

Abstract—Share market is a place where people buy/sell shares of publicly listed companies. It offers a platform to facilitate seamless exchange of shares. This buying and selling of shares take place through electronic medium. In this paper different forecasting packages and methods have been used to forecast the stock market price of different companies. We also look at Cognizant, HDFC, HCL, Infosys, SBI, ICICI share market from 2019-2021.

Index Terms—KNN, Prophet model, Neural Network, Share

I. INTRODUCTION

A stock is a financial instrument that represents ownership in a company or corporation and represents a proportionate claim on its assets and earnings. Stocks are also called shares or a company's equity [1]. Ours is a period of unpredictable economy. An organization earning large amount of money today might not be even in the frame of contention in a period of few years. Currently, the stock market has gained utmost attention from all possible sectors because of the high returns it promises. In recent years, trading has become dominated by computers, algorithms are responsible for making rapid split-second trading decisions.

Prophet Forecasting Model: Prophet [2] is a procedure for forecasting time series data based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects. It works best with time series that have strong seasonal effects and several seasons of historical data. Prophet is robust to missing data and shifts in the trend, and typically handles outliers well.

kNN regression time series forecasting: KNN model [3] can be used for both classification and regression problems. The most popular application is to use it for classification problems. For predicting values of new data points, the model uses 'feature similarity', assigning a new point to a values based on how close it resembles the points on the training set.

Feed Forward Neural network: It is a single layer neural network [3]. In this single hidden layer form there is only one layer of input nodes that send weighted inputs to a subsequent layer of receiving nodes. The function model approach is to use lagged values of the time series as input data, reaching to a non-linear autoregressive model.

Share of different companies like HCL, Infosys, ICICI, SBI, Cognizant, HDFC from 2019-2021 have been given. Share is pretty important for a company. The task is to predict future share/ stock prices of these companies.

In this paper, we learn the theory behind different models like prophet model, kNN regression and the mathematics behind these algorithms. Share data of different companies have been given in the CSV format. We compare the return of shares of these companies and project future share of these companies through different models.

II. MODELS

Prophet model, kNN regression model, feed forward neural network are discussed.

A. Prophet forecasting model

Prophet is an open-source tool from Facebook used for forecasting time series data. It is based on a decomposable additive model where non-linear trends are fit with seasonality. Certain terms are required to understand the model. [4]

Trend: It shows the tendency of the data to increase or decrease over a long period of time and it filters out the seasonal variations.

Seasonality: It is the variations that occur over a short period of time and is not prominent enough to be called a "trend".

The general idea of the model is similar to a generalized additive model [4]. The "Prophet Equation" fits trend, seasonality and holidays. This is given by,

$$y(t) = g(t) + s(t) + h(t) + e(t) \quad (1)$$

here, $g(t)$ refers to trend

$s(t)$ refers to seasonality

$h(t)$ refers to effects of holidays to the forecast

$e(t)$ It is the error term.

$y(t)$ is the forecast.

Prophet has two built-in models. One is the logistic growth model and the other one is piece-wise linear model. [4] In logistic growth model it is assumed that the dependent variable has an upper limit called carrying capacity that it approaches asymptotically but cannot outgrow. The logistic growth model is fit using the following statistical equation,

$$g(t) = \frac{C}{1 + e^{-k(t-m)}} \quad (2)$$

where, C is the carry capacity

k is the growth rate

m is an offset parameter

Similar to a linear regression, the time series variable is predicted by the time variable. In Prophet, this is called linear growth. The linear model is fit using the following statistical equations,

$$y = \begin{cases} \beta_0 + \beta_1 x & x \leq c \\ \beta_0 - \beta_2 c + (\beta_1 + \beta_2)x & x > c \end{cases} \quad (3)$$

where, c is the trend change point

By default, Prophet uses linear model. Choosing which model to use is very important as it is dependent on a variety of factors. If the data to be forecasted, has saturating and grows non-linearly and after reaching the saturation point, shows little to no growth or shrink and only shows some seasonal changes, then logistic growth model is the best option. If the data shows linear properties and had a growth or shrink trends in the past then, piece-wise linear model is a better choice. [4]

B. KNN regression time series forecasting

The KNN algorithm uses 'feature similarity' to predict the values of any new data points. This means that the new point is assigned a value based on how closely it resembles the points in the training set. KNN regression process consists of instance, features, and targets components.

The instance is used as a reference vector to find features that are the closest vectors to that instance [5]. The relevant distance metric is calculated by the Euclidean formula as shown below:

$$Distance\ metric = \sqrt{\sum_{x=1}^n (f_x^i - q_x)^2} \quad (4)$$

The q_x denotes the instance and f_x^i indicates the features that are ranked in order by the distance metric [5]. There are other different distance metric as well. The k parameter determines the number of k closest features vectors which are called k nearest neighbors.

Targets are the time-series data that come right after the nearest neighbors and their number is the value of the h parameter. The targets of the nearest neighbors are averaged to forecast the future h period. Features or targets can overlap the instance. This is because the time series data has no seasonality and is in a specific uptrend. The process is called MIMO(multiple-input-multiple-output) [5] strategy that is a forecasting method used as a default with KNN.

III. FEED FORWARD NEURAL NETWORK

A single hidden layer is used while building up the feed forward neural network model. The models looks like:

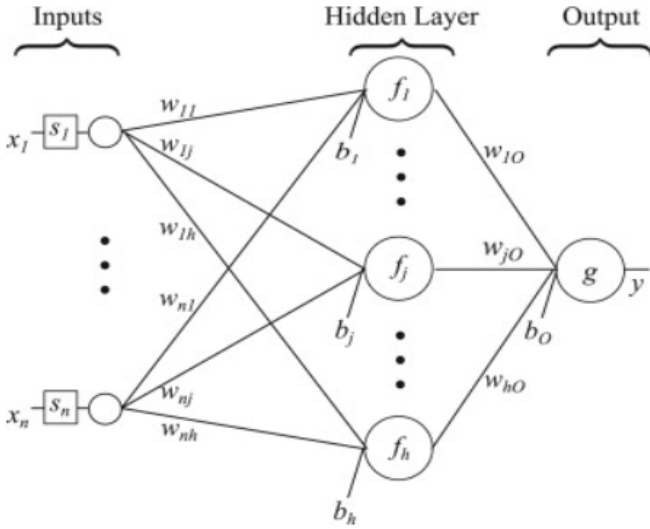


Fig. 1. Feed forward neural network

These models are called feedforward because information flows through the function being evaluated from x , through the intermediate computations used to define f , and finally to the output y . In this single hidden layer form there is only one layer of input nodes that send weighted inputs to a subsequent layer of receiving nodes [6]. The model approach is to use lagged values of the time series as input data, reaching to a non-linear autoregressive model.

For this approach, the specific number of hidden nodes is selected by this expression [4]:

$$N_h = \frac{N_s}{(\alpha(N_i + N_o))} \quad (5)$$

Where, N_i is number of input neurons.

N_o is number of output neurons.

N_s is number of training samples.

$\alpha = 1.5^{-10}$

IV. THE PROBLEM

The share data consists of date, highest/lowest value on that day, opening/closing value of that day and Volume of 6 different companies. These companies are Cognizant, HCL, HDFC, ICICI, Infosys and SBI. Cognizant, ICICI, HDFC and SBI have data from 2019-2021 and all other companies have data from May 2020 - 2021. The task is to predict future projections of the share value of these companies.

Two columns "Date" and "Close" value have been used for projections. For visualizations, another column "return" [7] is created which is $\log \frac{\text{Close share value on current day}}{\text{Close share value on previous day}}$. One null row is removed from all the datasets except the cognizant dataset. Columns "Close", "Volume", "Return" are used for visualizations.

Cognizant

Trend of Close share value

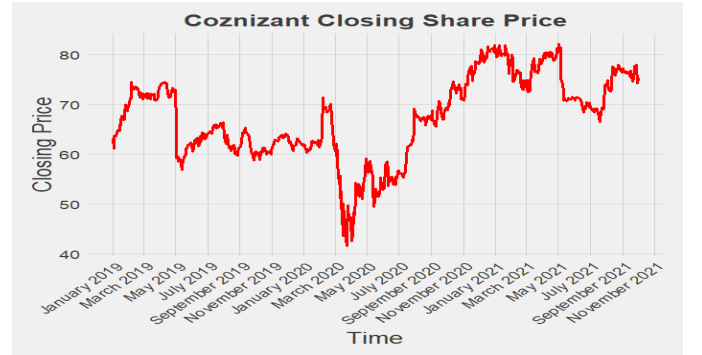


Fig. 2. Cognizant close share value

Volume(Amount of assets exchanged in a day)

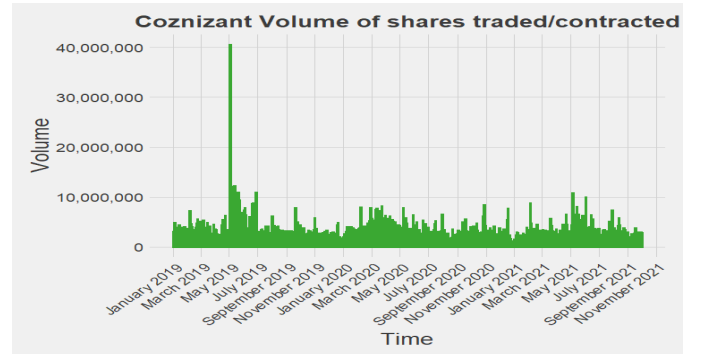


Fig. 3. Cognizant Volume

Log Return

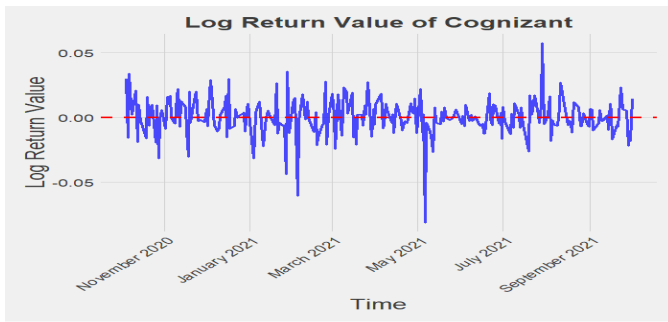


Fig. 4. Cognizant Log return

As expected, the share prices and log return are very volatile in nature. A major decrease in share prices is seen in March 2020 to May 2020 period (Fig 2), very likely due to first wave of COVID-19. Share prices also decreased during March 2021 to May 2021 period, the period of 2nd wave of COVID-19 but the effect was very less compared to 1st wave. Otherwise, there is no significant changes detected in the share prices. In general, the trend seems to be increase in share prices. The amount of shares exchanged (Fig 3), doesn't seem to change by much. During May 2019 there was some significant trading and exchange. Figure 4 shows the log return value of Cognizant. Overall the return value tends to be positive. A major negative return was seen in May 2021 period.

HCL

Trend of Close share value



Fig. 5. HCL close share value

Volume

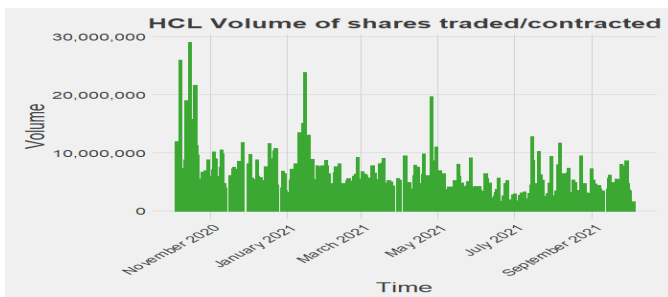


Fig. 6. HCL Volume

Log Return

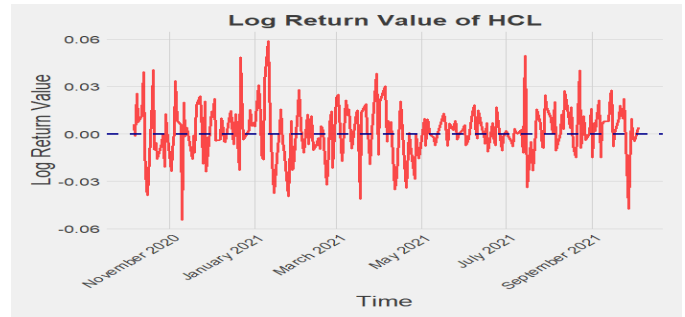


Fig. 7. HCL Log return

The share prices of HCL (Fig 5) show an increasing trend. Not much major decrease was seen in the given period. During COVID-19 waves as well not much change was detected. There is a small decrease seen in the amount of share traded/exchanges (Fig 6). Figure 7 shows the log return value of HCL. Overall the return value tends to be positive. A major negative return was seen in January 2021-May 2021 period.

HDFC

Trend of Close share value

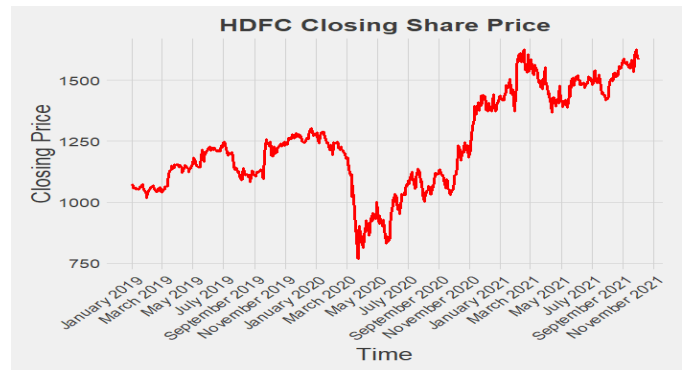


Fig. 8. HDFC close share value

Volume

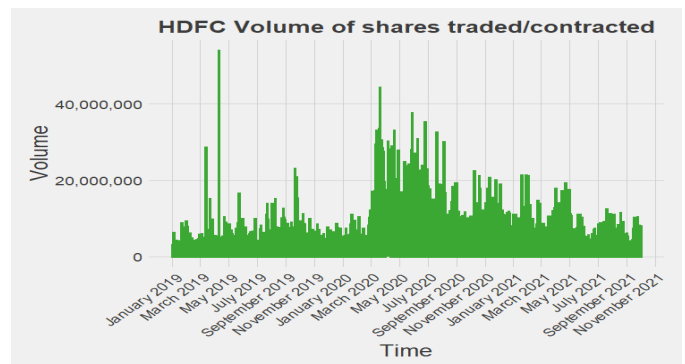


Fig. 9. HDFC Volume

Log Return

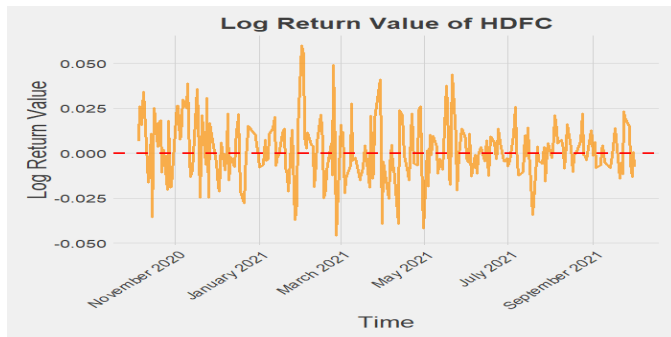


Fig. 10. HDFC Log return

The share prices of HDFC (Fig 8) also shows an increasing trend but a major decline was seen in March 2020-May 2020 period, likely due to COVID-19. The amount of shares exchanged (Fig 9) increased heavily in March 2020- July 2020 period also likely due to COVID-19, prices decreased and people bought more shares. The log return value (Fig 10) tends to be positive and it is very volatile during March 2021- May 2021 period.

ICICI

Trend of Close share value



Fig. 11. ICICI close share value

Volume

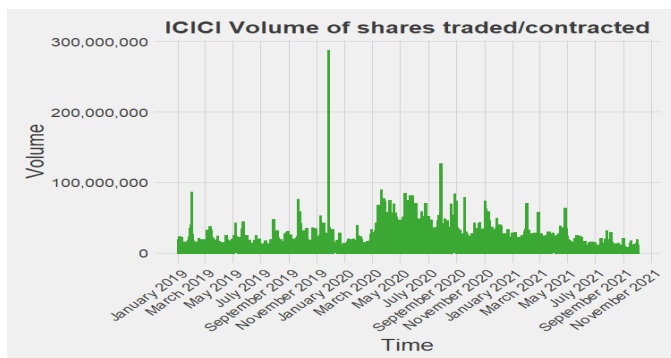


Fig. 12. ICICI Volume

Log Return

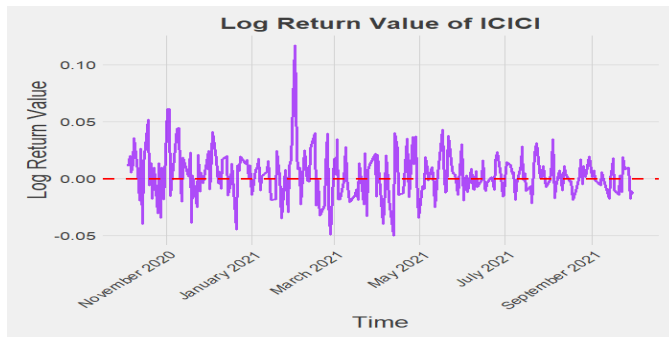


Fig. 13. ICICI Log return

Here also, the shares value of ICICI (Fig 11) shows an increasing trend with a major decline in prices during COVID-19 first wave. A small decline is also seen during 2nd wave of COVID-19 (March 2021 - May 2021 period). Almost the share exchange (Fig 12) remained constant but heavy increase in exchanges was seen in December 2019. The log return value (Fig 13) of ICICI is less volatile in later time compared to other companies.

Infosys

Trend of Close share value

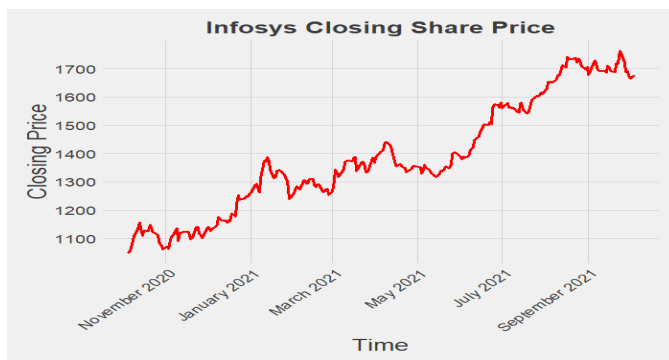


Fig. 14. Infosys close share value

Volume

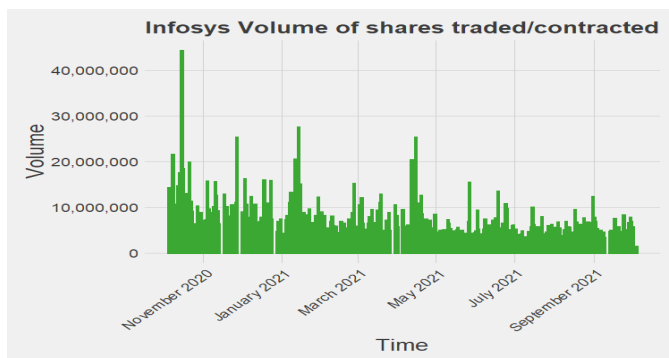


Fig. 15. Infosys Volume

Log Return

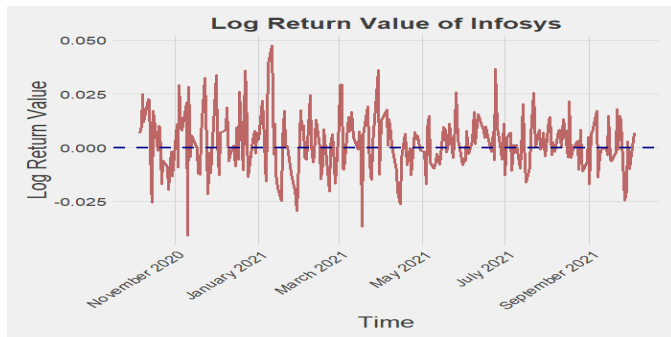


Fig. 16. Infosys Log return

The share prices of Infosys (Fig 14) also shows an increasing with no major sudden increase/ decrease changes observed. Currently, the amount of exchanges (Fig 15) of shares seems to have decreased. Sudden spikes in exchanges was observed in the figure. A major spike was observed in November 2020. The volatility in log return value (Fig 16) of infosys also seems to have reduced currently.

SBI

Trend of Close share value

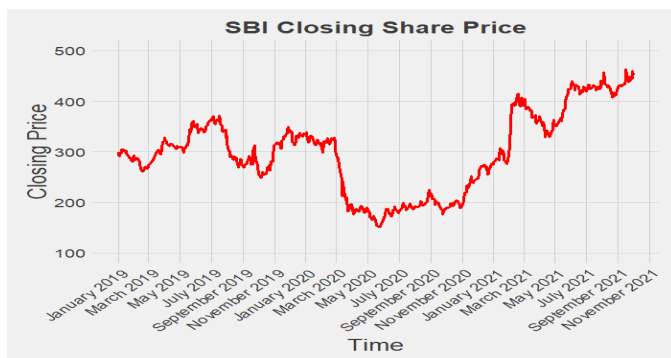


Fig. 17. SBI close share value

Volume

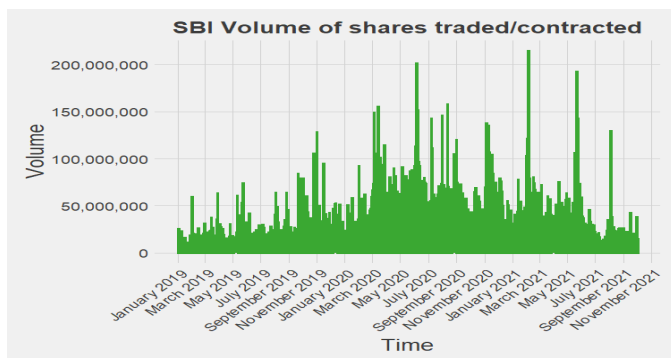


Fig. 18. SBI Volume

Log Return

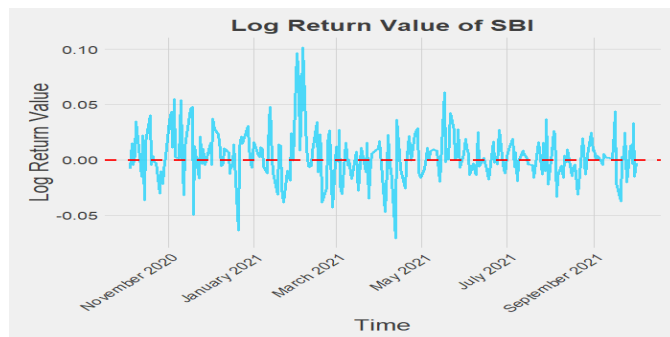


Fig. 19. SBI Log return

The share prices value (Fig 17) of SBI show an increasing trend currently. A major decrease in value was seen during March 2020 - May 2020 period, also a sudden small spike in values was seen in January 2021 period. A major exchange in shares (Fig 18) was seen in March 2020 - December 2020 period and then the amount decreased but sudden spikes in exchanges are seen frequently. The log return value (Fig 19) of SBI doesn't seem to be much volatile as well.

Overall, every company has a increasing trend of share prices with a major decline in prices seen during March 2020 - May 2020 period likely due to COVID-19. There was a decrease in prices observed during 2nd wave of COVID-19 (March 2021- June 2021) as well but decline was very much less compared to the 1st decline. The log return value of every company was volatile, some were less volatile and some were highly volatile. The banking companies such as ICICI, HDFC, SBI have more amount of shares exchanges compared to the service companies.

Forecasts of some companies are mentioned below. For the future projections of remaining companies see code documentation.

Cognizant Forecast

kNN model

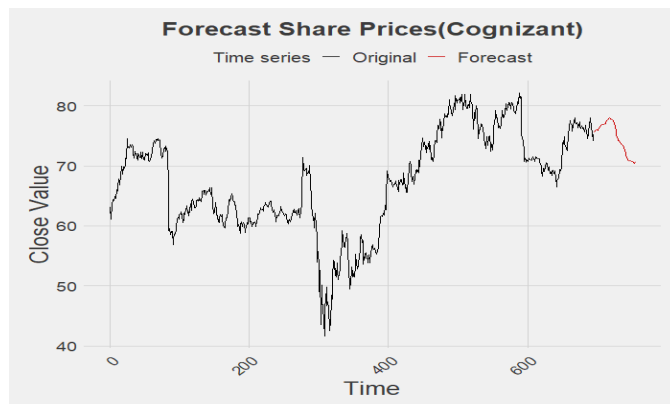


Fig. 20. Cognizant Forecast (kNN model)

Prophet Model

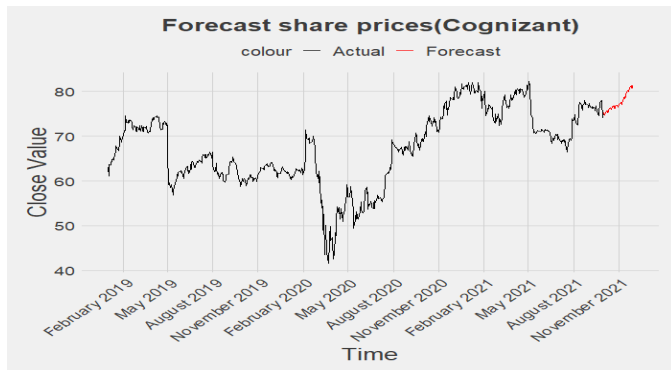


Fig. 21. Cognizant Forecast (Prophet model)

Prophet Model



Fig. 24. SBI Forecast (Prophet model)

Feed Forward Neural Network

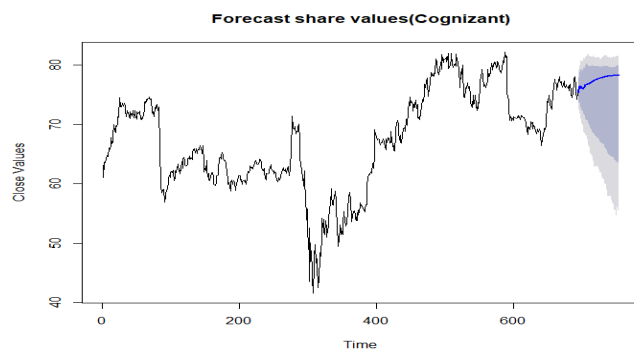


Fig. 22. Cognizant Forecast (Neural Network model)

Feed Forward Neural Network

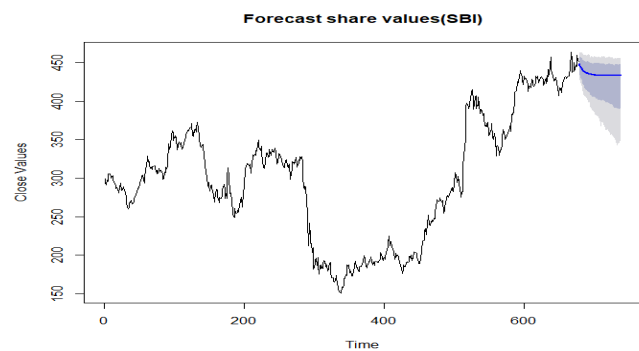


Fig. 25. SBI Forecast (Neural Network model)

The prophet model and neural network model both predicted increasing share prices but KNN model predicted decreasing prices. On the basis of evaluation metric Root mean squared error (RMSE) is considered. The RMSE of prophet model is 2.89, neural network model is 1.133, KNN model is 6.76.

Here, the neural network model predicts a slight decrease in prices, while Prophet model predicts an increase and knn model an increase and then small decrease in prices. The RMSE for prophet model is 18.28, neural network model is 7.11, KNN model is 18.98.

SBI Forecast kNN model

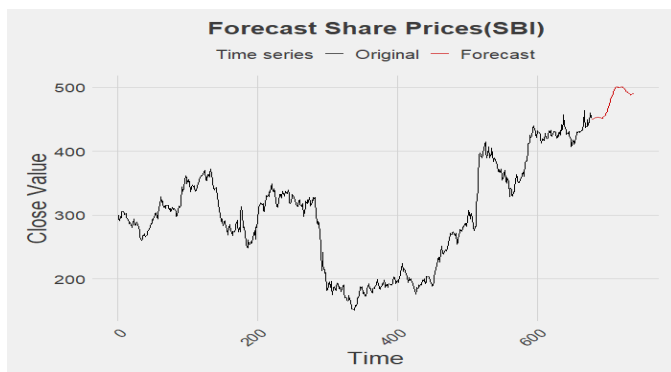


Fig. 23. SBI Forecast (kNN model)

HDFC

kNN model

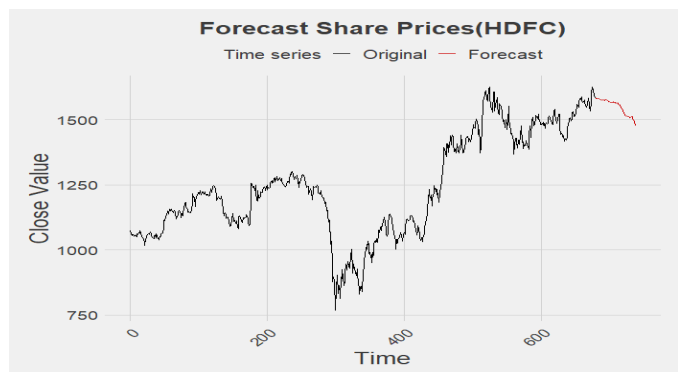


Fig. 26. HDFC Forecast (KNN model)

Prophet Model

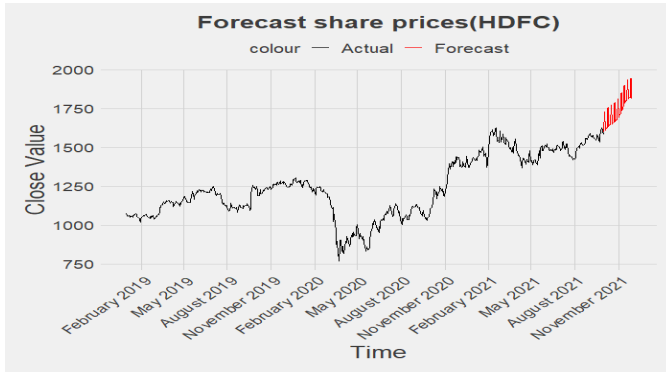


Fig. 27. HDFC Forecast (Prophet model)

Neural Network

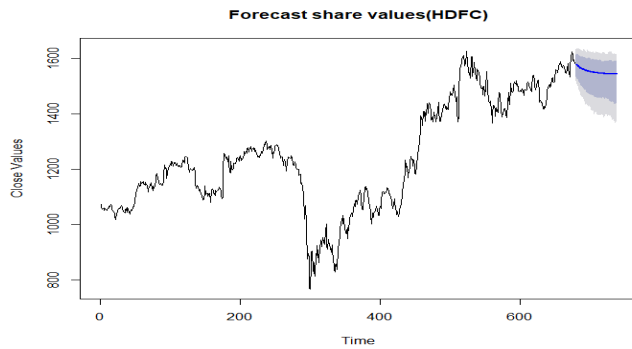


Fig. 28. HDFC Forecast (Neural Network)

Here, the prophet model predicts an increase in share while the KNN model predicts significant decrease and the neural network model predicts a slight decrease in prices. The RMSE for prophet model is 41.71, neural network is 20.21, KNN model is 96.41. Here also the least RMSE is of neural network and most is of KNN model. In general also, RMSE of neural network model is the least, then prophet model and maximum value of RMSE is observed in KNN model. This is same for all the other companies as well.

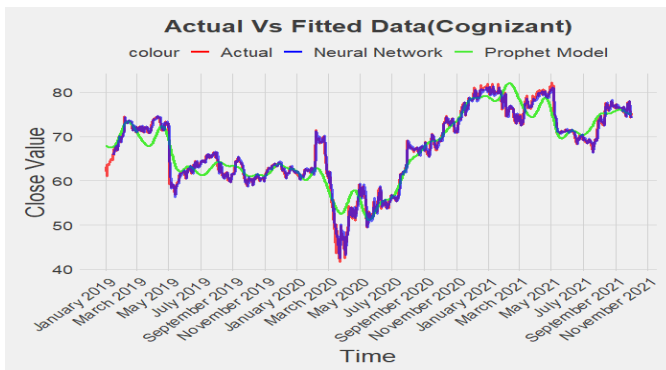


Fig. 29. Actual Vs Fitted Values (Cognizant)

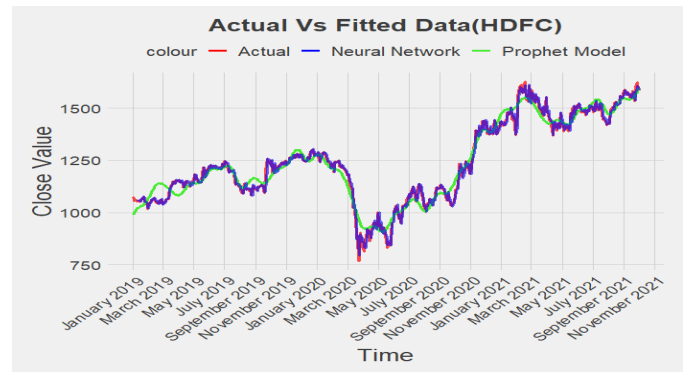


Fig. 30. Actual Vs Fitted Values (HDFC)

Figure 29 and 30 shows the fitted vs actual share value plot of Cognizant and HDFC respectively. From these plot it is seen that neural network fits the data very closely to the actual data, while the prophet model only captures the trend of actual data. The neural network model has high computational cost, so if one needs to perform the task as soon as possible, one should go for prophet model first as it is seen that this model captures the trend of actual data and that is fine initially, after tuning one make the model better. We learn that the neural network model works the best for the time series analysis but it is better to use prophet model if one has time constraint and computer memory constraint.

The prophet model actually provides some more insight on the data.

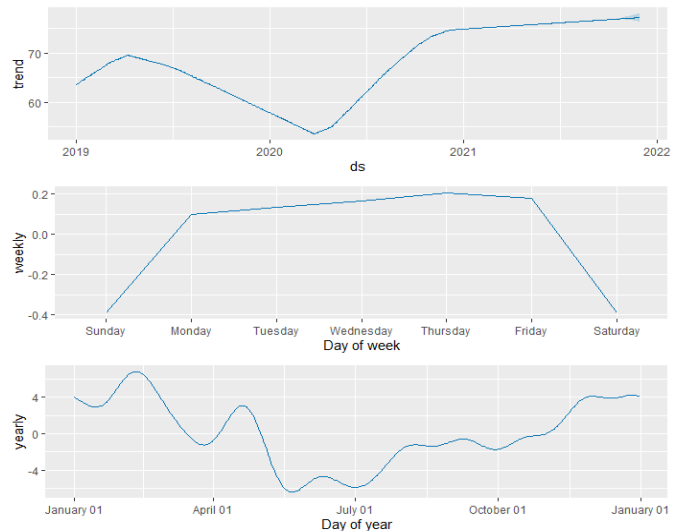


Fig. 31. Trends captured by prophet model(Cognizant)

In the 1st part of plot (Fig 31), it shows the general trend of share prices, 2nd part shows the trend on weekly basis, it is observed that share prices generally increases till Thursday and then decrease. It should be noted that values at Sunday and Saturday should be ignored as there is no calculation of prices on these days. The 3rd part of the plot shows the trend yearly,

it is observed that around May-June period, the prices were lowest, the time of COVID-19 1st wave . For other companies as well this type of plot is generated and they are present in the code documentation. These types of plot will help when more data is present. The cognizant dataset had data of 2 years while some companies have data of only 1 year so these kind of plot don't help very much but if data of previous 5 years was provided, much more better insights could be observed with the help of these plots.

If neural network model is used for forecasting, it is better to invest the shares in cognizant as it shows an increasing trend of share prices. There are many other factors while investing your shares in a company that should also be looked into. Due to COVID-19, all the companies suffered heavy losses. Factors like COVID-19 can never be predicted while forecasting.

V. CONCLUSION

If the models which were used are compared, neural network gave the best fit and KNN gave the worst fit. The prophet model was able to capture the trend of the data and if the required work requires less computational cost, this model can be effective and be used as well. From the projections which were made, it is better to invest in Cognizant. All the companies had a major decline in the share prices during March 2020- May 2020 period likely due to the effects of COVID-19. Banking companies such as ICICI, SBI, HDFC had a major decline in share prices during this period. This shows how certain factor can change the whole market.

Improvements in the work is possible. Other data such as highest prices, opening prices provided in the dataset can be used for other insights as well. If much more previous data was used, the models could have much more better fit. Other time series forecasting models like ARIMA can be used as well.

REFERENCES

- [1] <https://www.investopedia.com/terms/s/stock.asp>
- [2] <https://facebook.github.io/prophet/>
- [3] <https://rpubs.com/kapage/523169>
- [4] <https://www.geeksforgeeks.org/time-series-analysis-using-facebook-prophet/>
- [5] <https://www.r-bloggers.com/2020/09/time-series-forecasting-knn-vs-arima/>
- [6] <https://towardsdatascience.com/deep-learning-feedforward-neural-network-26a6705dbdc7>
- [7] https://rstudio-pubs-static.s3.amazonaws.com/78839_afca73ae18194eaf8f1b86d399dde969.html