**STOCK MARKET PREDICTION USING TIME SERIES FORECASTING MODELS**

**Abstract:**

Stock market prediction is one of the most challenging tasks since the financial time series is highly volatile, noisy, non-linear and dynamic in nature. We implement the Machine Learning concepts to predict the stock values based on the previous closed value of a Stock Price, using the Machine Learning algorithms like LSTM, ARIMA, Facebook Prophet models can help us to get better results for the prediction not only on the stock prices but also on the indices.

This Study aims to apply the Time series forecasting techniques to predict the stock price movement in the Indian Stock Market. We analysed the stock prices of the different sectors and applied these algorithms on the data of each symbol from each sector individually.

**Introduction:**

Stock Price Prediction means to determine the future value of the stocks or other financial instruments of an organisation. If you master the art to predict stock prices, you can earn a lot by investing and selling at the right time, and you can even earn by mentoring other people who want to explore trading. Over the years deep learning models such as Long-Term Short-Term Memory and statistical models such as Autoregressive Integrated Moving Average have shown promising results in predicting future stock prices. But the results from these models cannot be generalized as they fail to incorporate the dynamics of the market and influence of several external factors such as political, social, investor's emotion, etc on stock markets.

**Problem Definition:**

This research proposes a unique combination of the time series forecasting models like Facebook Prophet model, Long-Term Short-Term Memory model, ARIMA model etc to predict the adjacent closing price of stocks to fit both the seasonality and non-linearity component of stock price data.

**Literature Survey:**

*Combination of Facebook Prophet and Attention-Based LSTM with Multi- Source data for Indian Stock Market Prediction by Pavan Nagesh Technological University Dublin*

The primary objective of this research was to build a model that can accurately predict the adjacent closing price of NIFTY 50 stocks using a combination of both technical and fundamental analysis. The important aspect for building a stock market 6 forecasting model is gathering data that will reflect market and investor sentiment. This research collects historic stock price data and technical indicators to reflect market sentiments and news articles, tweets from Twitter data to reflect investor sentiments combining both technical as well as fundamental analysis.

*Using Neural Networks to Forecast Stock Market Prices, Ramon Lawrence*.

This paper is a survey on the application of neural networks in forecasting stock market prices. With their ability to discover patterns in nonlinear and chaotic systems, neural networks offer the ability to predict market directions more accurately than current techniques. Common market analysis techniques such as technical analysis, fundamental analysis, and regression are discussed and compared with neural network performance. Also, the Efficient Market Hypothesis (EMH) is presented and contrasted with chaos theory and neural networks. Finally, future directions for applying neural networks to the financial markets are discussed.

*Hybrid ARIMA-BPNN Model for Time Series Prediction of the Chinese Stock Market, Li Xiong, Yue Lu.*

Stock price prediction is a challenging task owing to the complexity patterns behind time series. Autoregressive integrated moving average (ARIMA) model and back propagation neural network (BPNN) model are popular linear and nonlinear models for time series forecasting respectively. The integration of two models can effectively capture the linear and nonlinear patterns hidden in a time series and improve forecast accuracy. In this paper, a new hybrid ARIMA-BPNN model containing technical indicators is proposed to forecast four individual stocks consisting of both main board market and growth enterprise market in software and information services sector.

Financial Indices Modelling and Trading utilizing deep learning techniques, Marios Mourelatos, Thomas Amorgianiotis, Christos Alexakos, Spiridon Likothanassis.

Prediction and modelling of the financial indices is a very challenging and demanding problem because its dynamic, noisy and multivariate nature. Modern approaches have also to challenge the fact that they are dependencies between different global financial indices. All this complexity in combination with the large volume of historic financial data raised the need for advanced machine learning solutions to the problem. This article proposes a Deep Learning approach

utilizing Long Short-Term Memory (LSTM) Networks for the odelling and trading of financial indices .

Hybrid Deep Learning Models for Stock Prediction, Mohammad Asiful Hossain, Rezaul Karim, Ruppa Thulasiram, Neil D B. Bruce, Yang Wang.

Stock market prediction has always caught the attention of many analysts and researchers. Popular theories suggest that stock markets are essentially a random walk and it is a fools game to try and predict them. Predicting stock prices is a challenging problem in itself because of the number of variables which are involved. This paper reviews all these points.

*Stock index forecasting based on a hybrid model, J.J. Wang, J. Z. Wang, Z. G. Zhang, and S. P Guo.*

This paper examines the prediction performance of ARIMA and artificial neural networks model with obtained stock information from New York Stock Exchange. The empirical results obtained reveal the prevalence of neural networks model over ARIMA model. The findings further resolve and clarify contradictory opinions reported in literature over the prevalence of neural networks and ARIMA model and the other way around.

**Dataset Description:**

To train the model, we are using the data from the API of yahoo financy. The data is the price history and trading volumes of the stocks in the index NIFTY 50 from [NSE (National Stock Exchange) India](https://www.nseindia.com/). All datasets are at a day-level with pricing and trading values The data spans from 1st January, 2000 to 31st March, 2021.

In this dataset, there are many columns but in our study, we will need the features like Date, Open, close and Symbol to predict the Adjacent close price of the day.

**Exploratory Data Analysis (EDA):**

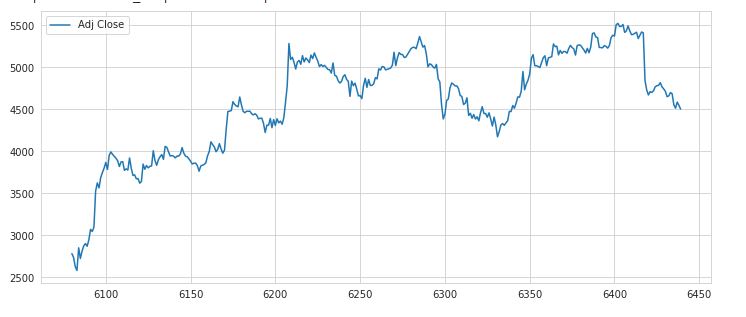
The stock Market data is pulled from the Kaggle and Yahoo Finance websites. Here we have studied the Pattern and behaviour of stock prices over the period of time. Using visualization or statistical tools data is explored in more detail to know the distribution of features, perform statistical analysis to come up with insights that help in the transformation of data to be used in the modelling stage.

The selected data is cleaned by removing inconsistencies, formatting the data types, handling missing entries using appropriate imputation methods, transforming data based on observation in the data exploration stage. Multiple data records collected from different sources are merged based on the requirement to create merged datasets

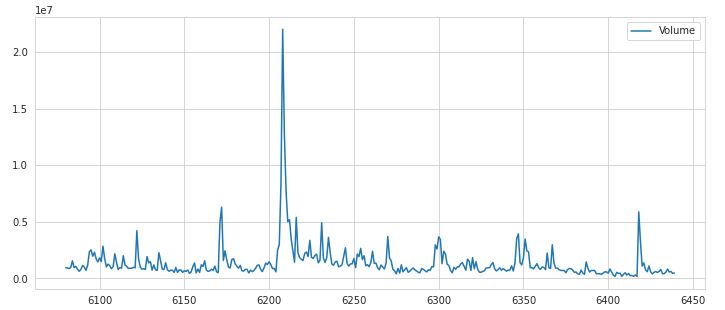
This strategy is based on the use of quantitative data which are historic price data of stocks and technical indicators that undergoes certain pre-processing steps before they are passed as an input to prediction models. Based on the period of prediction (short-term or long-term) data is collected on a daily, weekly, or monthly basis. The most common parameters in historic price data are open (Price of stock when the market opens), high (the highest price of a stock on that day, week, or month), low (lowest price of a stock on that day, week, or month), close (Price of stock when the market closes), volume (Volume of stocks sold) and adjacent close (Price of the stock after including dividends).

In this study we have explored the stock price pattern of all the four sectors (Banking, Pharmacy, Automobile and IT) in India. Each symbol from each sector is picked up to explore the pattern of data over the period.

Whenever the volume of the stock was increased, The price was also increased from that period. In fig 1 and fig 2, the closing price and volume of the DR Reddy’s is being shown. There is a clear indication here that whenever the volume was increased, the adj close price was also increased



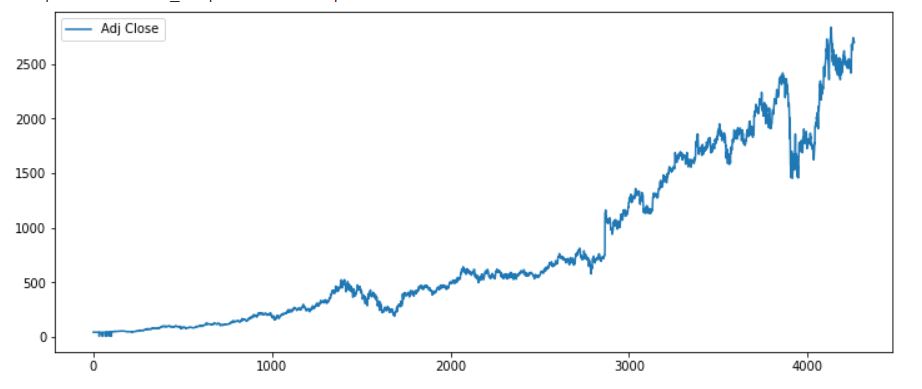
**Fig. 1 Adj. Close price of Dr. Reddy’s lab from March 15 2020**



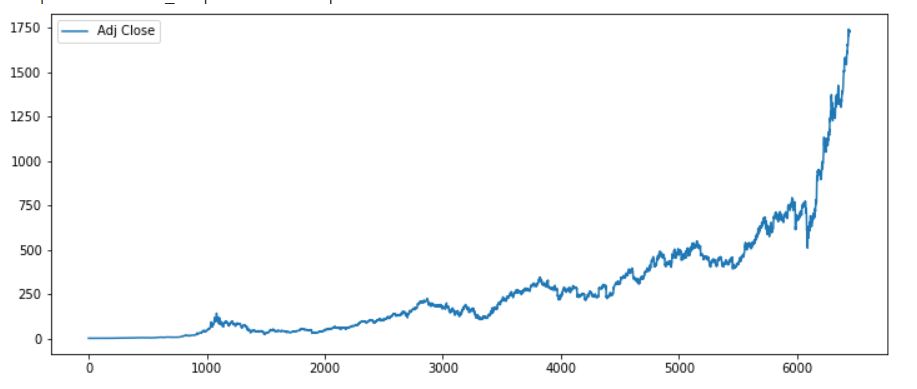
**Fig. 2 Volume of Dr. Reddy’s lab from March 15 2020**

We can see that Stock price and the volume both were low in the beginning of March. This is because it was the beginning of the pandemic and all the sectors were impacted because of the national wide lockdown.

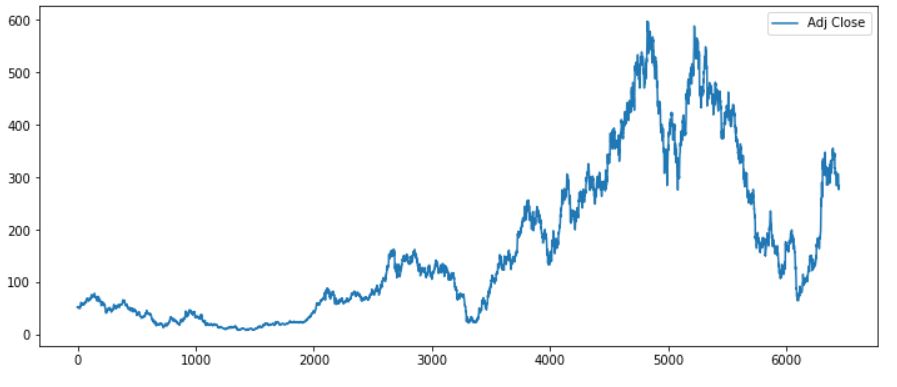
Here is the Stock price of each symbol from all the four sectors. The symbols include, HDFC bank from banking, Dr. Reddy’s from Pharmacy, Tata Motors from Automobile and Infosys from IT.



**Fig. 3 Adj. Close price of HDFC**

****

**Fig. 4 Adj. Close price of Infosys**



**Fig. 5 Adj. Close price of Tata Motors**



**Fig. 6 Adj. Close price of Dr. Reddy’s**

In Stock Markets, There are multiple factors involved for increasing/Decreasing of the stock price for the symbol. For example, The Dr. Reddy’s Stock price has started increasing when it signed a deal with Russian Pharmaceutical Company for Sputnik V. Once this news has come out, the volume has increased which resulted in sudden jump in the stock price.

**Methodology:**

In this study, We will be working on multiple algorithms and then compare the results of all the algorithms. This will give us the better results. Here are the list and description of each algorithms used in this study.

*The Facebook Prophet Model:*

Prophet 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.

The procedure makes use of a decomposable time series model with three main model components: trend, seasonality, and holidays. Similar to a generalized additive model, with time as a regressor, Prophet fits several linear and Non-linear functions of time as components. In its simplest form.

**y*(t)* = g*(t)* + s*(t)* + h*(t)* + *e(t)***

Where

g(t) is the trend models non-periodic changes (i.e. growth over time)

s(t) is the seasonality presents periodic changes (i.e. weekly, monthly, yearly)

h(t) ties in effects of holidays (on potentially irregular schedules ≥ 1 day(s))

e(t) covers idiosyncratic changes not accommodated by the model

In other words, the procedure’s equation can be written

***y(t)=piecewise\_trend(t)+ seasonality(t)+holiday\_effects(t)+noise***

*Trend***:** The procedure provides two possible trend models for g(t), "a saturating growth model, and a piecewise linear model."

*Saturating Growth Model:* If the data suggests promise of saturation i.e. one is wrestling constraints like: cubed footage, processing power, number of people w/ Internet access— setting growth='logistic' is the move.Typical modeling of these ***nonlinear, saturating trends*** is basically accomplished

where:

C is the carrying capacity

k is the growth rate

m is an offset parameter

**Rate of Change v. Time:**

Second, the market does not allow for stagnant technology. Advances like those seen over the past decade in handheld devices, app development, and global connectivity, virtually ensure that growth rate is not constant.

Because this rate can quickly compound due to new products, the model must be able to incorporate a varying rate in order to fit historical data.

We incorporate trend changes in the growth model by explicitly defining changepoints where the growth rate is allowed to change.

Suppose there are S changepoints at times sj, j = 1,…,S.

Prophet defines a vector of rate adjustments

δ ∈ Rs

Where

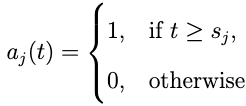
**δ***j* is the change in rate that occurs at time ***s****j*

The rate at any time ***t*** is then the base rate ***k***, plus adjustments up to that time

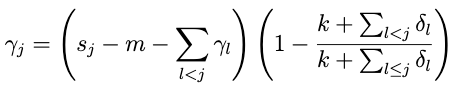
Image for postThis is represented more cleanly by defining a vector

∈ {0,1}s

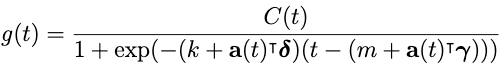
such that



The rate at time t is then k+a(t)ᵀδ. When the rate k is adjusted, the offset parameter m must also be adjusted to connect the endpoints of the segments. The correct adjustment at changepoint j is easily computed a

**

At last, the piecewise growth=‘logistic’ model is reached



An important set of parameters in our model is C(t), or the expected capacities of the system at any point in time. Analysts often have insight into market sizes and can set these accordingly. There may also be external data sources that can provide carrying capacities,such as population forecasts from the World Bank.

In application, the logistic growth model presented here is a special case of generalized logistic growth curves — which is only a single type of sigmoid curve — allowing the relatively straightforward extension(s) of this trend model to other families of curves.

*ARIMA Model:*

ARIMA, short for ‘Auto Regressive Integrated Moving Average’ is actually a class of models that ‘explains’ a given time series based on its own past values, that is, its own lags and the lagged forecast errors, so that equation can be used to forecast future values.

Any ‘non-seasonal’ time series that exhibits patterns and is not a random white noise can be modeled with ARIMA models.

An ARIMA model is characterized by 3 terms: p, d, q

where,

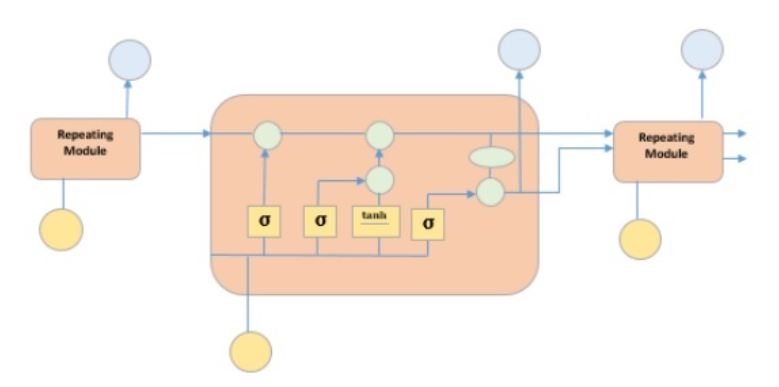
p is the order of the AR term

q is the order of the MA term

d is the number of differencing required to make the time series stationary

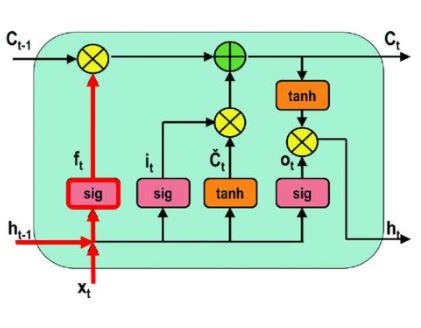
*LSTM Model:*

Long Short-Term Memory (LSTM) networks are a type of recurrent neural network capable of learning order dependence in sequence prediction problems. This is a behavior required in complex problem domains like machine translation, speech recognition, and more. LSTMs are a complex area of deep learning. It can be hard to get your hands around what LSTMs are, and how terms like bidirectional and sequence-to-sequence relate to the field.



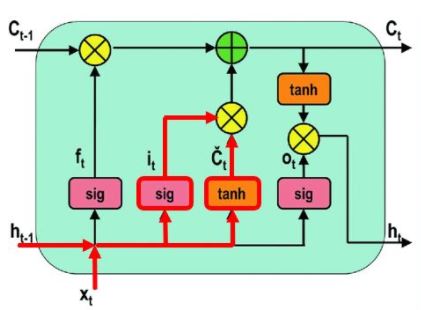
**Fig. 7 Structure of LSTM Model**

There are three different gates in an LSTM cell: a forget gate, an input gate, and an output gate.

****

**Fig. 8 The Forget Gate Operation in LSTM**

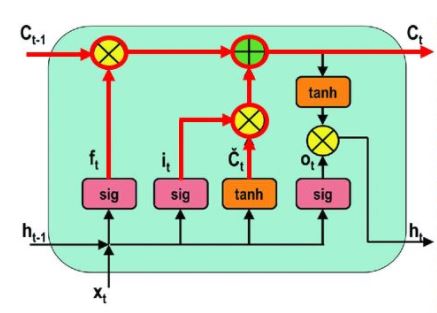
The forget gate decides which information needs attention and which can be ignored. The information from the current input X(t) and hidden state h(t-1) are passed through the sigmoid function. Sigmoid generates values between 0 and 1. It concludes whether the part of the old output is necessary (by giving the output closer to 1). This value of f(t) will later be used by the cell for point-by-point multiplication.

****

**Fig. 9 The Input Gate Operation in LSTM**

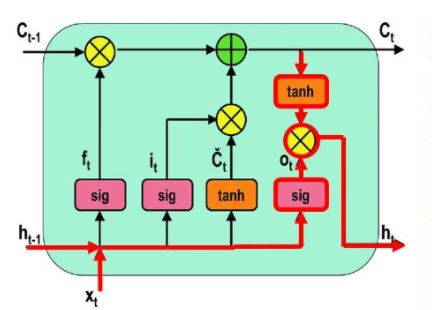
The input gate performs the following operations to update the cell status.First, the current state X(t) and previously hidden state h(t-1) are passed into the second sigmoid function. The values are transformed between 0 (important) and 1 (not-important).

Next, the same information of the hidden state and current state will be passed through the tanh function. To regulate the network, the tanh operator will create a vector (C~(t) ) with all the possible values between -1 and 1. The output values generated form the activation functions are ready for point-by-point multiplication.

****

**Fig. 10 The Cell State Operation in LSTM**

The network has enough information form the forget gate and input gate. The next step is to decide and store the information from the new state in the cell state. The previous cell state C(t-1) gets multiplied with forget vector f(t). If the outcome is 0, then values will get dropped in the cell state. Next, the network takes the output value of the input vector i(t) and performs point-by-point addition, which updates the cell state giving the network a new cell state C(t).

**

**Fig. 11 The Cell State Operation in LSTM**

The output gate determines the value of the next hidden state. This state contains information on previous inputs.

First, the values of the current state and previous hidden state are passed into the third sigmoid function. Then the new cell state generated from the cell state is passed through the tanh function. Both these outputs are multiplied point-by-point. Based upon the final value, the network decides which information the hidden state should carry. This hidden state is used for prediction.

Finally, the new cell state and new hidden state are carried over to the next time step.

To conclude, the forget gate determines which relevant information from the prior steps is needed. The input gate decides what relevant information can be added from the current step, and the output gates finalize the next hidden state.

**Results and Discussions:**

This section describes the portfolios constructed by the results

of the models. There were 400 stocks to select from and form a

portfolio each trading day. In this study, the portfolios were

constructed by ranking the prediction for all stocks each day

This section describes the portfolios constructed by the results

of the models. There were 400 stocks to select from and form a

portfolio each trading day. In this study, the portfolios were

constructed by ranking the prediction for all stocks each day

This section describes the portfolios constructed by the results

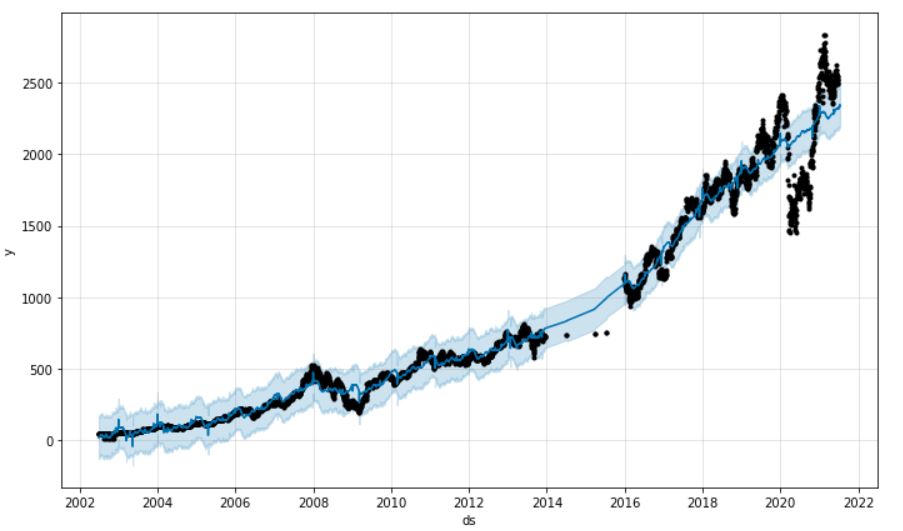
of the models. There were 400 stocks to select from and form a

portfolio each trading day. In this study, the portfolios were

constructed by ranking the prediction for all stocks each day

This section describes the portfolios constructed by the results of the models. The results of all the Three models in each sector will be discussed. We have developed Three Models and compared the results of each Model. The best result was from the LSTM Model because its nature. As this is stock Market Price prediction, The predicted result may not be much effective as multiple factors are involved in increase or decrease of the stock Market Price. The Holidays were also considered in the model building

Here is the result of FB prophet Model

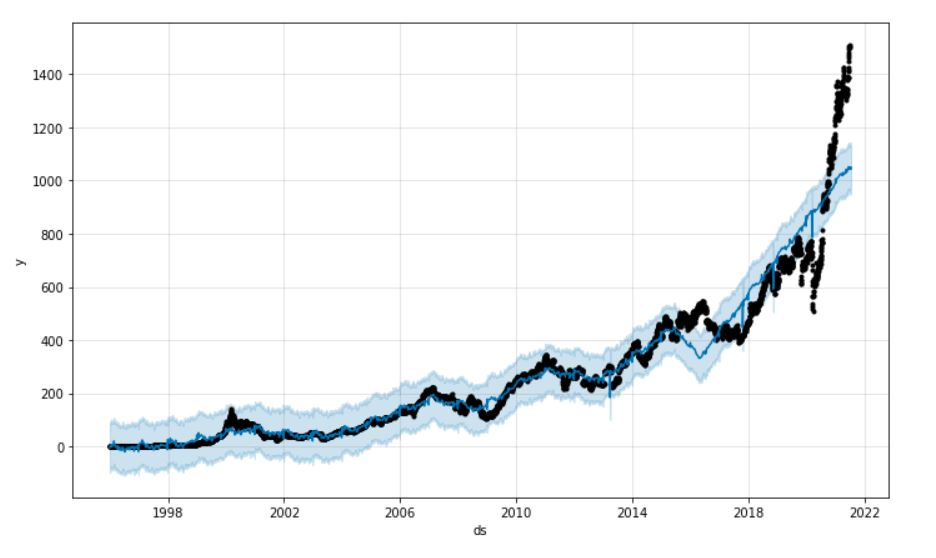


**Fig. 12 Comparison of Predicted results and Actual Results of Prophet Model on HDFC data**

In the above figure, the blue line is the predicted data of the Model. We have seen that at one point of time the predicted results were just based on the historical data and feature Engineering performed before the model but there is the sudden fall in the stock price, which is due to the covid19 pandemic.

When there was a national wide lockdown, almost all the sectors were impacted. We can see the Automobile industry was one of the major impacted area among these four sectors. Pharmaceutical sectors seems to be the least impacted sector during the beginning of the pandemic. Later once the Lockdown was over, the Stock Market was again in the peak. Currently we can see that all the prices in all the sectors are at its peak. The volumes after certain amount of time of the lockdown has increased as many investors realized the best time to invest in stocks.

Moving Forward our models predictions for all the symbols says that the price will increase. However, this cannot be the most accurate predictions, as all the factors needs to be considered which is almost an impossible.



**Fig. 13 Comparison of Predicted results and Actual Results of Prophet Model on Infosys data**

In Fig.12 and 13, We can clearly see the difference in the predicted data vs Actual data. There was a drop and sudden increase in the price. To improve this we have implemented LSTM models and the result was much better when compared to the other models. This is because the LSTM model can be tweaked as per the instructions of Financial Experts.

In the models like ARIMA and LSTM, more better features were added from the existing ones. The features like High low percentage and Percentage change has improved the results. The RMSE for these two models was less than 0.4 and for the Face book prophet Model, it was close to 0.5. This is because the facebook prophet model only works on the Date as the input feature and Adj Close and Predicted value.

**Future Work:**

Using Time series prediction techniques to forecast stock market prices will be a continuing area of research as researchers and investors strive to outperform the market, with the ultimate goal of bettering their returns. It is unlikely that new theoretical ideas will come out of this applied work. However, interesting results and validation of theories will occur as neural networks are applied to more complicated problems.

This study can be enhanced by integrating the our models with the Social media platforms to get the data and perform sentiment Analysis on this data. This can be further taken into the Dense layer to predict the stock price. This will give better results.

**References:**

1. Stock Price Prediction with Facebook Prophet Model - AMAN KHARWAL- AUGUST 9, 2020
2. Combination of Facebook Prophet and Attention-Based LSTM with Multi- Source data for Indian Stock Market Prediction by Pavan Nagesh Technological University Dublin
3. Agrawal, M., Khan, A. U., & Shukla, P. K. (2019). Stock Price Prediction using Technical Indicators: A Predictive Model using Optimal Deep Learning. International Journal of Recent Technology and Engineering (IJRTE), 8(2). doi: 10.35940/ijrteB3048.078219
4. Using Neural Networks to Forecast Stock Market Prices, Ramon Lawrence.
5. Hybrid ARIMA-BPNN Model for Time Series Prediction of the Chinese Stock Market, Li Xiong, Yue Lu
6. Stock index forecasting based on a hybrid model, J.J. Wang, J. Z. Wang, Z. G. Zhang, and S. P Guo
7. Selvin, S., Vinayakumar, R., Gopalakrishnan, E. A., Menon, V. K., & Soman, K. P. (2017). Stock price prediction using LSTM, RNN and CNN-sliding window model. nternational Conference on Advances in Computing, Communications and Informatics (ICACCI), 1643-1647. doi:10.1109/ICACCI.2017.8126078
8. Sousa, M. G., Sakiyama, K., Rodrigues, L., Moraes, P. H., Fernandes, E. R., & Matsubara, E. T. (2019). BERT for Stock Market Sentiment Analysis. IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI), 1597- 1601. doi:10.1109/ICTAI.2019.00231
9. Illustrated Guide to LSTM’s and GRU’s: A step by step explanation.
10. Generate Quick and Accurate Time Series Forecasts using Facebook’s Prophet (with Python & R codes) from Analytics Vidhya
11. <https://www.tutorialspoint.com/time_series/time_series_lstm_model.htm#:~:text=It%20is%20special%20kind%20of,layers%20interacting%20with%20each%20other>.
12. <https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/#:~:text=ARIMA%2C%20short%20for%20'Auto%20Regressive,used%20to%20forecast%20future%20values>.
13. T. Kimoto, K. Asakawa, M. Yoda, and M. Takeoka. Stock market prediction system with modular neural networks. In Proceedings of the International Joint Conference on Neural Networks, volume 1, pages 1–6, 1990
14. Manfred Steiner and Hans-Georg Wittkemper. Neural networks as an alternative stock market model. In Neural Networks in the Capital Markets, chapter 9, pages 137–148. John Wiley and Sons, 1995.
15. C. L. Wilson. Self-organizing neural network system for trading common stocks. In Proc. ICNN’94, Int. Conf. on Neural Networks, pages 3651–3654, Piscataway, NJ, 1994. IEEE Service Center.