear relationships. However, there are some drawbacks, such as lack of interpretability, the possibility of overfitting, and the need to choose the number of trees to include in the model.
- **KNN Regression** KNN regression is a non-parametric method that approximates the relationship between independent variables and continuous outcomes by averaging observations in the same neighborhood in an intuitive manner. The analyst must set the size of the neighborhood, or it can be chosen using cross-validation to find the size that minimizes the mean-squared error (Franco-Lopez et al., 2001). While the method is appealing, it quickly becomes impractical as the dimension grows, i.e. when there are a large number of independent variables.

1.1 Policy Recommendation

Following decisive monetary and fiscal policy responses, gradual easing of COVID-19-induced lockdown measures, and strong FPI inflows, the Indian equity market experienced sharp corrections in Q4:2019-20 amidst the tragic outbreak of COVID-19 (Dev et al., 2020). Furthermore, the stock market recovery in the pandemic was aided by policies recommended by government and monetary authorities, as well as the COVID-19 vaccine. Overall, the BSE Sensex increased by 68.0 percent to close at 49,509 on March 31, 2021, while the Nifty 50 increased by 70.9 percent to close at 14,691.

The Indian stock market got off to a strong start this year, posting its biggest monthly gain in 11 years in April 2020. During the month, the BSE Sensex rose 14.4% on expectations of a second round of government stimulus and the Reserve Bank's announcement of various liquidity measures. Market sentiment was lifted by

positive reports of COVID-19 drug trials and the easing of lockdown measures in major economies. Domestic equities took a breather after the Reserve Bank of India announced the "Aatma Nirbhar Bharat" stimulus package and a large policy repo rate cut on May 22.

Investors weighed the roll-out of corona virus vaccines in the country and upbeat corporate results for Q3:2020-21 against the persistent rise in COVID-19 cases around the world as domestic markets remained largely volatile in January 2021. On January 21, 2021, the benchmark hit 50,000 for the first time in history in intra-day trade before losing all of its gains due to concerns about reports of a new border standoff between India and China, weak global cues over stretched valuations in US equities, and cautious trading ahead of the Union Budget. The benchmark rose to a new high of 52,154 on February 15, 2021, reversing a downward trend that began at the end of January 2021, boosted by budgetary proposals, the Reserve Bank's optimistic outlook for GDP revival, and positive global market cues (Reserve Bank of India, 2021).

The remaining of the study proceeds as follows: Sect. 2 presents the details of the used data and methodology. In Sect. 3, the results are analyzed. The summary of the paper is highlighted in Sect. 4. Finally, in Sect. 5, a concluding remark is presented.

2 Data and Methods

This section describes the data that was utilized in the study as well as the methodology that was employed. It begins with collecting data, which includes vaccination, stock market, and death rate in India. The relationship between vaccination and the stock market index and death rate has been investigated using statistical data analysis. Furthermore, to make the dataset suitable for training, preprocessing and pre-analyzing approaches are used for Machine Learning prediction. The dataset is then trained using four algorithms or estimators. After that, the models are tested to check the link between vaccination with the stock market index and death rate.

2.1 Data

To review the effectiveness of COVID-19 vaccination on the stock market of the Indian economy, daily data is collected for the stock market index of S&P SENSEX from the Bombay Stock Exchange (BSE India 2021), for the time period of February 1, 2021, to July 30, 2021, which consists of Opening price, High price, Low price, and Closing price of the stock index. The sample of the data set is represented in Table 1.

On January 16, 2021, India began administering COVID-19 vaccinations. Covishield, a brand of the Oxford-AstraZeneca vaccine manufactured by the Serum Institute of India, and Covaxin, created by Bharat Biotech, were approved for emergency use in India at the start of the program. The Indian government approved the Russian Sputnik V vaccine (distributed locally by Dr. Reddy's Laboratories) as a third vaccination in April 2021, and it went into effect in May 2021.

Table 1 Sample of S & P BSE SENSEX dataset

Date	Open	High	Low	Close
01-Feb-21	46617.95	48764.4	46433.65	48600.61
02-Feb-21	49193.26	50154.48	49193.26	49797.72
03-Feb-21	50231.06	50526.39	49515.88	50255.75
04-Feb-21	50212.25	50687.51	49926.45	50614.29
05-Feb-21	51031.39	51073.27	50565.29	50731.63
01-Mar-21	49747.71	50058.42	49440.46	49849.84
02-Mar-21	50258.09	50439.82	49807.12	50296.89
03-Mar-21	50738.21	51539.89	50512.84	51444.65
04-Mar-21	50812.14	51256.55	50539.92	50846.08
05-Mar-21	50517.36	50886.19	50160.54	50405.32
01-Apr-21	49868.53	50092.48	49478.53	50029.83
05-Apr-21	50020.91	50028.67	48580.8	49159.32
06-Apr-21	49441.13	49582.26	48936.35	49201.39
07-Apr-21	49277.09	49900.13	49093.9	49661.76
08-Apr-21	49885.26	50118.08	49581.61	49746.21
03-May-21	48356.01	48863.23	48028.07	48718.52
04-May-21	48881.63	48996.53	48149.45	48253.51
05-May-21	48569.12	48742.72	48254.32	48677.55
06-May-21	48877.78	49011.31	48614.11	48949.76
07-May-21	49169.14	49417.64	49036.38	49206.47
01-Jun-21	52067.51	52228.65	51808.88	51934.88
02-Jun-21	51749.1	51913.92	51450.58	51849.48
03-Jun-21	52121.58	52273.23	51942.2	52232.43
04-Jun-21	52367.52	52389.02	51952.7	52100.05
07-Jun-21	52231.38	52378.69	52054.76	52328.51
01-Jul-21	52638.5	52638.5	52281.01	52318.6
02-Jul-21	52434.55	52527.9	52177.68	52484.67
05-Jul-21	52682.89	52919.17	52604.35	52880
06-Jul-21	52874.85	53129.37	52804.18	52861.18
07-Jul-21	52919.71	53105.41	52751.76	53054.76

Further, the Vaccination data is also collected from the MoHFW portal for the same time span as the stock market. (MoHFW 2021). Figure 1 shows a daily trend for vaccine dose graph, with the x-axis representing the date and the y-axis representing persons who received one dose, two doses, or the total dose.

Moreover, vaccination trend by age and vaccination trend by category graphs are provided in Figs. 2 and 3, respectively. The category graphs show the gender distribution, and the different types of vaccines used.

In addition to that, data for India's death rate is gathered via the MoHFW portal (MoHFW 2021) for the same time period as vaccination and stock market. Figure 4 shows the mortality rate of India where the date is represented in x-axis and death rate is given in y-axis.

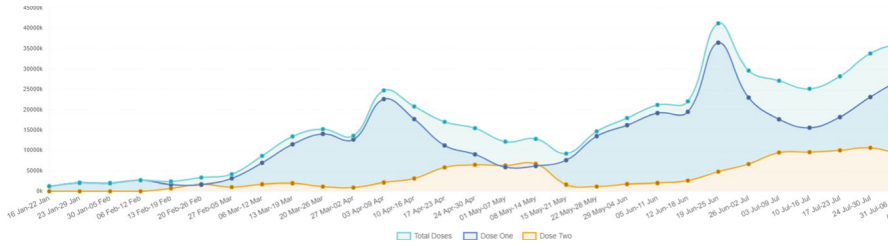


Fig. 1 Daily trend for vaccination dose in India

Fig. 2 Vaccination trend by age

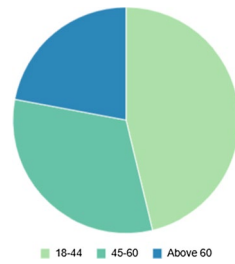


Fig. 3 Vaccination trend by category

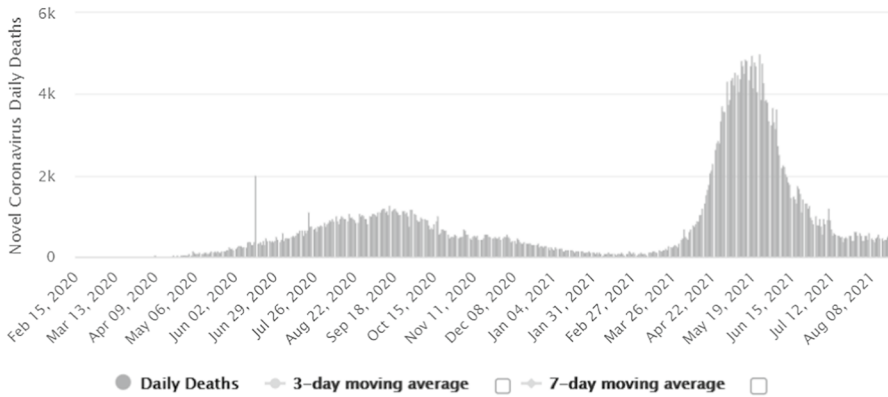
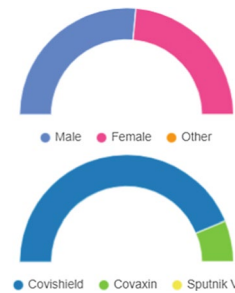


Fig. 4 Daily death rate due to COVID-19 in India

2.2 Methodology

This section includes Exploratory Data Analysis, Statistical Data Analysis, and Machine Learning-based prediction process for achieving the objectives. The EDA part depicted a statistical picture of the data distribution and relevant feature analysis; statistical data analysis shows the descriptive statistics and correlation, whereas the Prediction Process section detailed the technique for objective prediction.

2.2.1 Exploratory Data Analysis (EDA)

EDA is a type of statistical data analysis that focuses on graphical features (Abt, 1987). The fundamental goal of EDA is to gain a deeper understanding of a dataset’s hidden pieces of information. The relationship between features, arrangement, distribution, and stability of the feature association with the target variables are all advantages of EDA. It also illuminates the dataset’s key properties and statistical characteristics, such as the existence of missing values, distribution types such as skewness and kurtosis, class disproportion, and so on. However, if EDA is not performed properly, it can misguide a problem. Also, when EDA is dealt with high-dimensional data, it is not effective.

Descriptive statistics is a summary of a dataset using either the full dataset or a selection of the dataset. Measures of central tendency (mean, median, and mode), measures of variability (spread), and skewness and kurtosis may all be used to do descriptive statistics. The mean, median, and mode are calculated using central tendency measurements. The standard deviation, variance, minimum and maximum values are determined by the measures of variability, while the skewness and kurtosis define the data distribution. Thus, descriptive statistics are used to summarize quantitative data as shown in Fig. 5. There are 119 counts or samples in each column, with mean values that are suitable for the specified minimum and maximum values.

The correlation of two numerical variables is measured by statisticians and data analysts in order to get insight into their relationships. A correlation matrix is a matrix formed by the set of correlation values between pairs of characteristics in a dataset with numerous attributes. Figure 6 shows the correlation matrix on a heatmap. From the figure we can derive the insights that there is a high positive correlation between the total vaccination and closing price of the stock market, while negative correlation exists between the total vaccination and death rate.

	Open	High	Low	Close	Total vaccinated	Dose-1	Dose-2	total_cases	new_cases	total_deaths	new_deaths
count	119.000000	119.000000	119.000000	119.000000	1.190000e+02	1.190000e+02	1.190000e+02	1.190000e+02	119.000000	119.000000	119.00
mean	50918.747899	51169.992857	50543.260336	50872.01563	1.693540e+08	1.371253e+08	3.222863e+07	2.063256e+07	107837.823529	262557.798319	1476.22
std	1578.451966	1465.756003	1626.129726	1524.60054	1.389707e+08	1.104822e+08	2.882137e+07	8.540370e+06	119261.313104	106179.433035	1564.65
min	46617.950000	48020.790000	46433.650000	47883.38000	3.950156e+06	3.950156e+06	0.000000e+00	1.076624e+07	0.000000	154486.000000	0.00
25%	49755.825000	50039.495000	49237.140000	49626.54000	3.400598e+07	2.797076e+07	6.035224e+06	1.142428e+07	24687.000000	158950.000000	139.00
50%	50922.320000	51256.550000	50539.920000	50889.76000	1.577508e+08	1.286279e+08	2.912283e+07	2.066498e+07	48698.000000	226188.000000	780.00
75%	52455.870000	52588.665000	52047.635000	52336.48000	2.606481e+08	2.124981e+08	4.815001e+07	2.973155e+07	139769.000000	382696.500000	2748.50
max	53244.400000	53290.810000	52997.090000	53158.85000	4.615185e+08	3.598203e+08	1.016982e+08	3.161359e+07	414188.000000	423810.000000	7374.00

Fig. 5 Quantitative descriptive statistics

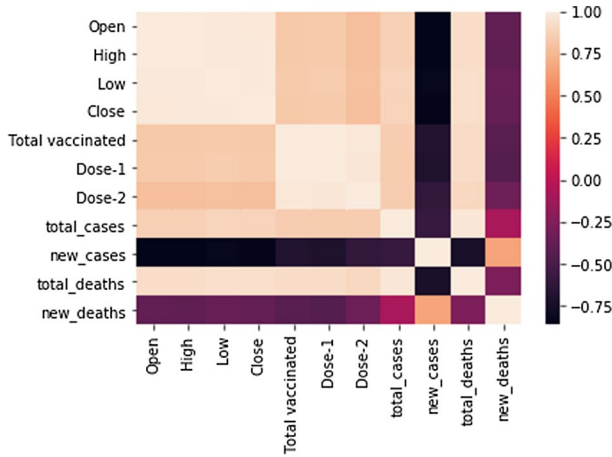


Fig. 6 Correlation matrix between vaccination, stock market, and death rate

Pairplot visualizes continuous or categorical data to discover the relationship between them. Figure 7 shows the pairplot of total vaccinated people with stock

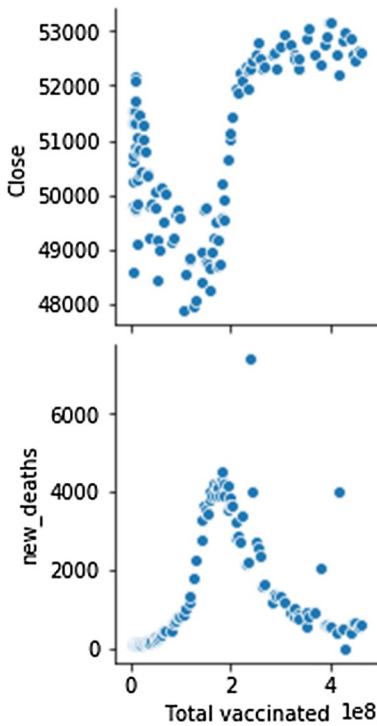


Fig. 7 Pairplot of total vaccination with closing price and death rate

index closing price and new deaths per day. From the figure, it can be observed that the closing price of the stock index is directly proportional to total vaccination while a new deaths per day is inversely proportional to total vaccination.

2.2.2 Statistical Data Analysis

Statistical analysis helps the researchers to quantify a wide range of phenomena, allowing them to study topics like social behaviour, political opinions, cellular biology, and forest growth rates objectively. On the other hand, statistical analysis is far from perfect and can lead to serious errors and false conclusions. To begin with, a statistical test is only as good as the data that it analyses. The statistical analysis will be misleading if researchers collect data using faulty or biased procedures. Another issue with statistical analysis is the proclivity for drawing rash conclusions about causal relationships. Researchers frequently discover evidence of high correlation between two variables, but this does not prove that one variable causes the other. In addition to that, statistical data is frequently secondary data, which makes it easy to misinterpret.

In this section, the relationship of vaccination with the stock market, as well as with the death rate in India, has been tested in two sections. According to that, for both the sections, we have defined two hypotheses as Null hypothesis (H_0) and Alternative hypothesis (H_a).

For vaccination and stock market:

- $H_0 \equiv$ There will be no significant relationship between vaccination and the stock market in India.
- $H_a \equiv$ There will be a significant relationship between vaccination and stock market in India.

For vaccination and death rate:

- $H_0 \equiv$ There will be no significant relationship between vaccination and the death rate in India.
- $H_a \equiv$ There will be a significant relationship between vaccination and the death rate in India.

There are two variables in each section, and all are continuous variables. For the vaccination and stock market section, variables are total vaccinated people and the closing price of the stock market where for vaccination and death rate section, the variables are total vaccinated people and daily new death rate in India. Moreover, in this study, Pearson correlation has been used to check the relationship between the variables. For both the sections, descriptive statistics and correlations of the variables are analyzed with Statistical Package for the Social Sciences (SPSS) software.

Fig. 8 Graphical representation of the Machine learning based methodology

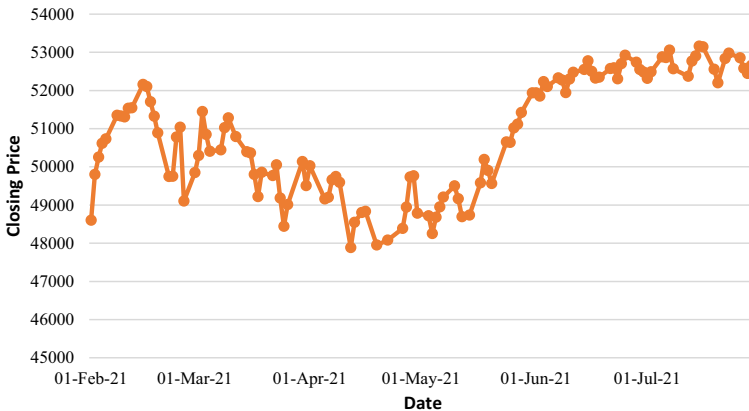
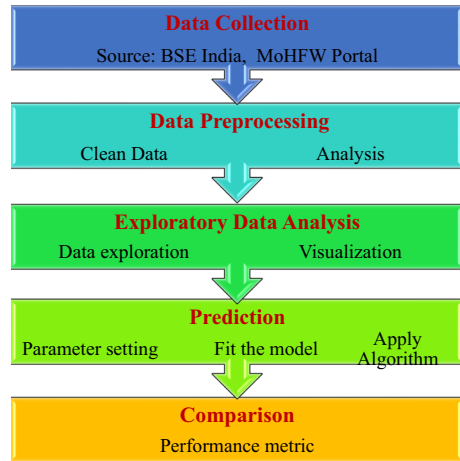


Fig. 9 Date wise closing price of S&P BSE SENSEX

2.2.3 Machine Learning based Prediction Process

Machine Learning is an important part of today’s technology. Many tasks such as pattern recognition and data prediction would be impossible without it. Moreover, Machine Learning algorithms are good at dealing with multidimensional and multivariate data, and they can do it in dynamic or uncertain environments. Also, as the amount of data increases, algorithms learn to make more accurate predictions in less time. However, Machine learning requires large dataset for training, which must be inclusive, unbiased, and of high quality. In addition to that, running models with different algorithms and determining the most accurate algorithm based on the results is a manual and time-consuming task. Also, it is highly susceptible to errors.

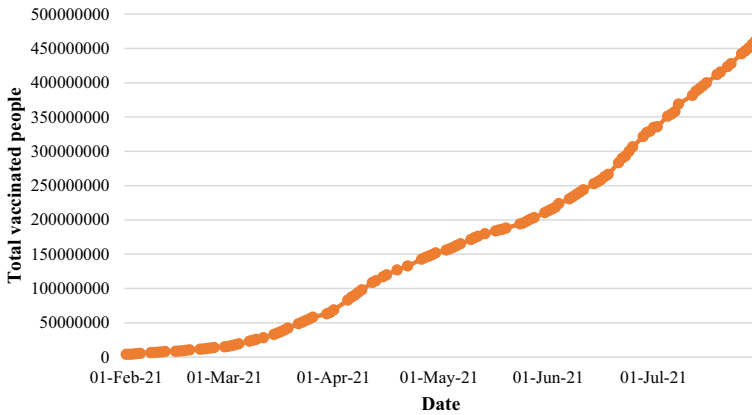


Fig. 10 Date wise vaccinated people in India

To predict the status of the Indian stock market and death rate with concern to COVID-19 vaccination, Machine Learning models like Linear regression, Support Vector Regression, KNN Regression, and Random Forest Regression are used in this research. The conceptual framework for Machine Learning prediction is depicted in Fig. 8.

Furthermore, to ensure consistency and completeness in the data set, stock index data is appended with COVID-19 vaccination data in one part, and in other part, the death rate data is appended with the vaccination data.

Figures 9 and 10 show the daily rate of stock market index and COVID vaccinated people, respectively.

In addition to that, Regression analysis is a collection of statistical approaches for estimating relationships between one or more independent variables and a dependent variable (Pasayat et al., 2020). Hence, in the stock market and vaccination section, the closing price of the stock index is considered to be a dependent variable, while others are supposed to be independent variables, and in the vaccination and death rate section, new death cases are assumed as dependent variables. Then, we divide our dataset into two parts as training dataset and testing dataset with 70% and 30% proportion respectively.

The models are trained using the training data set to predict the stock market closing price and mortality rate of India. After training, each of the models are tested with 30% testing data set to find out their performance level. Further, to estimate the accuracy and error, R^2 value, mean squared error, and mean absolute error have been used in each of the forecasting models.

- **R^2 Value** The coefficient of determination, often known as R-squared, is the proportion of the variance in the dependent variable. It is a scale-free score, meaning that regardless of how little or huge the values are, the R square value will be less than one. The formula to compute this is given in Eq. 1 (Liong & Sivapragasam, 2002).

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y})^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \tag{1}$$

where \hat{y} = predicted value of y ; \bar{y} = mean value of y .

- **Mean Squared Error (MSE)** The squared difference between the original and forecasted values in the dataset is represented by the Mean Squared Error. The variance of the residuals is calculated using the formula provided in Eq. 2 (Zheng et al., 2015).

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2 \tag{2}$$

- **Mean Absolute Error (MAE)** The average of the absolute difference between the actual and predicted values in the dataset is represented by the Mean absolute error. It computes the average of the dataset’s residuals by the formula represented in Eq. 3 (Lian et al., 2013).

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}| \tag{3}$$

Descriptive Statistics

	Mean	Std. Deviation	N
Total vaccinated	169353954.6	138970712.1	119
Close	50872.0156	1524.60054	119

Correlations

		Total vaccinated	Close
Total vaccinated	Pearson Correlation	1	.627**
	Sig. (2-tailed)		<.001
	N	119	119
Close	Pearson Correlation	.627**	1
	Sig. (2-tailed)	<.001	
	N	119	119

** . Correlation is significant at the 0.01 level (2-tailed).

Fig. 11 Statistical data analysis between vaccination and stock market

3 Result Analysis

This section empirically demonstrated the impact of COVID-19 immunization on the stability of the Indian stock market and the decrease of the mortality rate. The data on COVID-19 vaccination, stock market index and mortality rate in India for the same time period are gathered. Here, we calculate both the statistical-based correlation among the variables and the machine learning based-prediction.

3.1 Statistical Data Analysis Results

From the descriptive statistics of Fig. 11, we found the mean and standard deviation of the variables. In both the cases, sample size (N) is 119, which means any data has not been missed. Moreover, the correlation of each combination of the variables is given in the correlation matrix of Fig 11. When the total vaccinated people is correlated to the closing price of the stock market, the coefficient of the correlation (r) is 0.627. Also, the level of significance (α) in the correlation matrix is less than 0.001, which proves that the correlation between the total vaccinated people and the closing price of the stock market is significant.

Since the r value is positive, the total vaccinated people and closing price are positively correlated. In other words, we can say that if the total vaccination increases, then the closing price of the stock market is also increases. As we know, if the r value is > 0.5 , then the relationship is strong, and in the correlation matrix of Fig. 11, our r value is 0.627, which shows that total vaccination and closing price are strongly correlated. Hence, the null hypothesis got rejected here. In addition to that,

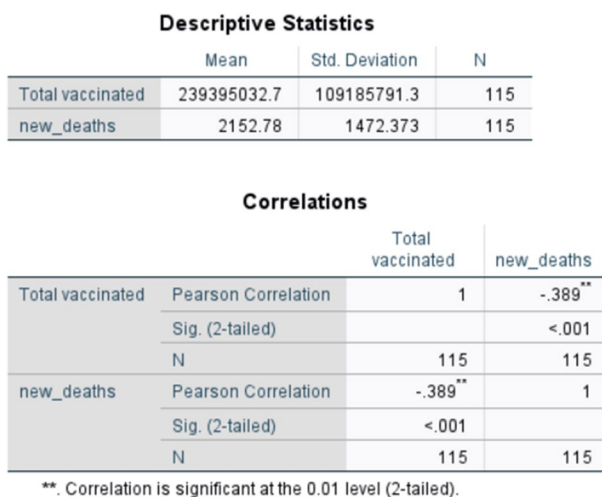


Fig. 12 Statistical data analysis between vaccination and death rate

$$\text{Coefficient of determination} = (r)^2 \times 100 \quad (4)$$

which tells how much variance the two variables share. Thus, here coefficient of determination is 39.3129. That means total vaccination and closing price of the stock market have 39.3129% of shared variance.

Furthermore, in Fig. 12, descriptive statistics and correlation of the variable in vaccination and death rate section are illustrated.

From the descriptive statistics, it can be observed that mean of the variables of total vaccinated people and death rate are 239,395,032.7 and 2152.78, respectively, whereas the standard deviation of the same are 109, 185, 791.3 and 1472.373. Also, for both the cases, $N = 115$, which means no data has been missed. Furthermore, from the correlation matrix of Fig. 12, it can be observed that, coefficient of correlation (r) = -0.389 when total vaccinated people is correlated with the new death rate. As the value of r is negative, total vaccinated people and death rate are negatively correlated, which means if total vaccination increases, then death rate decreases accordingly. Moreover, the level of significance is less than 0.001, which shows that the correlation between the variables is significant.

In addition to that, for the case vaccination and death rate, coefficient of determination is 15.1321, which states that the total vaccination and death rate have 15.1321% of shared variance.

3.2 Machine Learning based Prediction

Here, Machine Learning models are being used to investigate the link between the stock market and vaccination. To evaluate the performance of the models, metrics like R^2 score, mean squared error (MSE) and, mean absolute error (MAE) are taken into consideration. After model implementation, when the actual and predicted values are known, we have calculated the above-mentioned metrics. The accuracy of the model's performance is calculated by R^2 score, whereas the error in the performance is estimated by MSE and MAE.

Table 2 Percentage of various metrics score

Model	R^2 score	MSE	MAE
Linear regression	72	0.02	0.102
Support vector regression	91	0.006	0.058
Random forest regression	89	0.004	0.071
KNN regression	82	0.011	0.072

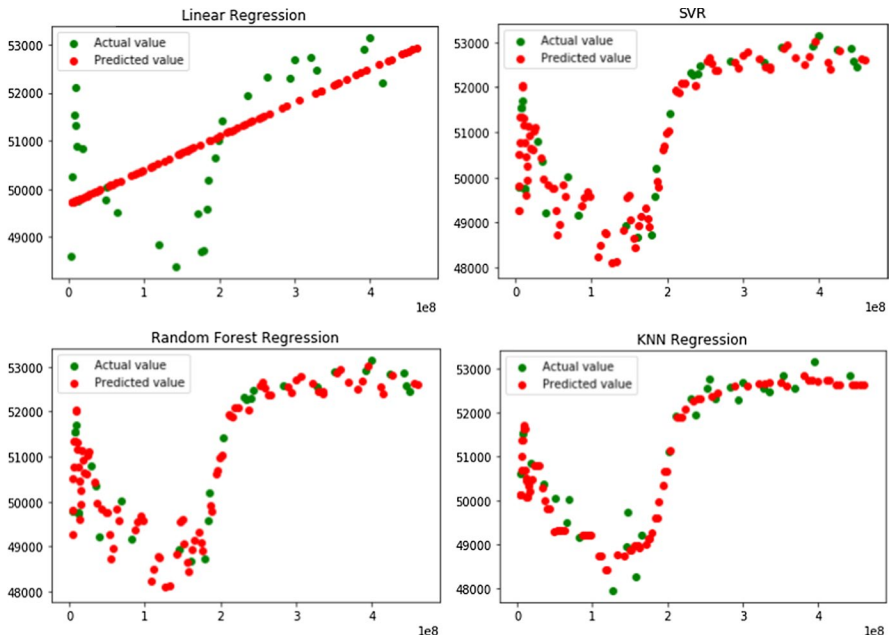


Fig. 13 Estimation vs actual values for each regression method in stock market regarding vaccination process

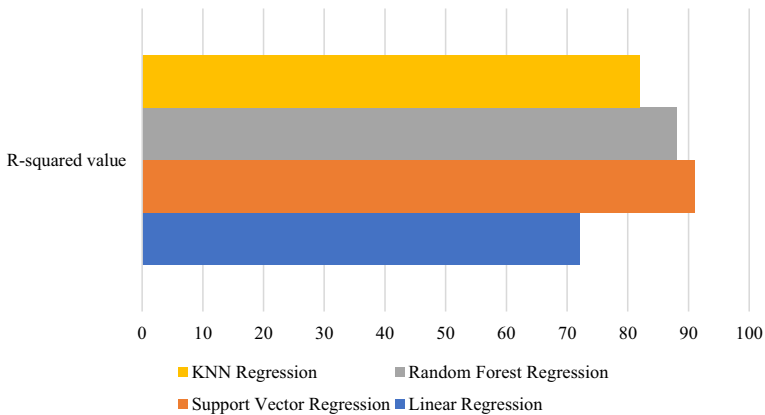


Fig. 14 R^2 value comparison between the models for vaccination and stock market

3.2.1 Vaccination and Stock market

The estimated values of the metrics in the prediction process of stock market regarding to Vaccination process are represented in Table 2.

From Table 2, the result found that, Linear Regression, Support Vector Regression, Random Forest Regression and KNN Regression models give accuracy (R^2

Table 3 Percentage of various metrics score

Model	R^2 score	MSE	MAE
Linear regression	70	0.018	0.105
Support vector regression	87	0.009	0.058
Random forest regression	89	0.078	0.067
KNN regression	85	0.013	0.079

score) of about 72%, 91%, 88%, and 82%, respectively, whereas the MSE are 0.02, 0.006, 0.004, and 0.011 and MAE of the models are 0.102, 0.058, 0.071 and 0.072, respectively.

The graph of actual value vs predicted value of different regression models in stock market concerned to vaccination process has been illustrated in Fig. 13. It can be seen that Support Vector Regression model gives the better prediction followed by Random Forest Regression, KNN regression and Linear Regression.

Comparison between the performances of the models is shown in Fig. 14. From this figure, it is observed that the Support Vector Regression model gives better accuracy of 91% followed by Random Forest Regression with 88%, KNN Regression with 82%, and Linear Regression with 72%. Despite the fact that the Random Forest Regression model has the second-highest accuracy, the difference between

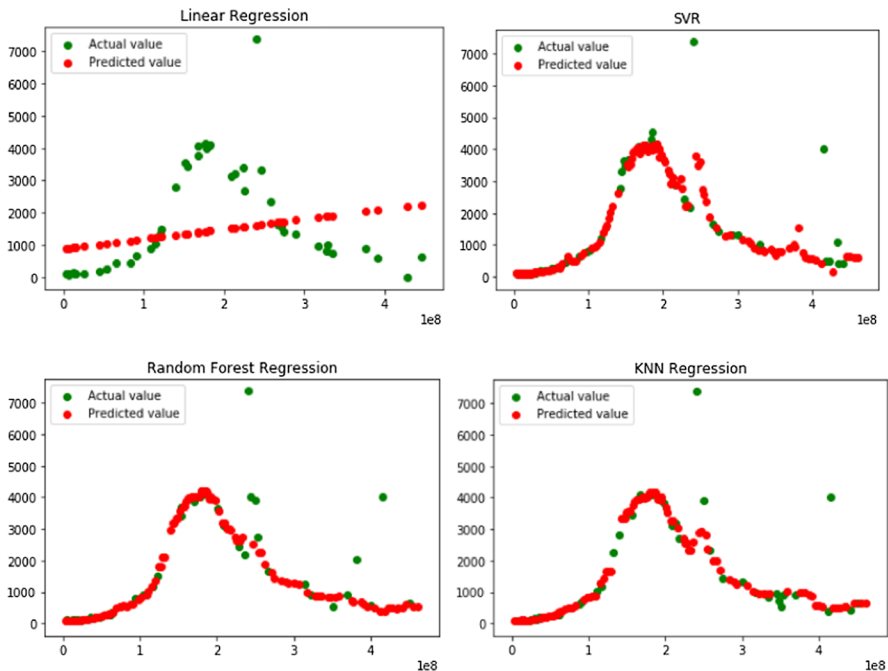


Fig. 15 Estimation vs actual values for each regression method in death rate regarding vaccination process

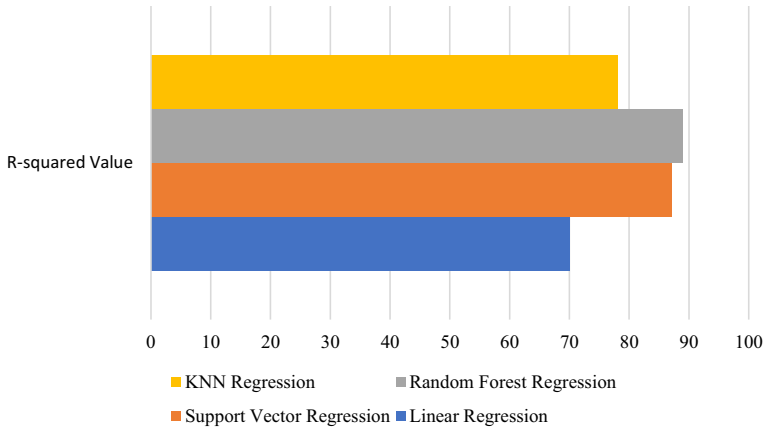


Fig. 16 R^2 value comparison between the models for vaccination and death rate

the Support Vector Regression and Random Forest model is small, implying that the Regression model can be used for stock market prediction in the future for a better comparison.

3.2.2 Vaccination and Death rate

Moreover, the estimated values of the metrics in the death rate prediction with concern to COVID-19 vaccination are provided in Table 3.

According to Table 3, the accuracy (R^2 score) of Linear Regression, Support Vector Regression, Random Forest Regression, and KNN Regression models is about 70%, 87%, 89%, and 78%, respectively, while the MSE is 0.018, 0.009, 0.078, and 0.013, and the MAE is 0.105, 0.058, 0.067, and 0.079, respectively.

Figure 15 depicts the graph of actual value vs. predicted value for various regression models in death rate regarding to vaccination process. As can be seen, the Random Forest Regression model outperforms the Support Vector Regression, KNN Regression, and Linear Regression models.

Figure 16 shows a comparison of the performance of the models. The Random Forest Regression model has a higher accuracy of 89%, followed by Support Vector Regression with 87%, KNN Regression with 78%, and Linear Regression with 70%, as seen in the graph.

4 Discussion

The global economy, where India is a large contributor, has been affected by the COVID-19 pandemic. Almost all the stock markets around the world have been affected by the COVID-19 pandemic. It pushed the world economy into a great crisis. Since the population of India is so large, the pandemic becomes more dangerous

to it. However, the delivery of the vaccination raises some hopes in both the stock market and the death rate.

This study has two purposes; one is to predict the stock market response to the vaccination process, and the second is to find the relationship between the mortality rate and vaccination in India. EDA has been employed to analyze the key features of the dataset with visual methods. Moreover, statistical data analysis has been used to demonstrate the relationship of vaccination with the stock market and death rate. As a result, it is found in vaccination and stock market part that total vaccinated people and closing price of the stock market are positively correlated, since r value is positive. In addition to that, as the value of $r > 0.5$, vaccination and stock market are strongly correlated. At the same time, in the vaccination and death rate section, vaccination process and death rate are negatively correlated as the value of r is negative.

Furthermore, different Machine Learning Regression models have been implemented to predict the stability of the stock market and decrease in death rate after vaccination process. The study reveals that in the part of stock market and vaccination, the Support Vector Regression model surpassed all other models in terms of accuracy, with an R^2 score of 91% and excellent prediction outcomes whereas in death rate and vaccination section, the highest accuracy produced by the Random Forest Regression model is 89%. The result has met our expectations thoroughly by giving the prediction of the stock market and mortality rate concerning the vaccination data.

The result summarized in the above section shows that increasing vaccination process decrease the volatility and chaos in the stock market. Moreover, with an increase in the number of daily vaccination, decrease the death rate in India. Therefore, the reaction of the mass vaccination process in India can be seen positively in the stock market with less mortality rate.

In addition to that, we have included some policy recommendation taken by Government in Introduction section. To summarise, the Reserve Bank took a number of steps to boost confidence in the financial markets and the economy. These efforts not only maintained orderly market conditions, but they also hastened the rate of recovery, as evidenced by high frequency economic indicators. The unwinding of some of the policy measures implemented in the aftermath of the pandemic necessitates a measured and deliberate approach. Financial market movements will be led in the future by progress in managing the COVID-19 pandemic, particularly through vaccine delivery, the pace of global and domestic economic recovery, and changes in global liquidity and financial conditions (Reserve Bank of India, 2021).

5 Conclusion

The Coronavirus has spread throughout India, posing serious health and economic threat. Though India is not as badly impacted as many other Eurozone countries, the disease's effects have made India no exception to the predicament. Nevertheless, in early 2021, the development and widespread distribution of a vaccine raised expectations of reviving the economy and stabilizing the stock market with reducing the death rate. This paper examines the stock market index value and mortality rate with

concern to COVID-19 vaccination in India. The link of immunization with the stock market and the death rate has been demonstrated using statistical data analysis. For the Machine Learning process, Linear Regression, Support Vector Regression, Random Forest Regression, and KNN Regression model has been implemented in this work. It is found that the stock market becomes stable after the introduction of the COVID-19 vaccination process. Furthermore, daily vaccination in the country also decreases the mortality rate. From the experiment, it is concluded that if the vaccination process goes on accordingly, better improvement can be seen in the stock market with a lower death rate. The limitation of our work is the relatively fresh and short data set.

In future studies, researchers may consider other indicators like the number of tests and the number of patients who recover. Also, a micro-level (i.e. company-based) analysis, as well as a macro-level (i.e. main stock exchange indices) analysis, could be used to examine the reaction to the vaccination process. Other statistical or econometric methods, such as Deep Learning Algorithms, Markov Switch Regression, Neural Networks, Multivariate Adaptive Regression Splines and Wavelet Coherence Approach, could be used in future research. In addition, the effect of vaccination on the stock market in other countries could be investigated. Moreover, new analyses of the stock market's reaction to the vaccination process, based on full data from the beginning to the end of the vaccination, could add to the existing literature. The findings of these studies could be helpful in formulating policies to mitigate the pandemic's negative effects on economies, financial markets, stock exchanges, and major indices. Furthermore, future research on this topic could also look into the effects of vaccines on other asset classes, such as currencies, commodities, and corporate bonds.

Funding This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

- Abt, K. (1987). Descriptive data analysis: a concept between confirmatory and exploratory data analysis. *Methods of Information in Medicine*, 26(02), 77–88.
- Acharya, V. V., Johnson, T., Sundaresan, S., & Zheng, S. (2020). *The value of a cure: An asset pricing perspective. Technical report*. National Bureau of Economic Research
- Alam, M. N., Md Shabbir, A., & Chavali, K. (2020). Stock market response during covid-19 lockdown period in india: An event study. *The Journal of Asian Finance, Economics, and Business*, 7(7), 131–137.
- Al-Awadhi, A. M., Alsaifi, K., Al-Awadhi, A., & Alhammadi, S. (2020). Death and contagious infectious diseases: Impact of the covid-19 virus on stock market returns. *Journal of behavioral and experimental finance*, 27, 100326.
- Ashraf, B. N. (2020). Stock markets' reaction to covid-19: Cases or fatalities? *Research in International Business and Finance*, 54, 101249.
- Baber, H., & Rao, D. T. (2021). What is the price of lockdown? the effects of social distancing on Indian economy and business during covid-19 pandemic. The Effects of Social Distancing on Indian Economy and Business During COVID-19 Pandemic (February 23, 2021). *Economic Horizons*.

- Barbate, V., Gade, R. N., & Raibagkar, S. S. (2021). Covid-19 and its impact on the Indian economy. *Vision*, 25(1), 23–35.
- Binrong, W., Wang, L., Wang, S., & Zeng, Y.-R. (2021). Forecasting the us oil markets based on social media information during the covid-19 pandemic. *Energy*, 226, 120403.
- BSE India. <https://www.bseindia.com/>, 2021. Accessed on August 1, 2021.
- Chan, K. Fong, Chen, Z., Wen, Y., & Xu, T. (2021). Covid-19 vaccines: Saving lives and the global stock markets. Available at SSRN 3785533
- Conlon, T., & McGee, R. (2020). Safe haven or risky hazard? bitcoin during the covid-19 bear market. *Finance Research Letters*, 35, 101607.
- Dasgupta, A., Sun, Y. V., König, I. R., Bailey-Wilson, J. E., & Malley, J. D. (2011). Brief review of regression-based and machine learning methods in genetic epidemiology: the genetic analysis workshop 17 experience. *Genetic Epidemiology*, 35(S1), S5–S11.
- Demir, E., Bilgin, M. H., Karabulut, G., & Doker, A. C. (2020). The relationship between cryptocurrencies and covid-19 pandemic. *Eurasian Economic Review*, 10(3), 349–360.
- Dev, S. M., & Sengupta, R. et al. (2020) Covid-19: Impact on the Indian economy. *Indira Gandhi Institute of Development Research, Mumbai April*
- Du, B., Zhao, Z., Zhao, J., Yu, L., Sun, L., & Lv, W. (2021). Modelling the epidemic dynamics of covid-19 with consideration of human mobility. *International Journal of Data Science and Analytics*. <https://doi.org/10.1007/s41060-021-00271-3>.
- Falato, A., Goldstein, I., & Hortaçsu, A. (2020). *Financial fragility in the covid-19 crisis: The case of investment funds in corporate bond markets*. National Bureau of Economic Research: Technical report.
- Feinhandler, I., Cilento, B., Beauvais, B., Harrop, J., & Fulton, L. (2020). Predictors of death rate during the covid-19 pandemic. In: *Healthcare*, vol. 8, p. 339. Multidisciplinary Digital Publishing Institute.
- Franco-Lopez, H., Ek, A. R., & Bauer, M. E. (2001). Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbors method. *Remote sensing of Environment*, 77(3), 251–274.
- Gharib, C., Meftah-Wali, S., & Jabeur, S. B. (2021). The bubble contagion effect of covid-19 outbreak: Evidence from crude oil and gold markets. *Finance Research Letters*, 38, 101703.
- Khattak, M. A., Ali, M., & Rizvi, S. A. (2021). Predicting the european stock market during covid-19: A machine learning approach. *MethodsX*, 8, 101198.
- Lian, C., Zeng, Z., Yao, W., & Tang, H. (2013). Displacement prediction model of landslide based on a modified ensemble empirical mode decomposition and extreme learning machine. *Natural Hazards*, 66(2), 759–771.
- Ling, D. C., Wang, C., & Zhou, T. (2020). A first look at the impact of covid-19 on commercial real estate prices: Asset-level evidence. *The Review of Asset Pricing Studies*, 10(4), 669–704.
- Liong, S.-Y., & Sivapragasam, C. (2002). Flood stage forecasting with support vector machines 1. *JAWRA Journal of the American Water Resources Association*, 38(1), 173–186.
- Mensi, W., Sensoy, A., Vo, X. V., & Kang, S. H. (2020). Impact of covid-19 outbreak on asymmetric multifractality of gold and oil prices. *Resources Policy*, 69, 101829.
- Milcheva, S. (2021). Volatility and the cross-section of real estate equity returns during covid-19. *The Journal of Real Estate Finance and Economics*, 8, 1–28.
- MoHFW. Vaccination details in india. <https://www.mohfw.gov.in/>, 2021. Accessed on August 1, 2021.
- Noi, P. T., Degener, J., & Kappas, M. (2017). Comparison of multiple linear regression, cubist regression, and random forest algorithms to estimate daily air surface temperature from dynamic combinations of modis 1st data. *Remote Sensing*, 9(5), 398.
- Pasayat, A. K., Pati, S. N., & Maharana, A. (2020). Predicting the covid-19 positive cases in India with concern to lockdown by using mathematical and machine learning based models. *MedRxiv*
- Pegu, G. (2020). Covid-19 pandemic, lockdown, and Indian economy. *International Journal of Management*, 11(10), 5.
- Reserve Bank of India. (2021). <https://rbi.org.in/scripts/AnnualReportPublications.aspx?Id=1315>.
- Roghani, A. (2021). The influence of covid-19 vaccine on daily cases, hospitalization, and death rate in tennessee: A case study in the united states. *medRxiv*
- Rouatbi, W., Demir, E., Kizys, R., & Zaremba, A. (2021). Immunizing markets against the pandemic: Covid-19 vaccinations and stock volatility around the world. Available at SSRN 3815000
- Sharif, A., Aloui, C., & Yarovaya, L. (2020). Covid-19 pandemic, oil prices, stock market, geopolitical risk and policy uncertainty nexus in the us economy: Fresh evidence from the wavelet-based approach. *International Review of Financial Analysis*, 70, 101496.

- Shataee, S., Kalbi, S., Fallah, A., & Pelz, D. (2012). Forest attribute imputation using machine-learning methods and aster data: Comparison of k-nn, svr and random forest regression algorithms. *International Journal of Remote Sensing*, 33(19), 6254–6280.
- Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, 14(3), 199–222.
- Song, W., & Yi, Y. (2021). Pin20 predicting the impact of vaccination strategies in the covid-19 pandemic using a susceptible-exposed-infectious-removed (seir) model. *Value in Health*, 24, S108.
- Topcu, M., & Gulal, O. S. (2020). The impact of covid-19 on emerging stock markets. *Finance Research Letters*, 36, 4.
- Wang, C., Horby, P. W., Hayden, F. G., & Gao, G. F. (2020). A novel coronavirus outbreak of global health concern. *The Lancet*, 395(10223), 470–473.
- Wang, L., Li, J., Guo, S., Xie, N., Yao, L., Cao, Y., et al. (2020). Real-time estimation and prediction of mortality caused by covid-19 with patient information based algorithm. *Science of the Total Environment*, 727, 138394.
- Wu, J. T., Leung, K., & Leung, G. M. (2020). Nowcasting and forecasting the potential domestic and international spread of the 2019-ncov outbreak originating in wuhan, china: a modelling study. *The Lancet*, 395(10225), 689–697.
- Yadav, D., et al. (2021). Covid19 impact ontwelve sectors of the Indian economy. *European Journal of Molecular & Clinical Medicine*, 7(8), 5660–5670.
- Zheng, Y., Rajasegarar, S., & Leckie, C. (2015). Parking availability prediction for sensor-enabled car parks in smart cities. In *2015 IEEE Tenth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*, pages 1–6. IEEE

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.