

Assessing the Influence of the Covid 19 Pandemic on Indian Pharmaceutical Companies

Atishay Jain, Naman Rastogi, Sushma Jain



Abstract: Our objective is to study, analyze and draw inferences on the movement of the stock prices of Indian pharmaceutical companies solely based on the COVID-19 pandemic in India. We specifically targeted pharmaceutical stocks because their share price is more directly dependent on the COVID-19 pandemic than companies in other sectors. As the demand for the primary products sold by pharmaceutical companies, i.e., medicines, is directly dependent on the COVID-19 pandemic, a common hypothesis is that the stock prices of pharmaceutical companies at a given time are significantly contingent upon the COVID-19 pandemic situation. We have tested this hypothesis by calculating the correlation between pharmaceutical stock prices and COVID-19 variables that measure the severity and provide an outline of the COVID-19 pandemic. The COVID-19 variables we have considered provide information regarding covid cases, deaths, testing, vaccination, positivity rate, virus reproduction rate, and government restrictions to counter the spread of the virus. Furthermore, as human emotion plays a significant part in deciding the share prices, we have considered public fear and awareness by considering the frequency by which the terms "Covid 19" and "Covid medicines" are searched on Google. We have considered the stock prices of 19 companies that contribute to the Nifty Pharma index as the target values upon which the impact of the COVID-19 pandemic is tested. We have selected the covid fields that have the most significant impact on the pharmaceutical stock prices and then calculated the correlation between the Covid fields and the stock parameters. The data we have considered for our study belongs to the period from 15th March 2020 to 17th February 2022. Link to the Github repository with the code for the presented research: <https://github.com/Atishaysjain/Predictive-Analysis-Project>.

Keywords: Pearson Correlation, COVID-19 Pandemic, Indian Stock Market, Comparative Analysis.

I. INTRODUCTION

A country's stock market is considered a reflection of that nation's overall economy. In normal circumstances, if the stock market performs well, the economy booms, unemployment decreases, and average household wealth increases. However, Since the start of the COVID-19 pandemic, the response of the stock markets around the world has not been commensurate with the economies of their respective countries.

Manuscript received on 31 July 2022 | Revised Manuscript received on 07 August 2022 | Manuscript Accepted on 15 August 2022 | Manuscript published on 30 August 2022.

* Correspondence Author

Atishay Jain*, SCOPE, Vellore Institute of Technology, Vellore (Tamil Nadu), India. Email: atishay.jain2019@vitstudent.ac.in

Naman Rastogi, SCOPE, Vellore Institute of Technology, Vellore (Tamil Nadu), India. Email: naman.rastogi2019@vitstudent.ac.in

Sushma Jain, Thapar Institute of Engineering and Technology, Patiala (Punjab), India. Email: sjain@thapar.edu

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

India witnessed a negative 8 percent change in its GDP in 2020. Nevertheless, NIFTY, the index of the National Stock Exchange of India, registered a growth of around 73 percent since the stock market crash in March 2020 till the end of 2020. Although many Indians lost their jobs, migrated to their home villages, and lost all their savings during the first two years of the COVID-19 pandemic, NIFTY still rose by 113 percent after the stock market crash from March 2020 to 18th February 2022.

The United States witnessed a similar situation. Dow Jones Industrial Average increased by 64 percent in 2020 since it crashed in March 2020, i.e., declaration of COVID-19 as a pandemic, while the US suffered a negative 3.41 percent GDP growth in the same year. Moreover, the Dow Jones climbed by 83 percent till 18th February 2022 after the Stock market crash of 2020, while many Americans lost their jobs during the same period.

The above numbers show that the traditional stock prediction models will fail in conjunction with the COVID-19 pandemic. Therefore, additional features describing the COVID-19 situation and public opinion on the COVID-19 scenario must be integrated into predicting models. Considerable work has been done to update stock predicting models by incorporating COVID-19 information.

To test the impact of the COVID-19 pandemic on the stock market, we have calculated the correlation between the pandemic and the stock market. We have limited our study to the pharmaceutical industry's niche sector as the pharmaceutical companies profits and revenues are directly dependent on the severity of the COVID-19 pandemic.

II. LITERATURE SURVEY

A modular Neural Network was utilized by [1] to predict Tokyo Stock Exchange Prices Indexes (TOPIX) in 1990. It was an accurate prediction system, and the stocks simulation showed a significant profit. Moreover, it advised people when to buy and sell stocks. Dow Jones Industrial Average index was predicted with an accuracy of 65 percent by [2]. The approach determined parameters for generating training patterns heuristically by auto-mutual information and false nearest neighbor methods. Then classifiers such as artificial neural networks, decision trees, and k-nearest neighbors were trained on the generated patterns. The goal was to show that not all periods are equally random. Some of them show better results as compared to others. The prediction accuracy over a long period was improved by [3]. The authors utilized 18 machine learning algorithms over three stock indexes to predict the closing price after 1,5 and 10 days, respectively.



They concluded that a fully convolutional neural network and extra trees classifier performed better for one day, whereas a Multi-level Perceptron performed better for 5-day and 10-day predictions. In addition, the 10-day forecast was more straightforward to calculate and can be used for long-term investment.

Bidirectional LSTM(BLSTM) and stacked LSTM(SLSTM) models were employed to predict stock market prices by [4]. Both of them are useful for short-term prediction. However, LSTM is preferred over traditional Recural Neural Networks(RNN) as it supports time steps of arbitrary sizes and does not have a vanishing gradient problem. The Standard & Poor 500 Index (S&P500) data from 1st January 2010 to 30th November 2017 from Yahoo finance was utilized. In addition, deep learning technology was preferred over shallow neural networks. Overall, BLSTM showed better performance for both short-term and long-term goals.

Many machine learning techniques like decision trees, random forest, adaptive boosting (Adaboost), gradient boosting, Xtreme gradient boosting(XGBoost), LSTM, etc., were employed by the [7]. The categories selected for experimental evaluations from the Tehran stock exchange were diversified financials, petroleum, non-metallic minerals, and primary metals. LSTM turned out to be the most accurate algorithm as it had a minor error and the best ability to fit.

Covid-19 has disrupted the stock market. A unique Covid19 stock market system(C19-SMS) was used by [5]. This system uses social media reports, Covid deaths, and cases in a particular area to extract sentiments to predict future stock prices. The extracted sentiments are utilized by SVM, random forest, and random tree classifier along with the current stock prices. Then stock prices for the following day are predicted. The approach has been tested using the Dow Jones Industrial Average (DJI) and the Tadawul All Share Index (TASI), returning an accuracy of 99.71%, in which the inclusion of Covid data increased the accuracy by 9.78%. This model can be improved by considering other factors directly influencing public emotion and sentiments.

[6] used the AutoRegressive Integrated Moving Average (ARIMA), the Generalized Auto Regressive Conditional Heteroscedasticity (GARCH), and the stacked Long Short Term Memory Deep Neural Network (LSTM DNN) to analyze and predict the daily number of Covid cases. Vaccination was also taken into account. Ten datasets were used: nine countries and one world dataset. LSTM DNN turned out to be the best compared to the other two models.

Fast Fourier Transform(FFT) curve fitting was used over Jakarta Stock Exchange (JKSE) by [8]. The paper covers descriptive statistics, multi-collinearity tests, hypothesis tests, determination tests, and prediction using FFT curve fitting. They used daily data of the Jakarta Stock Exchange (JKSE) Composite Index, interest rate, and exchange rate from 15th October 2019 to 15th September 2020 and recorded 224 observations. Data was collected from the Indonesia Stock Exchange (IDX), Indonesia Statistics Central Bureau, and Observation & Research of Taxation. They have used descriptive statistics, multi-collinearity tests, hypothesis tests, determination tests, and predictions using FFT curve fitting. The F-test result, interest rate, and

exchange rate have significantly affected the stock market index (JKSE). FFT curve predicted that fluctuations in the stock market would increase over some time. The coefficient of determination is 71.90%, which means that the independent variables simultaneously affect the dependent variable by 71.9% and the rest by other factors.

The Least Absolute Shrinkage and Selection Operator (LASSO) was used to predict 21 stock prices in the European market by [9]. [9] showed that LASSO is better than traditional regression models and is not susceptible to outliers and multi-collinearity. The European market was most affected by indices from countries like Spain, Germany, France, Singapore, Switzerland, and the S&P 500 index. However, there is a difference between the predictors before and after the pandemic. Before the pandemic, the European stock market was heavily affected by the gold market, EUR/USD exchange rate, Dow Jones index, and the stock markets of Switzerland, Spain, France, Italy, Germany, and Turkey. After the pandemic, France and Germany were the most significant predictors of the European Market.

Artificial Neural Network (ANN) has many limitations. So a hybrid model was proposed by [10], combining artificial intelligence networks and particle swarm optimization for better prediction. This model predicted the closing price of a particular stock for maximum profit and minimum risk. However, it did not give satisfactory results. So a system called PSOCoG was proposed. Identifying hyperparameter values with precision is necessary to design a neural network with minimal processing and search time and maximum prediction accuracy. PSOCoG was used to select the best hyperparameters to construct the optimal neural network. Finally, PSOCoG was combined with ANN to provide the best results. The model displayed better performance compared with ANN, SPSO, and ANN-SPSO models in terms of prediction accuracy.

[9] considered Stock market prediction as a binary problem with the objective to predict whether the closing price would be higher or lower than the opening price. [9] proposed two integrated frameworks of feature engineering, methodology to sort out the class imbalance, and predictive analyses to perform stock trend prediction. Kernel Principle Component Analysis (KPCA) is then used for feature engineering. Bootstrapping procedure is done to resolve class imbalance [11]. Then two models, Stacking and Deep Neural Network, are applied to the engineered and bootstrapped samples, respectively, to predict the trend of the stocks pre and post COVID-19 periods.

[12] concluded that LSTM could be used for the short-term prediction of stock prices(up to one year). [12] analyzed stock prices and covid cases of Indonesia. Data from the Bank of Central Asia (BCA) and Bank of Mandiri from Indonesia were collected from Yahoo finance. [12] calculated Open, High, Low, and Closing from various parameters.

[13] used SutteARIMA to predict US stock market behavior by first calculating the daily Covid cases. Data of confirmed Covid cases was collected from World meter. In addition, US stock market data(DJI) was collected from Yahoo finance. The Mean Absolute Percentage Error (MAPE) metric was employed for model evaluation. [13] concluded that SutteARIMA returned better results than ARIMA.

III. BACKGROUND

A. Covid Dataset

We have obtained the values of the COVID-19 statistics from the COVID-19 data repository of “Our World in Data”. The CSV file from the above-mentioned repository contained the values of numerous covid variables for all countries. We extracted the following covid variables for India: ‘total cases’, ‘new cases’, ‘new deaths’, ‘new tests’, ‘reproduction rate’, ‘positivity rate’, ‘total vaccinations’, ‘people vaccinated’, ‘people fully vaccinated’, ‘new vaccinations’ and, ‘stringency index’.

We then fetched the google trends data for the terms “covid 19” and “covid medicines” indicating the interest of the general population of India in the above-mentioned search queries.

B. Pharmaceutical Stock Dataset

We fetched the stock data of 19 of the 20 companies indexed in the Nifty Pharma index except for Gland Pharma Limited from the National Stock Exchange of India website. The CSV file of each of the companies contained the following attributes for each trading day: 'Prev Close', 'Open Price', 'High Price', 'Low Price', 'Last Price', 'Close Price', 'Average Price', 'Total Traded Quantity', 'Turnover', 'No. of Trades', 'Deliverable Qty', and '% Dly Qt to Traded Qty'. In addition, we fetched the combined equity data of Nifty Pharma index from “niftyindices.com”. Fig. 1 shows the Nifty Pharma index from 15th March 2020 to 17th February 2022.

Candlestick graph of NIFTY PHARMA

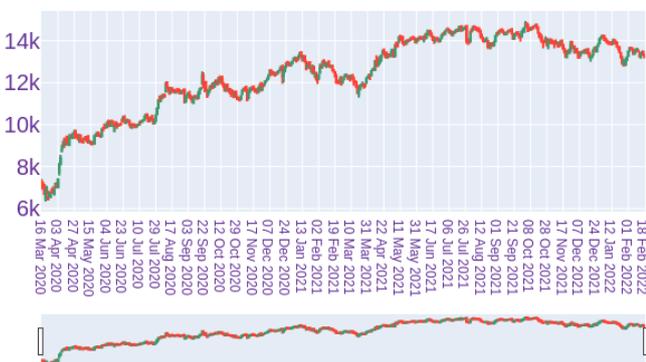


Fig. 1. Nifty Pharma Index from 15th March 2020 to 17th February 2022.

C. Pearson Correlation

Pearson Correlation indicates the strength of the relationship between two variables. The magnitude of the Pearson correlation coefficient is directly proportional to the strength of the relationship between the two variables. The stronger the relationship between two variables, the higher the confidence that we can determine a variable's value from

the other variable's value. We have utilized Pearson Correlation to determine the strength of the relationship between variables depicting the COVID-19 situation and the stock prices of Indian pharmaceutical companies to access the impact of COVID-19 on pharmaceutical stock prices.

D. Polynomial Interpolation

Polynomial Interpolation is an approach used for filling missing values in a series by considering the available values. Polynomial Interpolation utilizes a mathematical function in the form of a polynomial to calculate the missing values in the series. The degree of the Polynomial Interpolation is the power to which the term with the degree is raised. The polynomial is formulated to best fit the available data, i.e., to best overlap the available data points. We have utilized Polynomial Interpolation to fill in the missing values in the fetched COVID-19 data.

E. Forward Feature Selection

Forward Feature Selection is a technique utilized to obtain the most appropriate features to consider while building a machine learning model to predict a target variable or classify the target variable into different classes. Therefore, forward feature selection essentially returns a list of features with the highest impact on predicting or classifying the target variable. We have utilized forward feature selection to select a list of COVID-19 features from a pool of COVID-19 features that have a higher impact on pharmaceutical stock prices.

IV. DATA AND EMPIRICAL TESTS

A. Covid Data

Firstly, we filled in missing values present in the covid 19 data. Second-degree Polynomial Interpolation was used to fill in the missing data of new tests. While Linear Interpolation was utilized to fill in the missing values in the attributes of people vaccinated, new vaccinations, and people fully vaccinated.

Then, we performed smoothing of attributes like new cases, new deaths, new tests, positivity rate, and new vaccinations. Smoothing was performed by calculating the seven-day moving average of the variables. Smoothing was performed to mitigate the unwanted noise in the data resulting from an unintended representation of data. An example of a misleading data representation may be the less number of vaccines administered during the weekend. Such fluctuation in vaccine administration in the raw data may give an impression that there is a scarcity of vaccines during some parts of the week and hence may expect the market to behave accordingly. But in reality, fewer vaccines are administered during the weekend simply because the nurses and officials tasked with immunization are on holiday.

Fig. 2 illustrates the seven-day moving averages of the above-mentioned variables overlapped on the raw data of the variables.



Assessing the Influence of the Covid 19 Pandemic on Indian Pharmaceutical Companies

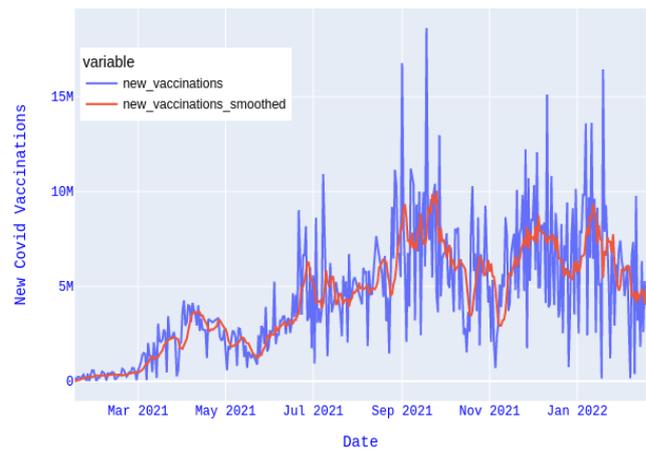
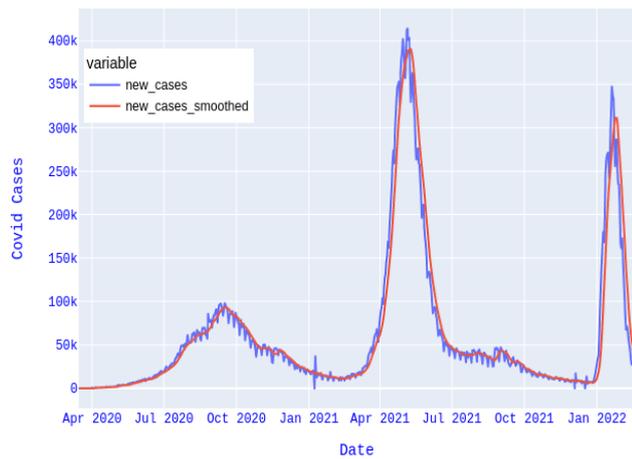
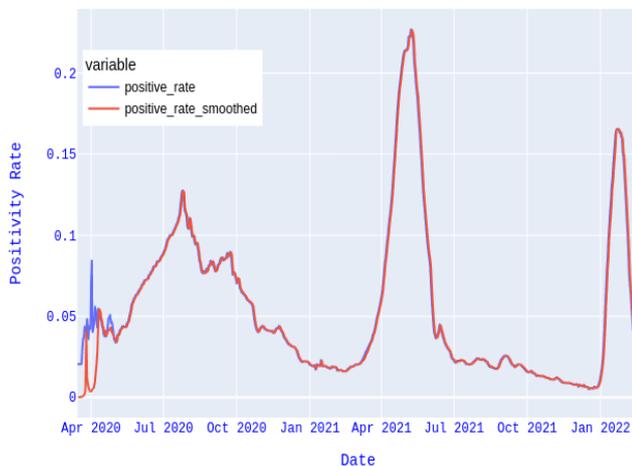
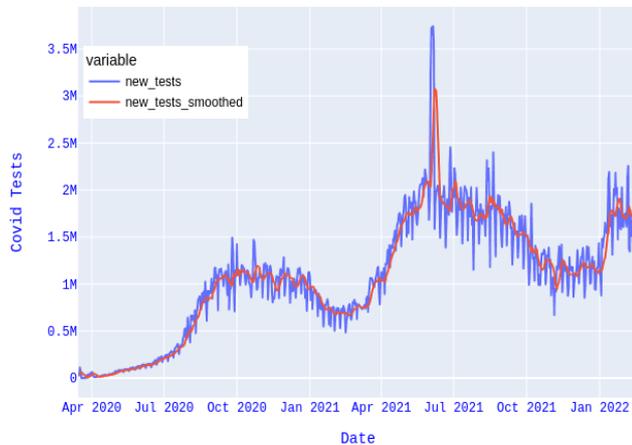
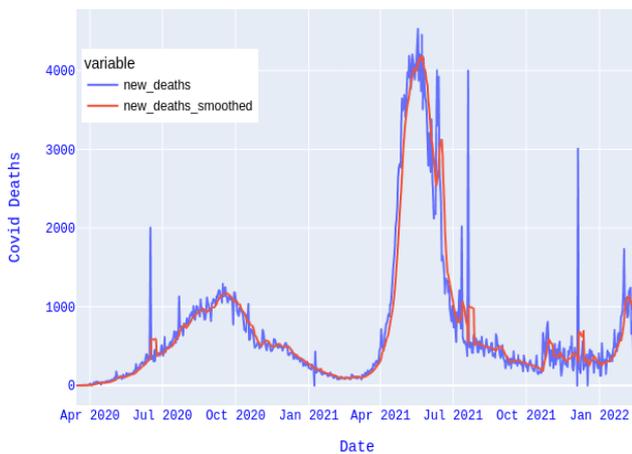


Fig. 2. Seven-day moving average of new cases, new deaths, new tests, positivity rate, and new vaccinations overlapped on their respective raw variables.



Next, we calculated the daily change in new cases, new deaths, new vaccinations, and their respective seven-day moving averages. Finally, we also considered the daily changes in reproduction rate and positivity rate. The daily changes in the above-mentioned fields indicate the change in the severity of the covid 19 pandemic. Hence, they are viable indicators to consider for evaluating the effect of the pandemic.

The stock market is closed during weekends, national holidays, and some festivals in India. Thus, the change in the opening price for stocks on a day when the market reopens after it was closed will not be impacted by the daily change in the new covid cases, new deaths, new vaccinations, reproduction rate, and positivity rate, but, rather be affected by the cumulative change in the above fields from the last day the market was open. For example, if the market was open on Friday, then it was closed for the weekend and reopened on Monday. To account for the impact of the change in the severity of the covid 19 situation on the opening stock price of Monday, we must consider the change in new covid cases, new deaths, new vaccinations, reproduction rate, and positivity rate over the weekend rather than the daily change of the above-mentioned fields. We have hence accounted for the cumulative change of the above fields over the period when the stock market was closed.

The raw covid 19 statistics change may not be proportional to the public sentiment towards the pandemic. For example, an increase in new covid 19 cases per day from 10,000 a day to 50,000 a day will trigger a greater panic than an increase of new covid cases from 310,000 a day to 350,000 a day even though the rise in new covid cases per day is 40,000 in both the scenarios. Therefore, we calculated the common logarithm, i.e., log to the base 10 of all covid 19 fields except reproduction rate, positivity rate, stringency index, google trends popularity index of “covid medicine” and “covid 19”, and the derivatives like moving average or daily change of the above-mentioned fields.



Fig. 3 shows the log of the new cases, new deaths, and new tests.

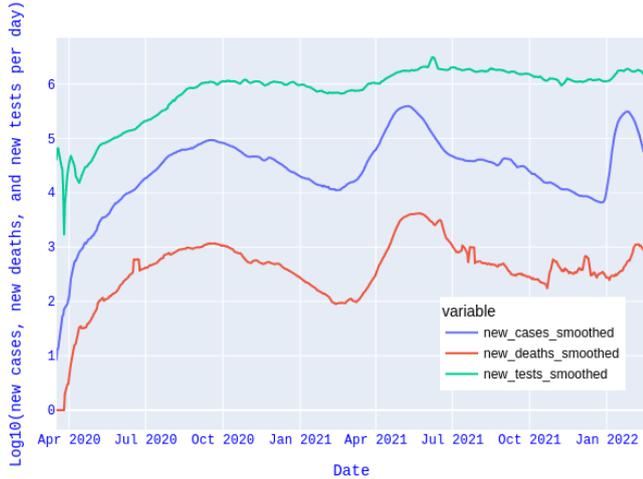


Fig. 3. Logarithm to the base 10 of new Covid cases, new Covid caused deaths, and new tests performed.

B. Pharmaceutical Stock Data

We added variables like High-Low, Oprice-Pclose, and close-open. High-Low stored the difference between the highest traded price during a day and the lowest traded price during a day for a given stock. High-Low indicates the fluctuation in the stock price during a day. The intuition behind adding High-Low is that a sudden change in covid 19 situation during a day may lead to higher fluctuation in the market. Hence considering High-Low is necessary to test this hypothesis. Oprice-Pclose is the difference between the opening price of a stock on a given day and that stock's closing price on the previous day. The covid data of a given day is announced in the early morning of the next day.

Table- I: Correlation table between the seven most impactful covid variables and financial variables of pharmaceutical stock companies.

Parameters	Avg price	No of trades	High-low	Oprice-Pclose	Close-open	Binary Oprice-Pclose	Binary close-open	Total traded quantity	Turnover
Total_cases	0.033	-0.0071	-0.0036	-0.0045	0.00047	-0.0072	0.0012	-0.0043	-0.0023
New_cases	0.0085	0.0043	0.0029	-0.00042	0.001	0.00021	0.00094	0.007	0.0075
New_deaths	0.013	0.003	0.0011	7.3e-05	0.0017	0.0036	0.00097	0.0055	0.0062
New_tests	0.039	-0.002	-0.0041	-0.0027	0.0019	-0.0028	0.0016	0.0015	0.0038
Reproduction_rate	-0.022	-0.00014	0.0071	-0.00029	0.0014	0.0015	0.0017	-0.0017	-0.0027
Positivity_rate	-0.012	0.0045	0.002	0.0013	0.0013	0.0017	-4.9e-05	0.006	0.0053
Total_vaccination	0.014	-0.0083	-0.0033	-0.0043	-0.00073	-0.01	-0.00057	-0.0071	-0.0061

V. RESULT AND DISCUSSION

From Table 1, we can see that out of all the financial features, the average price is most highly correlated with the covid features like total cases, new tests, and total vaccination. This can be attributed to the general trend of increment in the stock prices of pharmaceutical companies from March 2020 to February 2022. Total covid cases and total vaccinations are the cumulative summations of new covid cases and new vaccinations, respectively. Hence they will always be increasing. Thus we see a high correlation between average price with total covid cases and total vaccinations.

Therefore the difference between the opening price of a stock on the following day will be dependent on the covid data of the previous day. Therefore, to test this hypothesis, we have added the field Oprice-Pclose. While close-open is simply the difference between the closing price and the opening price of a stock.

C. Empirical Tests

We performed forward feature selection to obtain COVID-19 features that would be most appropriate to consider while building a machine learning model to predict pharmaceutical stock prices using COVID-19 variables. Doing so would return us the COVID-19 features that would have the highest impact on the pharmaceutical stock prices. Using forward feature selection, we calculated that the top 7 features that most strongly impact the pharmaceutical stock prices are total Covid cases, new Covid cases, new Covid deaths, new tests performed, virus reproduction rate, positivity rate, and total vaccinations.

D. Pearson Correlation

We calculated the Pearson correlation between the seven most impactful covid variables, which were selected using forward feature selection with financial variables like 'Average Price', 'No. of Trades', 'High-Low', 'Oprice-Pclose', 'close-open', 'binary-Oprice-Pclose', 'binary-close-open', 'Total Traded Quantity, and 'Turnover'. Table 1 illustrates the correlation in the form of a correlation table.

We observe that the correlation values range from -0.02 to +0.04. The magnitude of these values is shallow. The less magnitude of the correlation values indicates that there isn't a strong correlation between covid 19 statistics and the pharmaceutical stock prices when the data is considered over a span of almost two years, from March 2020 to February 2022.



VI. CONCLUSION

Our study concludes that the Covid features having the highest correlation with the pharmaceutical companies' stock prices of India are total Covid cases, new Covid cases, new Covid deaths, new Covid tests performed, virus reproduction rate, positivity rate, and total vaccinations. We presented the correlation table depicting the Pearson correlation between the above-mentioned Covid fields and the financial parameters of the Nifty pharmaceutical stock companies.

REFERENCES

1. T. Kimoto, K. Asakawa, M. Yoda and M. Takeoka, "Stock market prediction system with modular neural networks," 1990 IJCNN International Joint Conference on Neural Networks, 1990, pp. 1-6 vol.1. [\[Crossref\]](#)
2. Qian, B., Rasheed, K. Stock market prediction with multiple classifiers. *Appl Intell* 26, 25–33 (2007). [\[Crossref\]](#)
3. Z. Li and W. Jiang, "Evaluation of Machine Learning Techniques for Stock Market Movement Prediction," 2020 Management Science Informatization and Economic Innovation Development Conference (MSIED), 2020, pp. 248-253.
4. K. A. Althelaya, E. M. El-Alfy and S. Mohammed, "Evaluation of bidirectional LSTM for short-and long-term stock market prediction," 2018 9th International Conference on Information and Communication Systems (ICICS), 2018, pp. 151-156. [\[Crossref\]](#)
5. Almeahmadi, Abdulaziz. "COVID-19 pandemic data predict the stock market." *Computer Systems Science and Engineering* (2021): 451-460. [\[Crossref\]](#)
6. im, Meejung. "Prediction of COVID-19 confirmed cases after vaccination: based on statistical and deep learning models." *SciMedicine Journal* 3.2 (2021): 153-165. [\[Crossref\]](#)
7. Nabipour, Mojtaba, et al. "Deep learning for stock market prediction." *Entropy* 22.8 (2020): 840. [\[Crossref\]](#) [\[PMid\]](#) [\[PMCID\]](#)
8. GOH, Thomas Sumarsan, Henry HENRY, and Albert ALBERT. "Determinants and prediction of the stock market during COVID-19: Evidence from Indonesia." *The Journal of Asian Finance, Economics, and Business* 8.1 (2021): 1-6.
9. Khattak, Mudeer Ahmed, Mohsin Ali, and Syed Aun R. Rizvi. "Predicting the European stock market during COVID-19: A machine learning approach." *MethodsX* 8 (2021): 101198. [\[Crossref\]](#) [\[PMid\]](#) [\[PMCID\]](#)
10. Jamous, Razan, Hosam ALRahhal, and Mohamed El-Dariby. "A New ANN-Particle Swarm Optimization with Center of Gravity (ANN-PSOCoG) Prediction Model for the Stock Market under the Effect of COVID-19." *Scientific Programming* 2021 (2021). [\[Crossref\]](#)
11. Ghosh, Indranil, and Tamal Datta Chaudhuri. "FEB-stacking and FEB-DNN models for stock trend prediction: A performance analysis for pre and post Covid-19 periods." *Decision Making: Applications in Management and Engineering* 4.1 (2021): 51-84. [\[Crossref\]](#)
12. Budiharto, W. Data science approach to stock prices forecasting in Indonesia during Covid-19 using Long Short-Term Memory (LSTM). *J Big Data* 8, 47 (2021). [\[Crossref\]](#) [\[PMid\]](#) [\[PMCID\]](#)
13. Singh, P.K., Chouhan, A., Bhatt, R.K. et al. Implementation of the SutteARIMA method to predict short-term cases of stock market and COVID-19 pandemic in USA. *Qual Quant* (2021). [\[Crossref\]](#) [\[PMid\]](#) [\[PMCID\]](#)

AUTHORS PROFILE



Atishay Jain is a 4th-year undergraduate student at Vellore Institute of Technology pursuing his undergraduate degree in Computer Science with a Specialization in Data Science. His current research interests include Machine Learning, Deep Learning, Computer Vision, and Cryptography. His other fields of interest include Data Science and Statistics.



Naman Rastogi is a 4th-year undergraduate student at Vellore Institute of Technology pursuing his undergraduate degree in Computer Science with a Specialization in Data Science. Naman is particularly interested in Data Science and has done many projects. Naman wants to explore Data Science in the field of health, particularly Diabetes.



Dr. Sushma Jain works as an Assistant Professor in the Computer Science and Engineering Department, Thapar Institute of Engineering & Technology, Patiala. She has done her Ph.D. from the Thapar Institute of Engineering & Technology, Patiala. She has more than 21 years of teaching experience. She has supervised 15 M.Tech. dissertations and contributed 15 articles in

Conferences and 14 papers in Research Journals. Her areas of interest are Networking and Data Analytics using Machine Learning

