Abstract:

after covid 19 ear Indian got boom in stock market because during lockdown people started to learn passive income and investing in stock is one of them so number of tread per day has increase as well as number of investors also increased.in this paper we analysed data of 7 pharmaceutical companies listed in National stock Exchange(NSE) from 2015 to 2020 and using different machine algorithm like knn, Linear Regression, Prophet we analyse how prediction goes wrong or right due to covid-19 pandemic and how it affects to the stocks.

Literature review:

1. Statistical Modelling of Data on Teaching Styles[1981]

This paper presents the detailed statistical modelling of an extensive body of educational research data on teaching styles and pupil performance. Clustering of teachers into distinct teaching styles is carried out using a latent class model, and comparison of these latent classes for differences in pupil achievement is examined using unbalanced variance component ("mixed") models. Differences among the classes are altered by the probabilistic clustering of the latent class model compared to the original findings of the Teaching Styles project, and the statistical significance of the differences is substantially reduced when allowance is made for the correlation among children taught by the same teacher.

2. more aspects of polya tree distribution for statistical modelling[1994]

empirical Bayes models using Dirichlet processes are generalized to Polya trees. An example from Berry and Christensen is reanalyzed with a Polya tree model.

3. Realistic Statistical Modelling of Financial Data[2000]

this paper has dealt with problems in the area of interplay between statistics and financial econometrics. It has described most of the features that characterize financial data.

4. The Random Walk Hypothesis in the Emerging Indian Stock Market[2002]

38 most actively traded stocks in the Bombay Stock Exchange.

this paper rejects the random walk model of efficient price formation for the Indian market, further research is suggested whereby other plausible models should be analysed to confirm the empirical evidence presented in this paper.

5.An Introduction to Statistical Modelling[2004]

Statistical modelling is a huge subject. In the space we have available I will concentrate on why you do modelling and what can be achieved. I consider what sort of questions it can answer, what sort of data looks like a ‘regression’ problem and what steps we can take to ensure we get valid results. I have written this introduction from the advanced perspective of the generalized linear model (McCullagh and Nelder, 1989) and have included a substantial discussion on the developing approach of multilevel modelling because of its major potential in the analysis of social research questions.

6.Do product innovation and news about the R&D process produce large price changes and overreaction? The case of pharmaceutical stock prices[2012]

We modelled market-adjusted daily changes in stock prices of the 17 biggest pharmaceutical firms worldwide for the period from 1989 to 2008 to detect large price changes (outliers), using an Autoregressive Moving Average–Generalized Autoregressive Conditional Heteroscedasticity (ARMA–GARCH) dynamic econometric model.

we conclude that large price changes do not lead to subsequent price reversals, either in the short or the long term, for the pharmaceutical stock

7. Intellectual Capital, Innovation and Firm Performance of Pharmaceuticals: A Study of the London Stock Exchange[2017]

IC which leads to innovation, growth and greater business performance. Greater IC efficiency also contributes to achieve superior business performance. this study is also beneficial for the investors to make appropriate investment decisions.

8. Statistical Analysis for Financial Data: A Case Study of Four Stocks[2019]

In this paper author investigates different models for predicting stock returns through a case study involving four stocks from 2013 to 2018. By examining these methods, the predictive ability of several common, and some less common, financial models are compared.

9. STATISTICAL MODELLING OF MALAYSIA TRADING GOLD PRICE USING EXTREME VALUE THEORY APPROACH[2020]

in this study, the application of EVT in the field of finance has been illustrated to gold price data from the year 2002 until 2019. GEV distribution is used to model the monthly maxima of negative log returns of gold prices. The assessment of the Q-Q plot suggests that GEV distribution provides a good fit for the series of data using the MLE method.

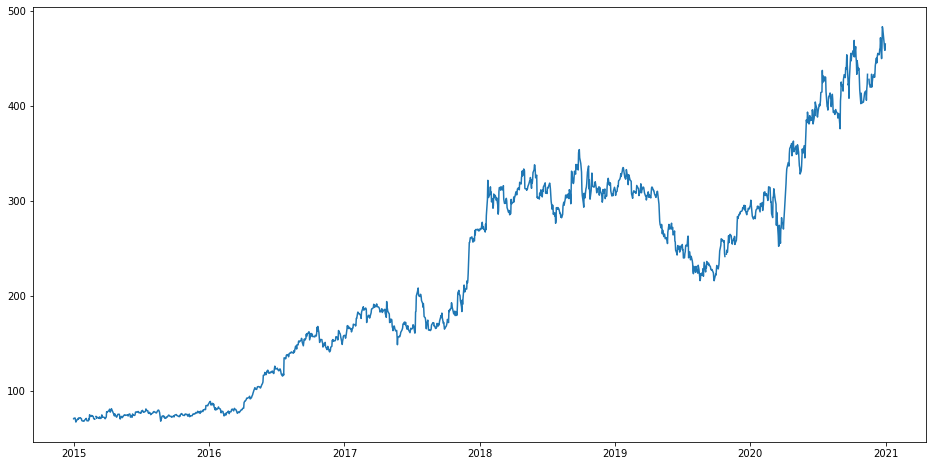
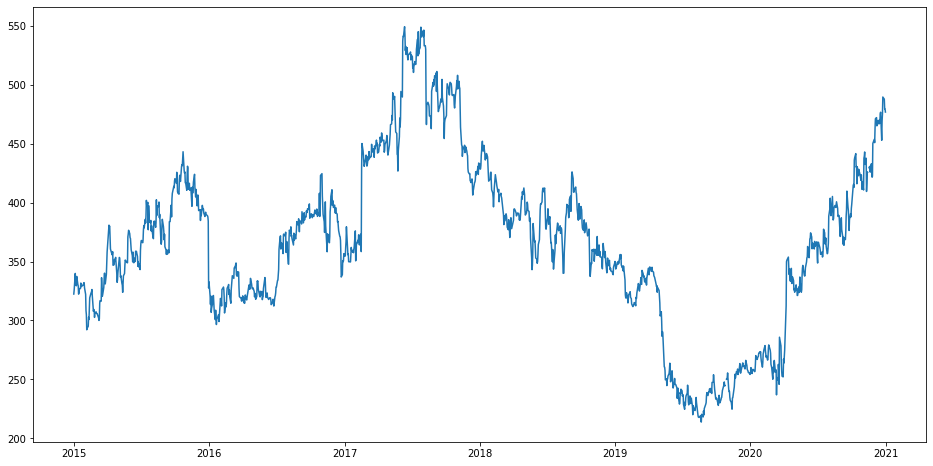
10.Sector Wise Stock Market Performance during Pre and Post Covid Era[2021]

Dr. Avijit Sikdar [1] has analysed stoke price data from 50 listed companies of 5 sectors and compared closing average prices, average daily return, average volatility, daily number of trading and average delivery using paired sample t-test.

Dataset Description

The dataset in CSV format was downloaded from Yahoo finance. For BIOCON, CIPLA, our dataset had day-wise data from 01- 01-2015 to 01-01-2020. There were 7 variables available for each dataset namely: ’Date’, ’Open’, ’High’, ’Low’, ’Close’, ’Adj’, ’Close’, ’Volume’. Open and close columns represent the starting price and closing price of the Lecture Notes in Computer Science: Authors’ Instructions 3 currency on particular days. Maximum, minimum, and last price of the currency is described in the ’High’, ’Low’ and ’Last’ columns of the dataset. Also, the market is closed on weekends and bank holidays so the data is not available for 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.

IMAGE



(a) BIOCON (b) CADILA

4.3 K-Nearest Neighbour (KNN)

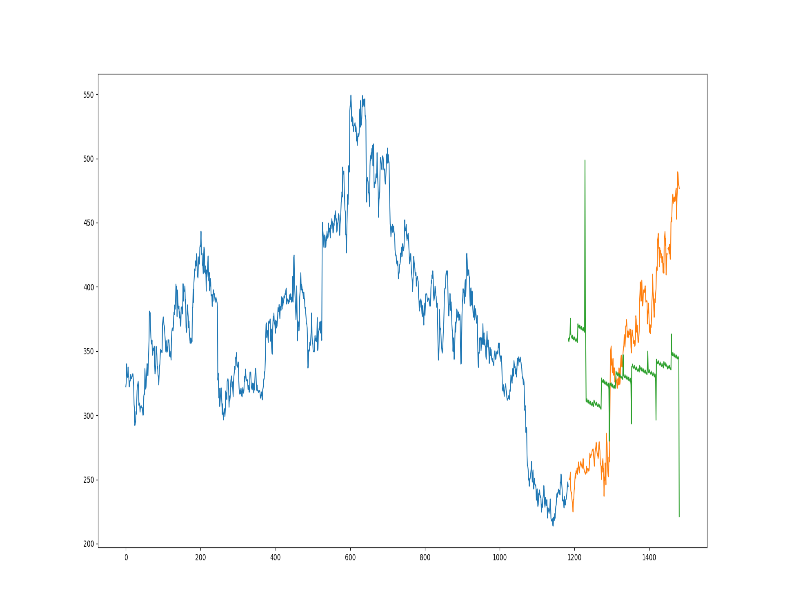
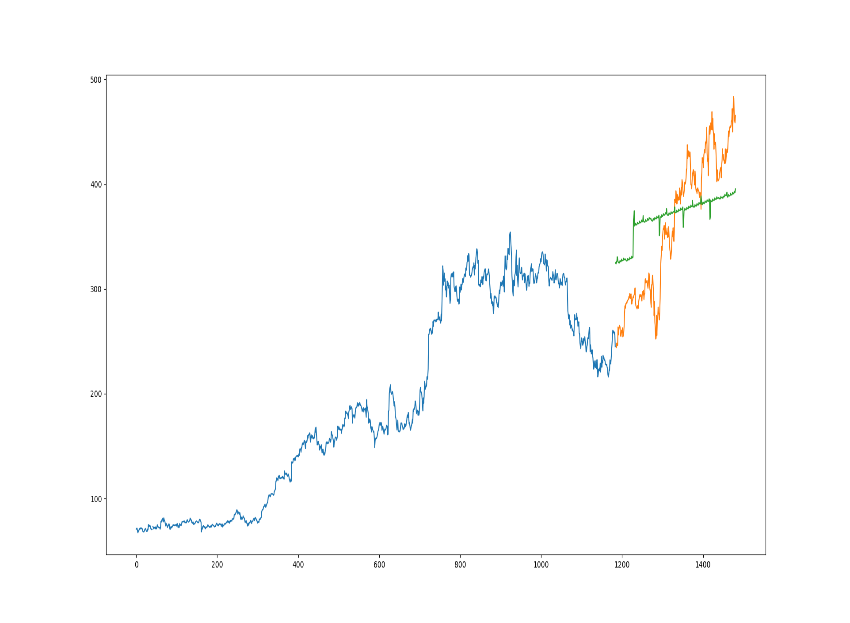
KNN works by plotting all the data points and then calculating the value for a new unknown data point as the average of some nearest neighbouring data Lecture Notes in Computer Science: Authors’ Instructions 5 points. The number of nearest neighbouring data points to look for is determined by a constant K. KNN assumes that a similar class of objects lie together and uses the distance formula to identify the nearest neighbouring element. For KNN also we have used the same features as Linear Regression.

4.5 Fbprophet

Fbprophet was developed by Facebook for time-series prediction, it is very easy to implement as it has only two inputs Date and the target variable. In prophet, we have directly given the dates column and the closing price column as an input and it does all the other work on its own. It does not require data preprocessing. Fbprophet also takes into account seasonal trends for better predictions.

4.2 Linear Regression

Linear regression works by forming an equation that has a relationship between the dependant variable and the independant variable. It searches for the equation that best fits the training dataset and then uses it to predict unknown values. In our problem statement, we have the date as an independent variable. From the date, we extract features like day, month, year and then fit a linear regression model. Moreover, we have also added other features on the idea that the first and the last day of the week would have more impact on the closing price of the stock than any other day. So this feature puts a value as 1in the column, if the day of the week is 0 or 4 otherwise it will keep it as 0. Then the dataset was split into train and test and after training was complete RMSE value along with a graph that shows predicted and actual values was also plotted.

Image

(a)CADILA (b)BIOCON

Reference:

[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).