Price Prediction for Pharmaceutical Stocks During Covid-19 Pandemic

Karan V. Padariya(✉), Harsh T. Parikh, and Ankit K. Sharma

Institute of Technology, Nirma University, Ahmedabad, Gujarat, India
{19bic033,19bic036}@nirmauni.ac.in, ankit.sharma@irmauni.ac.in

Abstract. Dramatic price changes in pharmaceutical equities reflect unexpected scientific information gained throughout the pharmaceutical R&D process, such as clinical trial results, recalls and withdrawals, and the approval of new treatments. During the Covid19 shutdown, major pharmaceutical firms were studying and producing vaccinations, pills, and other medicines to combat the coronavirus outbreak. This activity has a substantial impact on the global and Indian markets. Stock price prediction is an important issue in finance and economics that has piqued the interest of scholars throughout the years in developing better predictive algorithms. To evaluate Sun Pharma Ltd. data, we employed machine learning techniques such as K-nearest neighbor (KNN), Linear Regression, and Fbprophet during the time frame 2016 to 2020. It is the second-largest Indian pharmaceutical firm in terms of stock volume. Statistical modeling algorithm Fbprophet outperforms standard regressing algorithms like K-nearest neighbor and Linear regression on time series data.

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 on 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 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

S. Tamane et al. (Eds.): ICAMIDA 2022, ACSR 105, pp. 61–68, 2023.
https://doi.org/10.2991/978-94-6463-136-4_8

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 the global problem of the covid-19 pandemic. In this work, we used different ma-chine learning models to forecast stock prices like KNN, Linear Regression, and Fbprophet.

To better understand how these three algorithms predict the stock price and can adapt to changes in the stock price trend. We have taken time series data from Sun Pharma Ltd. from 2015 to 2020. Python programming language was used to implement these machine-learning methods in PyCharm software. Section 2 briefly describes the empirical literature on machine learning algorithms and the linear R&D process in the pharmaceutical industry, and different statistical models. In Sect. 3, we describe the data used for this work. In Sect. 4, the algorithms we used to predict the stock price are described. In Sect. 5, we illustrated the results obtained, and the Conclusion was stated in Sect. 6.

2 Literature Review

Saebold et al. [1] discussed the relationship between statistics and python that shows how statistical models package fill a gap in this relationship. Empirical Bayes models based on Dirichlet processes are presented by Lavine et al. [2] and generalized to Polyatrees. The Polya tree model is used to reanalyze a Berry and Christensen case. Rydberg et al. [3] addressed 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 modeling is a vast field, according to Jones et al. [5] Because of its significant potential in the study of social research problems, we have includ-ed a comprehensive discussion on the emerging method of multilevel modeling (McCullagh and Nelder, 1989) in the advanced view of the generalized linear model (McCullagh and Nelder, 1989).

Pérez Rodrguez et al. [6] 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. The intellectual capital research by Amin, Shahid, and colleagues is useful for investors in making suitable investment selections [7]. Through a case study covering four equities from 2013 to 2018, McKnight et al. [8] al predicted stock re-turns. The financial models are contrasted by analyzing these approaches. Ali N et al. [9] have demonstrated Extreme value theory's applicability in 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, the daily number of trading, and average delivery from 50 listed businesses in five sectors. Mr. Ayodele A. Adebiyi and Aderemi O. Adewum et al. [11] have used the Auto Regressive Integrated Moving Average (ARIMA) algorithm for stock price prediction for short-term investments. But [12, 13] ARIMA is very good for daily price prediction not for the long-term artificial neural network (ANN) is better than ARIMA.ANN gives very good accuracy on daily basis as well using data of high, low,

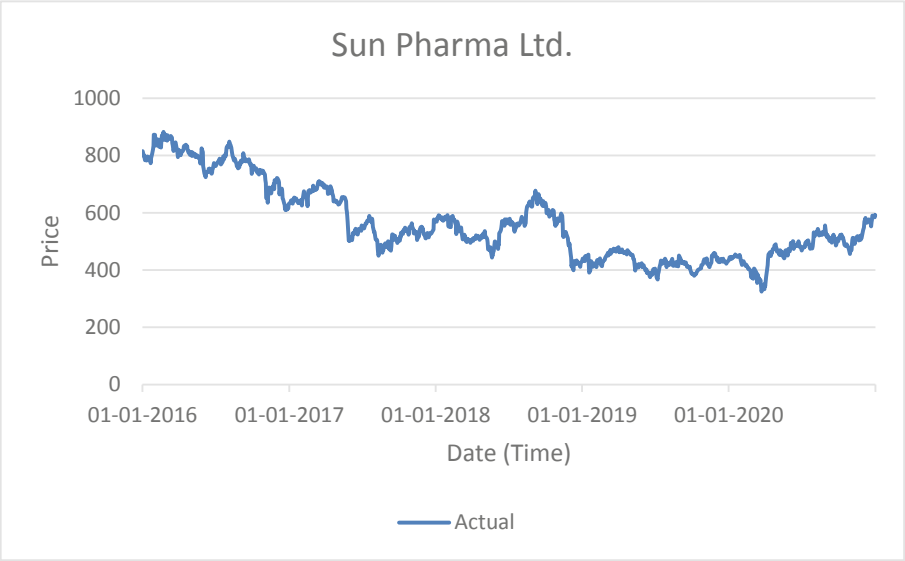

Fig. 1. The closing price of Sun pharma stock 2015–2020 (Year vs Price)

open, and close prices only. the stock price depends on many factors and it fluctuates with news related to it so we can make a deep learning algorithm that understands the news and can product change [14].

3 Dataset Description

Yahoo Finance provided the dataset in CSV format. Our dataset for Sun pharma Ltd. included day-by-day data from 01-01-2016 to 31-12-2022. Each dataset has seven variables: 'Date', 'Open', 'High', 'Low', 'Close', 'Adj', 'Close', and 'Volume'. The open and close columns reflect currencies' initial and closing prices 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 week-ends and bank holidays, so no data is available to them. All the prices were in Indian rupees (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 graph of the closing price is shown in Fig. 1.

4 Proposed Methods

Block diagram of the proposed flow is shown in Fig. 2.

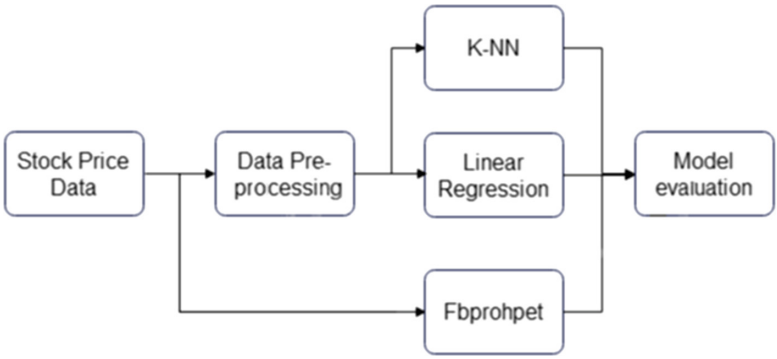

Fig. 2. Block diagram of proposed Method

4.1 K-Nearest Neighbor (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 neighboring data points. A constant K determines the number of nearest neighboring 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 neighboring 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 the Fbprophet, and it does all of the other work on its own. It does not necessitate any data preprocessing. Fbprophet also considers seasonal trends function, growth functions and holiday/event function to make more accurate predictions.

4.3 Linear Regression

Linear regression works by constructing an equation that describes the relation-ship 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 the date's features, such as day, month, and year, and then fit a linear regression model. Furthermore, we have added other features based on the idea that the first and last day of the week has 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.

Table 1. Accuracy of All three implemented algorithms

Machine Learning Models	Accuracy (RMS)
Linear Regression	105.71
K-Nearest neighbor	240.56
Fbprophet	81.23

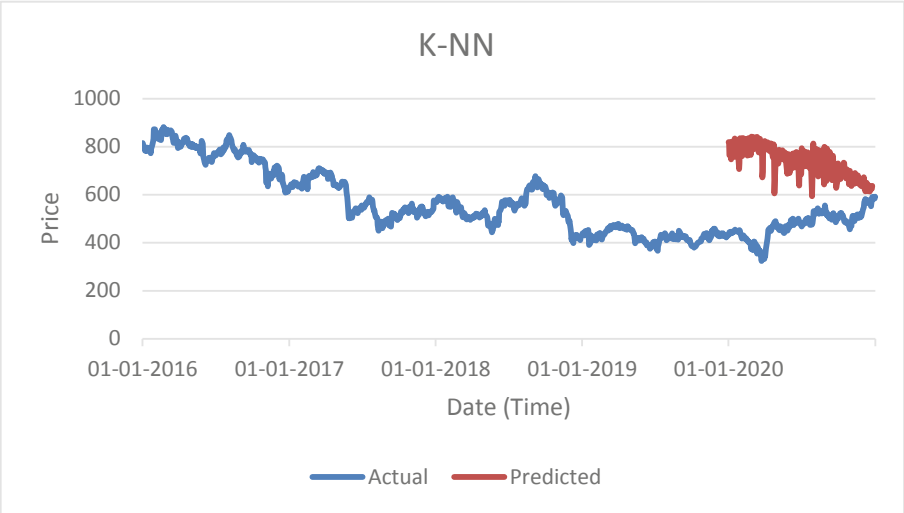

Fig. 3. Prediction using K nearest neighbor. (Year vs price)

5 Results

The Root Mean Square (RMS) Accuracy is a typical method for calculating a model's error in predicting statistical data. RMS is the quadratic mean and is used to average out regular variations. Figure 4 shows that linear regression works effectively until 2020 when going with the downtrend, which is before the pandemic, but after the pandemic trend changed upward, the prediction went incorrect. The prediction fluctuations are too much in the KNN method, as illustrated in Fig. 3, resulting in a poor outcome and a high RMS accuracy. Figure 5 shows that Fbprophet gives comparable patients as the actual price and has the lowest RMS Accuracy. Table 1 shows RMS accuracy value for Linear regression is 105.71, the value comes out to be 240.56 in the case of KNN and it turns out to be 81.23 when Fbprophet is applied to Sun pharma data. Hence Fbprophet out-performs as the average difference between the actual and predicted value is the smallest.

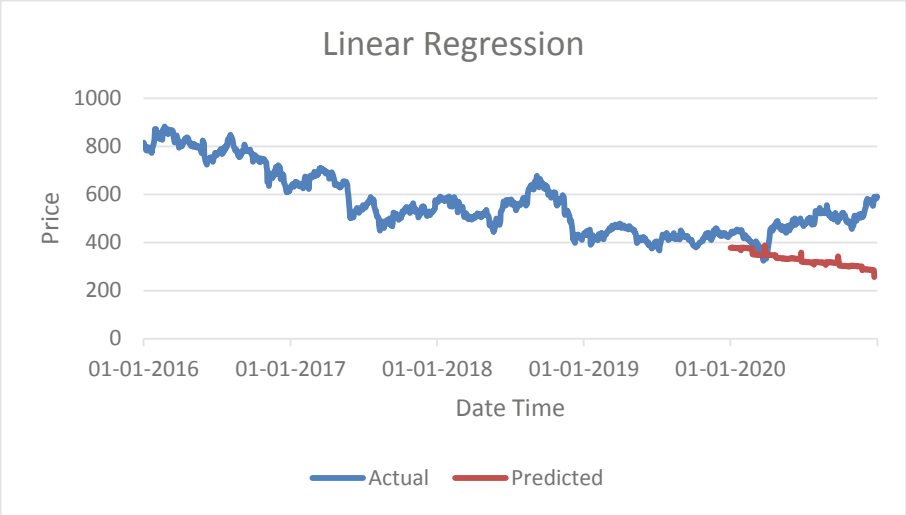

Fig. 4. Prediction using Linear Regression. (Year vs price)

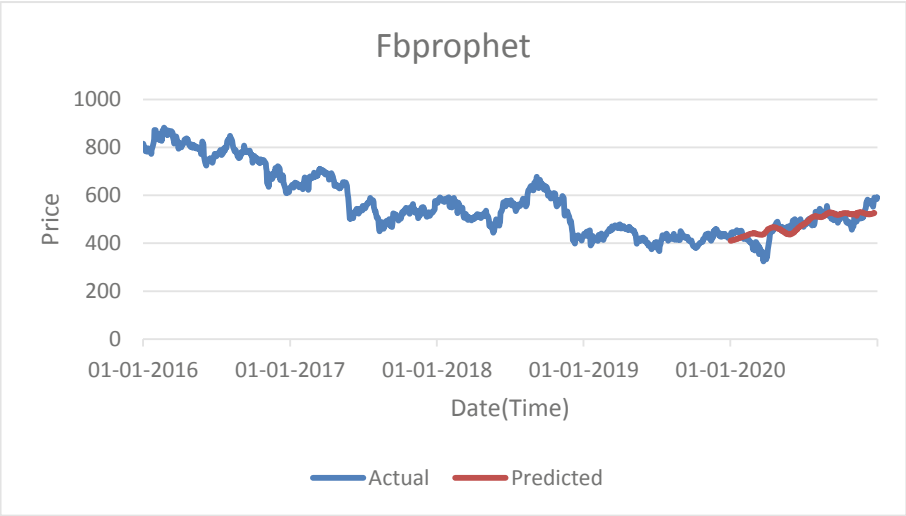

Fig. 5. Prediction using the Fbprophet. (Year vs price)

6 Conclusion

Three machine learning model were used to forecast pharmaceutical stock prices. Cryptocurrency and stock prices can be forecasted using this study's findings. K nearest neighbor and linear regression did not provide us with expected results, showing a large deviation from true values, hence having a larger Root mean squared error (RMSE) value. Meanwhile, the Fbprophet has the least RMSE value among the three algorithms.

Making Fbprophet the most efficient algorithm specific to this time series problem, it can adeptly change the trend of the price as it understands seasonality.

References

1. Seabold, Skipper, and Josef Perktold. "Statsmodels: Econometric and statistical modeling with python." In Proceedings of the 9th Python in Science Conference, vol. 57, no. 61, pp. 10-25080. 2010.
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).
11. Ariyo, Adebiyi A., Adewumi O. Adewumi, and Charles K. Ayo. "Stock price prediction using the ARIMA model." 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation. IEEE, 2014.
12. Nguyen, Nguyet, and Mohammad Islam. "Comparison of Financial Models for Stock Price Prediction." 2021 Joint Mathematics Meetings (JMM). AMS, 2021
13. Vijh, Mehar, et al. "Stock closing price prediction using machine learning techniques." Procedia computer science 167 (2020): 599-606.
14. Mohan, Saloni, et al. "Stock price prediction using news sentiment analysis." 2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService). IEEE, 2019.