

Statistical modeling of pharmaceutical stocks during covid-19 Era.

Harsh.T.Parikh¹ and Karan.V.Padariya²

^{1,2}Nirma University, Ahmedabad Gujarat 380014, India

Abstract. Do dramatic price movements in pharmaceutical stocks represent unanticipated scientific knowledge generated throughout the pharmaceutical R&D process, such as pre- and clinical trial outcomes, recalls, and withdrawals, as well as the approval of new drugs?

Is it true that stock prices first overreact to such news? During the Covid-19 shutdown, individuals began to learn about passive income, and stock investment is one of them. We used several machine algorithms such as knn, Linear Regression, and Prophet to analyses data from 7 pharmaceutical firms listed on the National Stock Exchange (NSE) from 2015 to 2020.

Keywords: Stock market, Price Prediction, Machine Learning, Statistical Modeling , K-nearest neighbor , Linear Regression , Fbprophet

1 Introduction

The emergence of the Covid-19 epidemic had a massive and unprecedented impact on both the global and Indian economies.

The stock market, which is used as a gauge of a country's economic activity, is harmed. This epidemic had a significant impact on not just public health but also the financial markets. Since the Covid-19 outbreak, the market value of Standard & Poor (S&P) 500 indices has dropped by 30%. During this time, the Sensex has also dropped by about 32%.The epidemic has caused uncertainty and danger all across the world, hurting both mature and rising economies such as the Spain, Italy, Brazil, and India

The pharmaceutical business is an active researcher. Managers and investors in pharmaceutical and biotech companies are likely to be interested in this research.

The R&D process is especially delicate. There are two key stages in every R&D process: discovery and innovation. Because the development of new medications is a lengthy (and expensive) process, these tend to be separated in time in the pharmaceutical business. The favorable relationship between innovation and firm profitability has been well demonstrated in the economic literature. Especially when pharmaceutical firms are coming up with different solutions to global problem of covid-19 pandemic We used different machine learning models to forecast stock prices in this work, including KNN, Linear Regression, and Fbprophet.

2 Literature Overview

[1]Aitkin et al. offer statistical modelling of a large amount of educational research data on teaching styles and pupil performance. A latent class model is used to group instructors into various teaching methods, and unbalanced variance component ("mixed") models are used to compare these latent classes for variations in child accomplishment.[2] Empirical Bayes models based on Dirichlet processes are presented by Lavine et al and generalized to Polytrees. A Polya tree model is used to reanalyze a Berry and Christensen case. Rydberg et al. [3] addressed on issues relating to the interaction of statistics and financial econometrics.[4]Poshakwale et al. state that this work rejects the random walk model of efficient price creation for the Indian market, and that more research is needed to corroborate the empirical data given in this paper. Statistical modelling is a vast field, according to Jones et al. Because of its significant potential in the study of social research problems, we have included a comprehensive discussion on the emerging method of multilevel modelling (McCullagh and Nelder, 1989) in the advanced viewpoint of the generalised linear model (McCullagh and Nelder, 1989).

[6]Pérez-Rodriguez et al. used an Autoregressive Moving Average–Generalized Autoregressive Conditional Heteroscedasticity (ARMA–GARCH) dynamic econometric model to model market-adjusted daily changes in stock prices of the 17 largest pharmaceutical companies from 1989 to 2008 to detect large price changes.[7]The intellectual capital research by Amin, Shahid, and colleagues is useful for investors in making suitable investment selections.[8]Through a case study covering four equities from 2013 to 2018, McKnight et al predicted stock returns. The financial models are contrasted by analysing these approaches.[9]Ali N et al. have demonstrated the applicability of EVT in the realm of finance using gold price data from 2002 to 2019. Dr. Avijit Sikdar et al. [10] used a paired sample t-test to compare closing average prices, average daily return, average volatility, daily number of trading, and average delivery from 50 listed businesses in five sectors.

3 Dataset Description

Yahoo Finance provided the dataset in CSV format. Our dataset for BIOCON, CIPLA, CADILA, ZYDUS, LUPIN, included day-by-day data from 01-01-2015 to 01-01-2020. Each dataset has seven variables: 'Date,' 'Open,' 'High,' 'Low,' 'Close,' 'Adj,' 'Close,' and 'Volume.' The open and close columns reflect the initial and closing prices of currencies on specific days. The dataset's 'High,' 'Low,' and 'Latest' columns describe the currency's maximum, minimum, and last price. Furthermore, the market is closed on weekends and bank holidays, so no data is available to them.. All the prices were in INR and as we aimed to predict the closing price at a specific date, we have only used the closing price and date from our dataset. The graphs for all the pharmaceutical stocks is shown in figure 1 and figure 2.

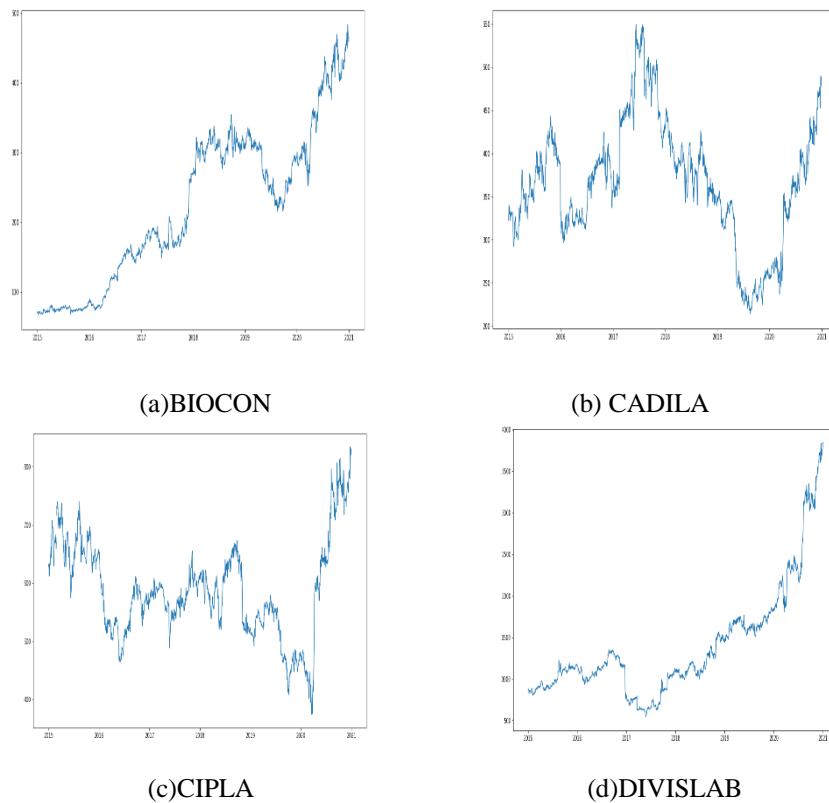
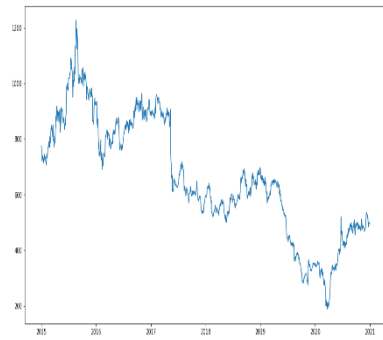
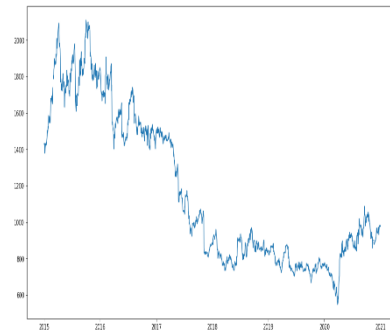


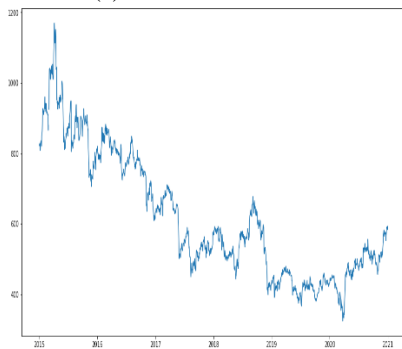
Figure 1. Closing price of pharma stock from 01-01-2015 to 01-01-2020



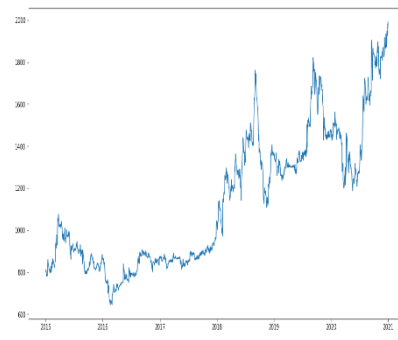
(e)GLENMARK



(f)LUPIN



(g) SUNPHARMA



(h)ZYDUSWELL

Figure 2. Closing price of pharma stock from 01-01-2015 to 01-01-2020

4 Proposed Method

4.1 K-Nearest Neighbour (KNN)

KNN works by first plotting all of the data points and then calculating the value for a new unknown data point as the average of some nearest neighbouring data points. A constant K determines the number of nearest neighbouring data points to look for. KNN assumes that objects of a similar class are clustered together and employs the distance formula to locate the nearest neighbouring element. We used the same features for KNN as we did for Linear Regression.

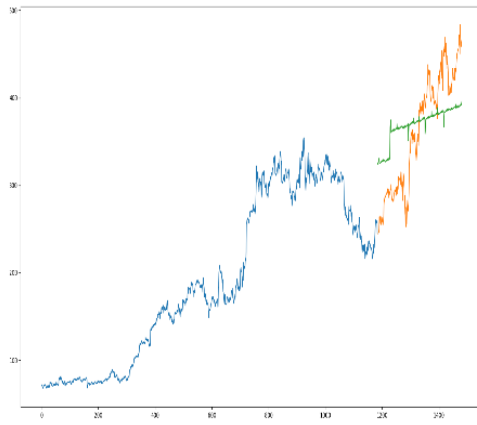
4.2 Fbprophet

Fbprophet is a time-series prediction tool developed by Facebook. It is simple to use, with only two inputs: the date and the target variable. We have directly given the dates column and the closing price column as input to prophet, and it does all of the other work on its own. It does not necessitate any data preprocessing. Fbprophet also considers seasonal trends to make more accurate predictions.

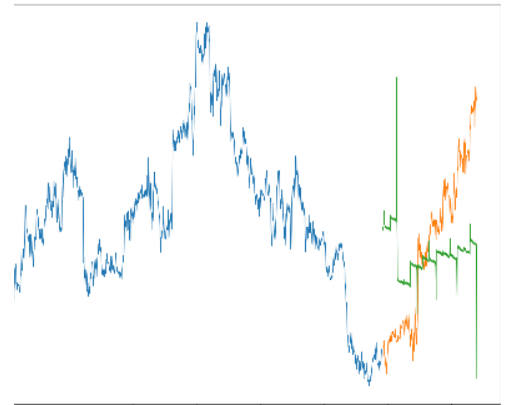
4.3 Linear Regression

Linear regression works by constructing an equation that describes the relationship between the dependent and independent variables. It looks for the best equation to fit the training dataset and then uses it to predict unknown values. The date is an independent variable in our problem statement. We extract features such as day, month, and year from the date and then fit a linear regression model to it. Furthermore, we have added other features based on the idea that the first and last day of the week have a greater impact on the closing price of the stock than any other day. So, if the day of the week is 0 or 4, this feature assigns a value of 1 to the column; otherwise, it assigns a value of 0.

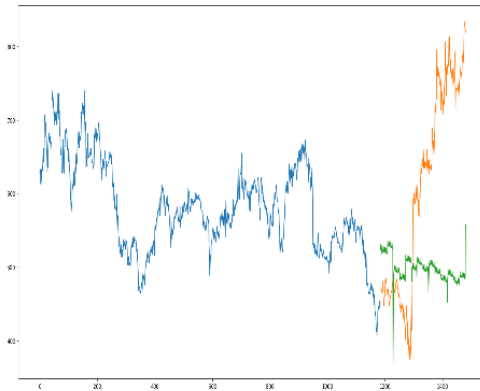
5 Results



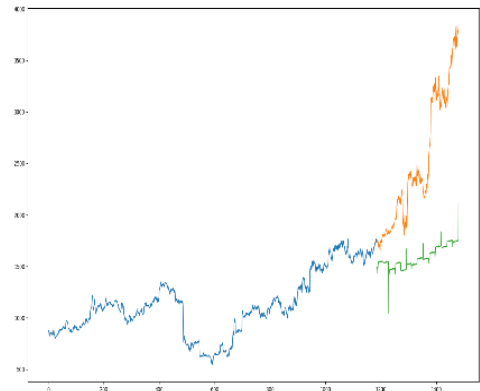
(b) CADILA



(a) BIOCON

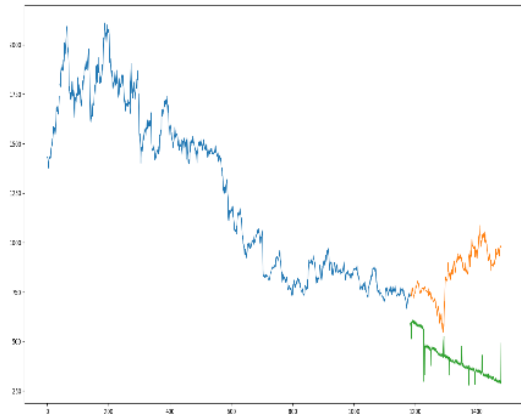


(c) CIPLA



(d) DIVISLAB

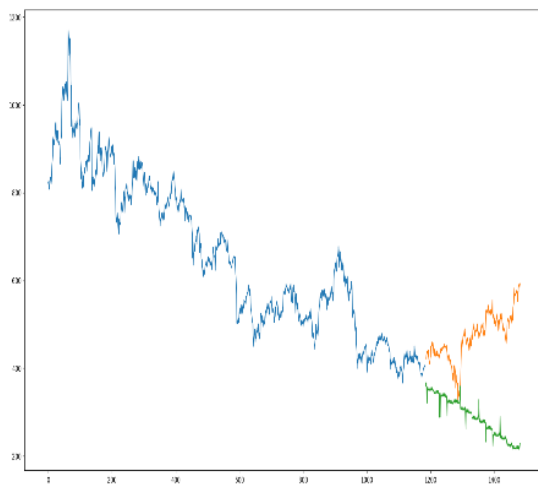
Fig. 3. Prediction of pharma stock price using linear regression



(e) GLENMARK



(f) LUPIN

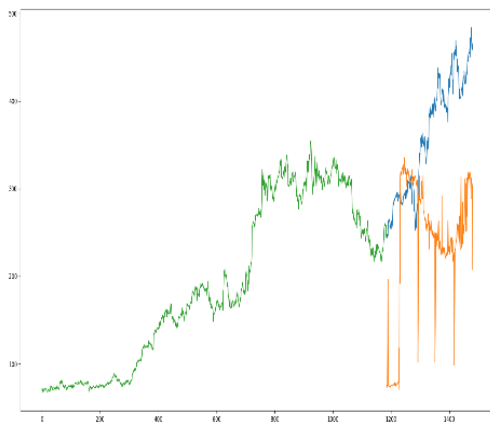


(g) SUNPHARMA

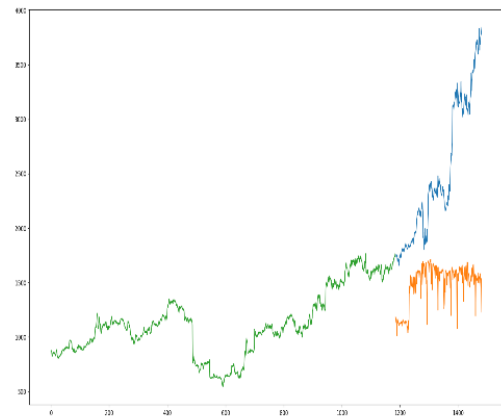


(h) ZYDUSWELL

Fig. 4. Prediction of pharma stock price using linear regression



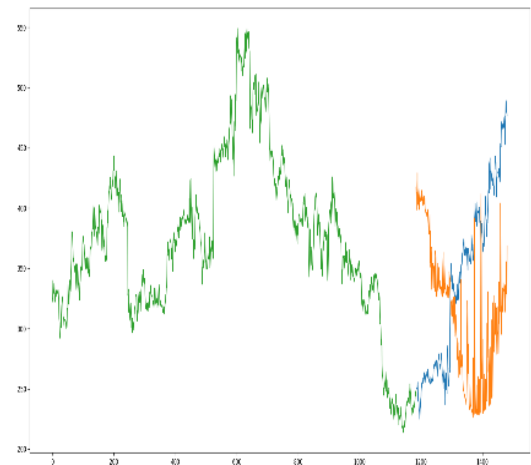
(a) BIOCON



(b) CADILA

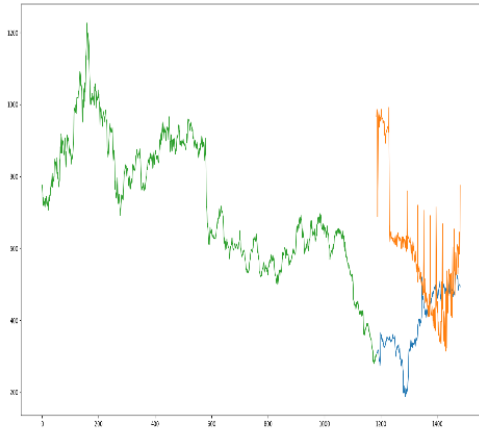


(c) DIVISPAB

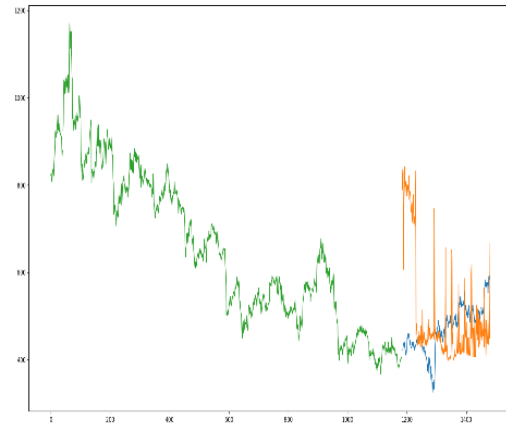


(d) CIPLA

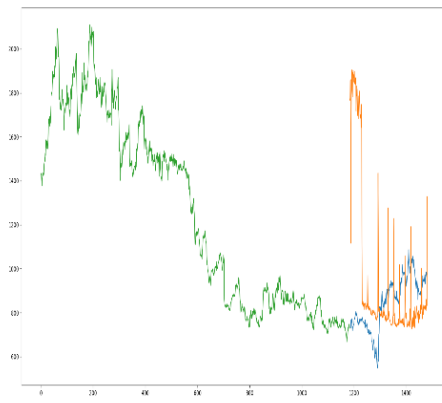
Fig. 5. Prediction of pharma stock price using KNN



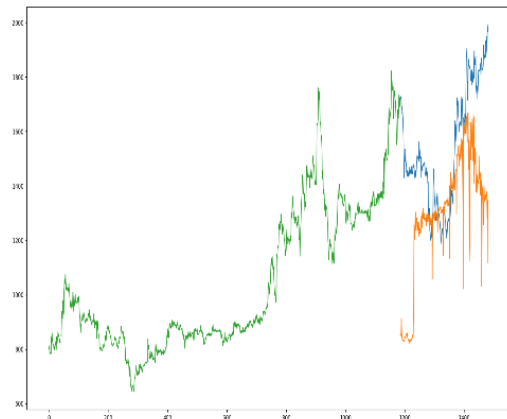
(e)GLENMARK



(f)LUPIN

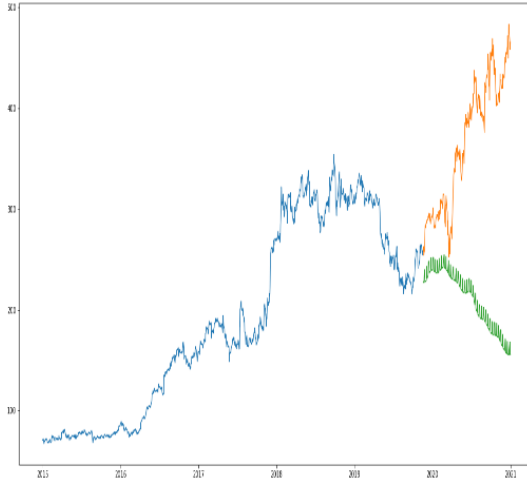


(g) SUNPHARMA

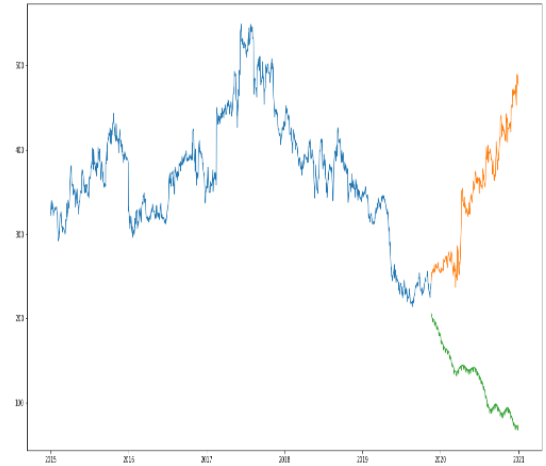


(h)ZYDUSWELL

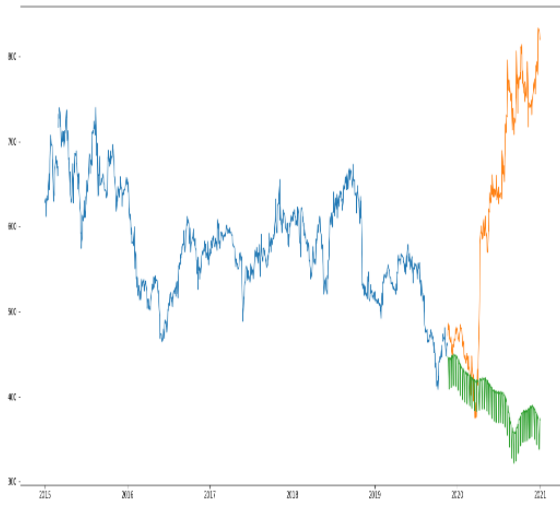
Fig. 6. Prediction of pharma stock price using KNN



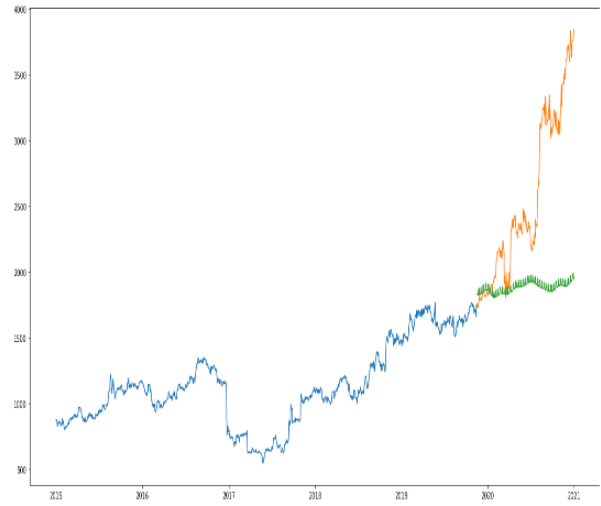
(a) BIOCON



(b) CADILA

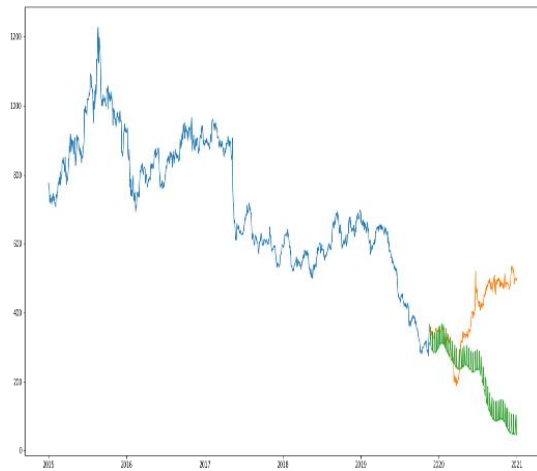


(d) DIVISPAB

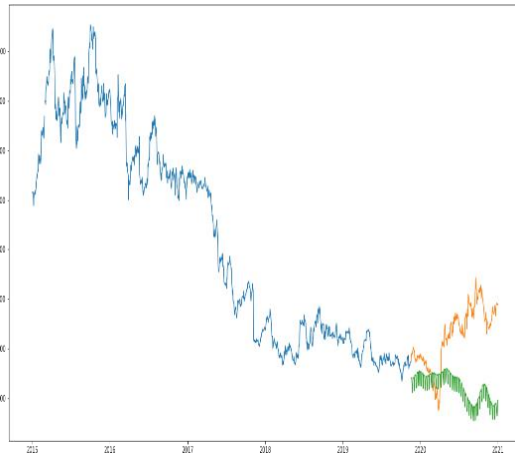


(e) CIPLA

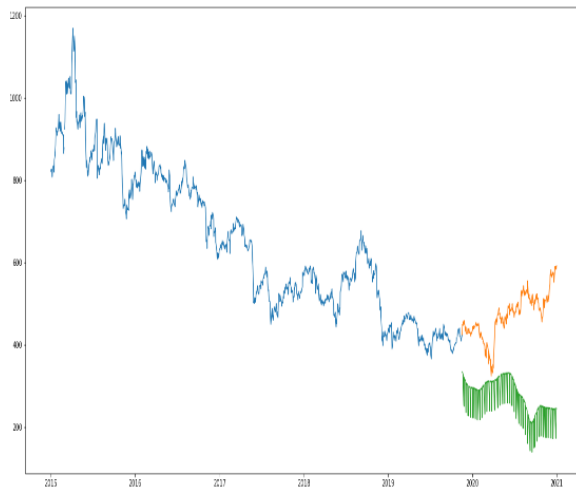
Fig. 7. Prediction of pharma stock price using prophet



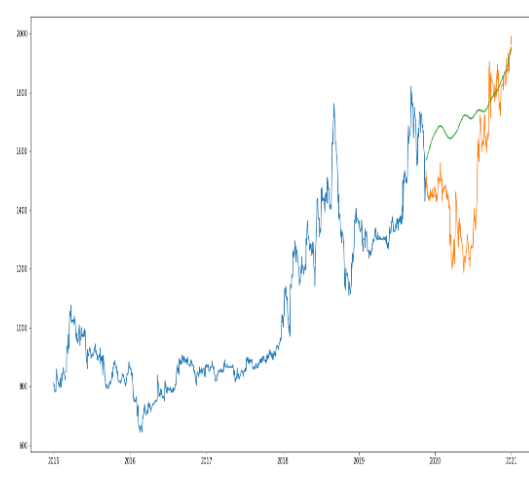
(d) DIVISPAB



(e) CIPLA



(g) SUNPHARMA



(h) ZYDUSWELL

Fig. 8. Prediction of pharma stock price using prophet

6 Conclusion

To foresee Pharmaceutical stock prices, a variety of machine learning and statistical models were used. Cryptocurrency and Stock prices can be forecasted using this study's findings. KNN has provided us with better results, while Linear Regression and prophet have shown large deviation from the true prices. This study can also be applied to sector wise stock price prediction and other time series forecasting problems.

References

- [1]Aitkin, Murray, Dorothy Anderson, and John Hinde. "Statistical modelling of data on teaching styles." *Journal of the Royal Statistical Society: Series A (General)* 144.4 (1981): 419-448.
- [2]Lavine, Michael. "More aspects of Polya tree distributions for statistical modelling." *The Annals of Statistics* 22.3 (1994): 1161-1176.
- [3]Rydberg, Tina Hviid. "Realistic statistical modelling of financial data." *International Statistical Review* 68.3 (2000): 233-258.
- [4]Poshakwale, Sunil. "The random walk hypothesis in the emerging Indian stock market." *Journal of Business Finance & Accounting* 29.9-10 (2002): 1275-1299.
- [5]Jones, Kelvin. "An introduction to statistical modelling." *Research methods in the social sciences*. Sage, London (2004): 236-251.
- [6]Pérez-Rodríguez, Jorge V., and Beatriz GL Valcarcel. "Do product innovation and news about the R&D process produce large price changes and overreaction? The case of pharmaceutical stock prices." *Applied Economics* 44.17 (2012): 2217-2229.
- [7]Amin, Shahid, and Shoaib Aslam. "Intellectual capital, innovation and firm performance of pharmaceuticals: A study of the London Stock Exchange." *Journal of Information & Knowledge Management* 16.02 (2017): 1750017.
- [8]McKnight, Sarah. *Statistical Analysis for Financial Data: A Case Study of Four Stocks*. Diss. Montana State University, 2019.

[9]Ali, N., N. Zaimi, and N. MOHAMED Ali. "Statistical modelling of Malaysia trading gold price using extreme value theory approach." *Advances in Mathematics: Scientific Journal* 10.1 (2020): 1857-8365.

[10]Sikdar, Avijit. "Sector wise stock market performance during pre and post covid era." *International Journal of Engineering and Management Research* (2021).