Stock Price Prediction Using Deep Learning and Sentimental Analysis

R. Sathishkumar[1] R. Girivarman[2] S. Parameswaran[2] V. Sriram[2]

[1] Assistant Professor, Department of Computer Science & Engineering, Manakula Vinayagar Institute Of Technology, Puducherry,

[2]Student, Department of Computer Science & Engineering, Manakula Vinayagar Institute Of Technology, Puducherry.

***Abstract*–** The objective of this paper is to find an optimal technique for predicting stock price using the combination of fundamental and technical analysis using deep learning and sentiment analysis. we propose a method to predict stock price using sentiment analysis giving specific keywords and extracting news articles from google news which gives us the fundamental analysis is predicting the stocks and in order to perform technical analysis for the same, we check the accuracy of various statistical, machine learning and deep learning algorithms and models. Stock market is an important part of the economic system, which gives a major hand in the growth of the country through improving the economy. The stock market is the place where the trade of stocks of publicly listed companies takes place. The stock market impacts the GDP (Gross domestic product) of the country. GDP measures the output of all goods and services in an economy, which is a financial indicator of the country. As the stock market falls and rises, so does the sentiment of the country's economy. The keywords of the sentiment analysis includes WORLD ECONOMY, INDIAN ECONOMY, NIFTY(National Stock Exchange Fifty), SENSEX(Stock Exchange Sensitive Index), SEBI (Securities and Exchange Board of India)and the models subjected to technical analysis includes linear regression, moving average, ARIMA(Autoregressive integrated moving average) model a statistical machine learning model, fakebook's prophet model, and Long Short-Term Memory(LSTM).

Keywords – Stock market, financial indicators, LSTM, GDP, ARIMA, prophet, world economy, Indian economy, google news sentimental analysis, Nifty, Sensex, SEBI inflation, monetary, artificial intelligence, deep learning

1. INTRODUCTION

1.1 Stock Market:

The stock market is a place where shares of a public listed companies are traded. A stock symbolizes ownership in a company. Purchasing a company's stock means purchasing ownership in that company. Hence the investors who invest in a company's stock become a part of the company's overall profit or loss [1]. The stock of a company is also called as a script and the script changes according to various factors. The factors such as the demand of the product or service from the company, rumors in news, financial activities, political events, and various other mathematical factors[2].The above mentioned sentiment factors are called as fundamental analysis for stock prediction and the mathematical data according to historical data is classified as technical analysis.

1.1.1 Fundamental Analysis:

The two distinct trading philosophies for stock market prediction are fundamental and technical analysis [3]. While technical analysis focuses on the study of market actions through the use of charts, fundamental analysis concentrates on the economic forces of supply and demand that cause the stock price to move higher, lower, or remain the same [4]

There are few factors such as international environments, economic factors, political changes that influences the stock market across all types of stock. These can be put under a keyword such as world economy since news about the same would however affect economy or trade of the word one such example could be the recent Coronavirus disease (COVID-19) which affected the international environment and thus the international economy. By subjecting the keyword world economy into sentiment analysis, the rate at which the above-mentioned factor has affected the world economy can be found and further used for the prediction.

Few international factors having impacts to the market both positively and negatively are summarized below:

1. A key member of the economic world is subjected to a change in leadership.
2. Political instability across the countries
3. Symptoms of war between countries
4. Imposition of restriction on international trade of any country by US.

Similarly the keyword Indian economy can be used to evaluate the sentiment of the national factors , the factors affecting world trade or world economy not necessarily be affecting the Indian economy or sometimes impact India in an different aspect an example of this could be the recent Saudi-Russian oil war which had a negative impact to the world trade whereas it had a positive impact to the Indian economy. More ratio of data is collected for this when compared to the previous keyword thus giving greater weightage to this keyword

National Factors are those within India impacts the stock market in India. Few are summarized below:

1. Political changes in India and the states
2. Monthly IIP – Index of Industrial Production
3. Inflations Factors
4. Natural Disaster – Earthquake, Flood, etc.

There are multiple sectors such as banking, consumer durables, infrastructure including construction & power, information technology, automotive. Each sector is having it owns dependency with respect to market fluctuations. This can be categorized under the stock exchanges in which the script exists either NSE or BSE this can be further narrowed

down by getting to know how the trade of the top performing companies in each stock exchange. The keywords to perform the above mentioned are as following

- *NSE (National stock exchange)*: The National Stock Exchange of India Limited (NSE) is the prominent stock exchange of India which is located in Mumbai. The NSE was established in 1992. the first dematerialized electronic exchange in the country. It provides a neoteric, completely automated screen-based electronic trading system which provided easy trading facility to the investors spread over the length and breadth of the country.

- *NIFTY 50*: the top 50 stocks which are traded frequently in the national stock exchange limited by investors in India and around the world is seen as a barometer of the Indian capital markets. Nifty 50 index was launched in 1996 by the NSE. It is estimated that about 4% of the Indian economy / GDP is derived from the stock exchanges in India.

- *BSE (Bombay stock exchange)*:it was Established in 1875, the BSE (formerly known as Bombay Stock Exchange Ltd.) is Asia's oldest stock exchange. It is the world's 10th largest stock exchange. Located in the Dalal street in Mumbai. On 31st of August 1957, the BSE was officially recognized by the Indian Government under the Securities Contracts Regulation Act. In 1980, which turned out to be the first stock exchange in India.

- *BSE SENSEX*: it is also known as the S&P Bombay Stock Exchange or Sensitive Index or simply the SENSEX).it is a free-float market-weighted stock market index of 30 top traded and financially sound companies listed on Bombay Stock Exchange. The elemental 30 companies which are the largest and most actively traded stocks are the face of various industrial sectors of the Indian economy.

- *SEBI (Securities and Exchange Board of India)*: SEBI is the only regulator of securities market in India. It was established in 1988 and provided with Statutory Powers on 30 January 1992 through the SEBI Act, 1992. The function of SEBI is to protect the interests of the investors in securities, to encourage the development of stock exchange, and to regulate, the securities market.

A stock specific factor is one which that influences the market by the means of the performance of an industry or a company or a organization. In order to calculate the impact factor of script oriented sentimental calculation we choose the name of the script added with the share and price for knowing the appropriate news and collecting articles. The list of factors present in stock specific variables are:

1. Financial results include revenue, profit margin, Earning Per Share, Dividend Distribution.
2. New business deals
3. Formation of new joint ventures with multinational companies (Merge and New Acquisition)
4. Change in leadership like MD or CEO.

1.1.2 Technical analysis:

Technical analysis is the method of evaluating the future performance of securities by the identification of pattern and trend from past[5]. The statistical nature of analysis differs from functional analysis in the same regard. Fundamental analysis is future oriented, with focus on future. Earnings and risks while technical analysis is an analysis of the past behavior of the market which then is used to predict the future [6]. Technical analysis is done through various sort of methods such as. [5][7]

1. Moving average
2. Candlestick Agent
3. Linear Regression
4. ARIMA
5. Prophet
6. LSTM

In this paper we would study each and every method mentioned above irrespective of if it is a statistical or machine learning or deep learning model, and compare them and conclude to one model from which we take a model for the evaluation of technical analysis and at last would calculate in order to get the prediction of stock price with an greater accuracy. Moving average, moving average crossover agent, candle stick agent, linear regression comes under statistical modelling and used traditionally as technical analysis for stock price prediction. ARIMA (Auto Regressive Integrated moving average) and prophet comes under machine learning model. LSTM (Long short-term memory) falls under deep learning model. Apart from these above mentioned statistical, machine learning and deep learning models some other predictive models are verified in previous papers [8][9][10][11]. those include random forest [8], decision trees [9] and support vector machines[10]. and all of those provides a less accuracy than ARIMA and prophet model in machine learning thus we prefer evaluating other models as referenced in the starting of 1.1.2.

Technical analysis is a very subjective way of analysis with various variations available for the parameters used in Indicators of Technical Analysis. It is a rule-based technique with little scope for personal judgement, personal or social judgement affects the stock price to a greater extent thus we do fundamental analysis using sentiment analysis.

From brokers to intelligent trading system. Stock market prediction system has improvised a lot. To understand the influence of Machine Learning in Stock Market Prediction and financial Analysis the below section discusses about deep learning.

1.2 Deep Learning

Deep learning is a function of artificial intelligence that mostly imitates the workings of the human brain in handling and processing of data and creating patterns for the purpose of decision making. Deep learning is a subset of machine learning which is again the subset of artificial intelligence (AI) that has networks that are capable of learning unsupervised data which is either unlabeled or unstructured. Also known as deep neural network deep or neural learning.

Deep learning utilizes a hierarchical level of artificial neural networks to carry out the process of machine learning, a self-adaptive algorithm that gets increasingly better analysis and patterns with experience or with newly added data. The artificial neural networks are loosely built much like the

human brain, with nodes similar to biological neurons that are connected together like a web. traditional programs follow a linear pathway where the data is analyzed and built-in such manner whereas deep learning follows a non-linear approach through building a hierarchical function of networks.

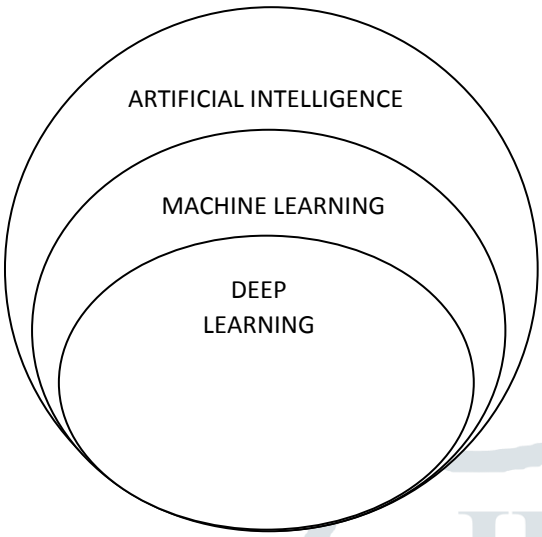


Fig 1: Differentiating Artificial intelligence, machine learning and deep learning

1.3 RELATED WORKS

1.3.1 A Prediction Approach for Stock Market Volatility Based on Time Series Data

The Authors Sheikh Mohammad Idrees, m. Afshar Alam, and Parul Agarwain in the year 2019 used ARIMA(Autoregressive Integrated Moving Average) a machine learning oriented statistical model to predict the volatility of a stock by predicting the stock price of the stocks for an appropriate period of time.

The predicted time series has been compared with the actual time series, which showed roughly a deviation of 5% mean percentage error for both Nifty and Sensex on average. for the validation of the predicted time series they have used the ``ADF test and the L-jung box tests" The system took long time to train but produced good accuracy.[12]

1.3.2 Predicting the direction of stock market prices using random forest

The Authors Luckyson Khaidem, Snehanshu Sahain and Sudeepa Roy Dey in the year 2016 used rando forest an supervised algorithm in machine learning to predict the stock market prices.

The system provided an accuracy of around 85-95% for certain stocks they used for evaluating. In this paper they have done a long-term prediction.[8]

1.3.3 Stock price prediction using support vector regression on daily and up to the minute prices

The authors had used SVR and measured its performance on various Brazilian, American and Chinese stocks with different characteristics, for example, small cap or blue chip. They used predictive variables and calculated it using TA indicators on asset prices. The results showed that the magnitude of the mean squared errors for the three common kernels in the literature, by using specific algorithm training strategies with different price frequencies of days and minutes. The results are contrasted with those of a random walk-based model. This study shows that using a fixed training set on daily prices [10].

1.3.4 Predicting Stock Prices Using Technical Analysis and Machine Learning

Mr. Jan Ivar Larsen in the year 2010 used technical analysis and machine learning to predict stock prices. a stock price prediction model has been created using concepts and techniques in technical analysis and machine learning. The resulting prediction model has been employed as an artificial trader that was used to select stocks to trade on any given stock exchange. The performance of the model was evaluated on stocks listed on the Oslo Stock Exchange. [7]

1.3.5 Efficacy of news sentiment for stock market analysis

From the research of the existing system we could classify the system into two parts. The phase one in the system would be the description of news based on sentiment analysis that gives end result as a polarity value either as positive or negative. In the phase two the system propagates the estimated model that takes count of the decisive and the abrogating news, variation with factual data as input.

From the study of the existing system the fractional count of sentiment values and variation gave a favorable prediction difference than the existing systems which has consummated accuracy level between 65% to 86.12%.

In future the system will take into account of the social media statistics, reports, sites for long term that directly influences the stock market by taking into account of large number of news data instances.[13]

1.3.6 Prediction Models for Indian Stock Market

Aparna Nayak.M, Manohara Pai M and Radhika M. Pai in the year 2016 had built an efficient predictive model of stock market prediction where the trend for the next day is predicted. By considering various patterns like continuous up/down, volume traded per day and including sentiment of the company a model was built and tested with different stock market data available open source.

The conclusion of their research work resulted such that On the considered dataset, Decision Boosted Tree was performing better than Support Vector Machine and Logistic Regression [2].

2. FORECASTING MODELS

This is to be noted that stock price is time dependent thus the prediction of stock price needs a time dependent prediction model and below mentioned are various time dependent models.

2.1 Simple Moving average

Moving average is the average of security price or exchange rate over a specific period of time. It is basically the mean of all values. It is called moving average because it keeps moving with time. For example, if the closing price of last 30 days is taken, summed together and then divided by 30,

the simple moving average is obtained. But these average keeps changing each day and the new price will be updated in the calculation of moving average every new day [14] .

A simple moving average is calculated by taking the average of prices over a specified period of time. Normally the closing price of stock is used in the average calculation [15] The five-day moving average means taking five days stock prices, summing it and then dividing the total by five. As mentioned above, the moving average keep changing every new day (.

Daily Closing Prices: 37,16,21,29,25,25,37

First day of 5-day SMA:
(37 + 16 + 21 + 29 + 25) / 5 = 25.6

Second day of 5-day SMA:
(16 + 21 + 29 + 25 + 25) / 5 = 23.2

Third day of 5-day SMA:
(21 + 29 + 25 + 25 + 37) / 5 = 27.4

The moving average of the first day is of the last five days. The moving average of second day replace the first price (37) with the new price (16). the moving average The moving average of the third day continues by replacing the first price of it (16) with the new one (21). It is evident from this example that prices have been gradually increasing from 37 to 37 in a time period of seven days. Also note that it has also been increased from 25.6 to 27.4 in three-day calculation period and the outcome of each moving average value is just below the last price. For the first day moving average one equals 25.6 and the last price is 27.4. So, the lag exists in simple moving average.
This is to note that the moving average calculated is for the yes bank rounded of to 2 digits from 05/03/2020 to 16/03/2020, and this is evident that moving average doesn't provide a prominent accuracy. The output of the stock price prediction for TATAGLOBAL using moving average is shown below in Fig .2 where in the graph blue denotes the graph for training and orange denotes the true close value for prediction and green is the predicted close price.

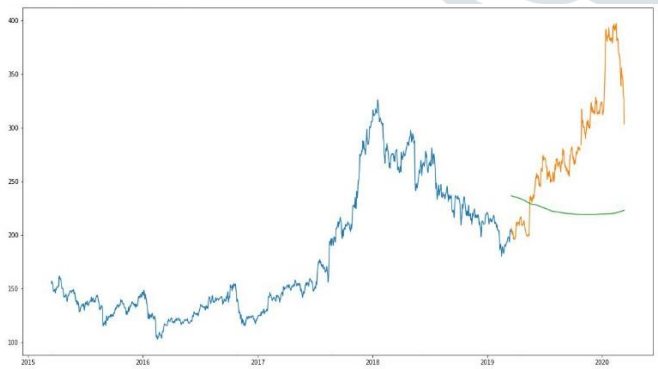


Fig 2: stock price prediction using moving average for the script TATAGLOBAL

2.2 Candle Stick Agent

Candlestick pattern analysis is a technique used in technical analysis to predict short-term trend reversals (Murphy,1999). Candlestick patterns are formations of typically one to three candlesticks. The Candlestick Agent holds a library of candlestick patterns that, when detected in the price history for some stock, generate buy and sell signals depending on the detected pattern [7].

Candle Stick is a technical analysis used chart which depicts the movement of stocks and thus provides a insight in predicting the stock and this is completely instinctive and doesn't provide a visual view of close price in future.

Fig. 3: Illustration of candle stick agent for the script TATAGLOBAL (Source: moneycontrol.com)

2.3 Linear Regression

Linear regression is one of regression method to be used for classifying numerical class [15]. It creates linear function by calculating weight values (w) for each feature (β). The function can be seen as follow:

$$x = \beta_0 w_0 + \beta_1 w_1 + \beta_2 w_2 + \cdots + \beta_n w_n \qquad (1)$$

X are regression value for one instance data. To have a clear understanding about linear regression, here is the illustration of linear regression[l7].

There are many ways to evaluate a linear regression model. Model which fitted most data has a normal distribution in its residual. the coefficient of determination value (R2) can also be evaluated just by looking to it. R2 is square of coefficient of correlation between predicted values and actual values. R2 ranged from 0 to 1. The greater the R2 value close to 1 is the greater the data that are fitted by the model ,

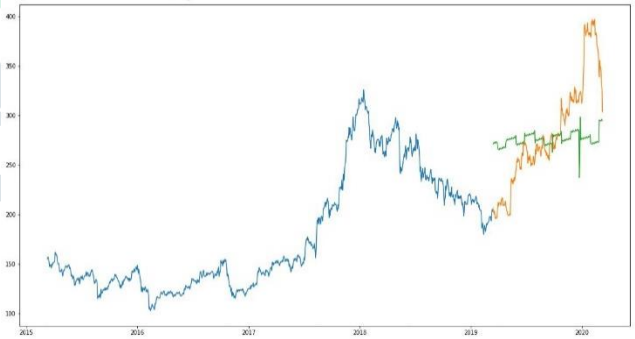


Fig 4: stock price prediction using linear regression for the script TATAGLOBAL

The output of the stock price prediction for TATAGLOBAL using linear regression is shown above in Fig .4 where in the graph blue denotes the graph for training and orange denotes the true close value for prediction and green is the predicted close price. It is evident from the graph that it is not an optimal model to implement technical analysis through it.

2.4 ARIMA (Auto Regressive Integrated Moving Average)

This model is also often called by the famous ``Box-Jenkins model". The ``ARMA model" is best suited for the stationary time series data but the thing is that most of the time series data from real world shows nonstationary behaviour. This model claims that a non-stationary series

could be changed to stationary by means of differencing it [18], [19]. The common form of an ``ARIMA model'' for y_t is as:

$$\text{“}y_t = (c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \epsilon_t) + (\mu + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + .. + \theta_q \epsilon_{t-q} + \epsilon_t)\text{”}$$

$$y_t = c + \epsilon_t + {}_{i=1}\Sigma^{p} \phi_1 y_{t-1} + {}_{j=1}\Sigma^{q} \theta_j \epsilon_{t-j} \quad (2)$$

where y_t is ``differenced time series'' that might have been differenced once or more.

This model is called as ARIMA (p,d,q)' model, in which the parameters p, d, q represents the following:

`p' is the ``order of auto-regressive part'',
`d' is the ``degree of differencing'',
`q' is the ``order of the moving average part''.

The ``ARIMA model'' combines 3 basic methods these are as follows:

2.4.1 AUTO-REGRESSIVE (AR)

This species that the values of the given time series is to be regressed with its own lagged value. This is specified through the ‘p’ value of an ARIMA model.

2.4.2 DIFFERENCING PART (I FOR INTEGRATED)

Integration is implied as the inverse of differencing. It is the degree on differencing that needs to be done on data. In order to transform a ``non-stationary time series into a stationary one'', the series needs to be differenced. Differencing can be implied as:

D(i) = Data (i) - Data(i-1).

This differencing of the time series is represented by the `d' value of the ``ARIMA model''. There are some situations that may arise with the value of `d', likewise below:

When `d' D 0, it signi_es that the series under consideration is stationary, so we don't require to take the difference of it.

If `d'D1, it signi_es that the current series is not stationary, we need to take the _rst difference of the series.

If `d' D 2, it signi_es that the series under consideration has been differenced two times.

2.4.3 MOVING AVERAGE (MA) in ARIMA

The ``moving average'' component of an ARIMA model is denoted by `q'. This simply refers to the total number of lagged values of the error term. For instance, when `q' D 1 it means that there exists an error term and there is autocorrelation with one lag.

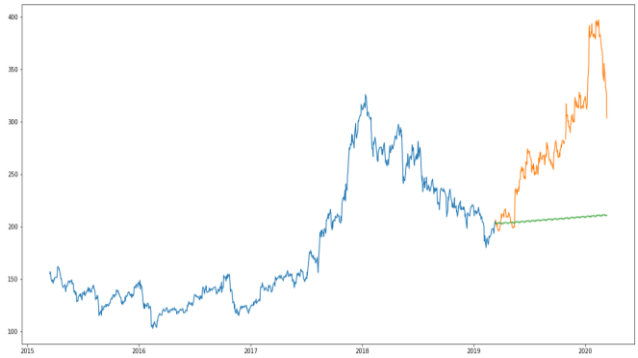


Fig 5: stock price prediction using ARIMA(p,d,q) model for the script TATAGLOBAL

The output of the stock price prediction for TATAGLOBAL using ARIMA model is shown above in Fig .5. where in the graph is similar to that of mention in 2.3. It is evident from the graph that it is not an optimal model to implement technical analysis through it.

2.5 Prophet Model

Stock price is time dependent factor thus a time series and prophet model is exclusively built by the tech giant Facebook to forecast time series data. Prophet model as it is prominently used for time series forecasting has many benefits such as weekly, daily, hourly observations can be predicted with a least or few month of history and it could detect important holidays in the respective calendars. the main objective of creating prophet model is to make the work of forecasting easy for the non-experts in the respective fields.

Prophet model is documented to be an additive regression model with four main components:

- Prophet detects changes in trends from the data by selecting the checkpoints.
- Yearly seasonal component is modelled using Fourier series
- Weekly seasonal component is modelled by dummy variables
- A list of holidays can be generated by the user profile

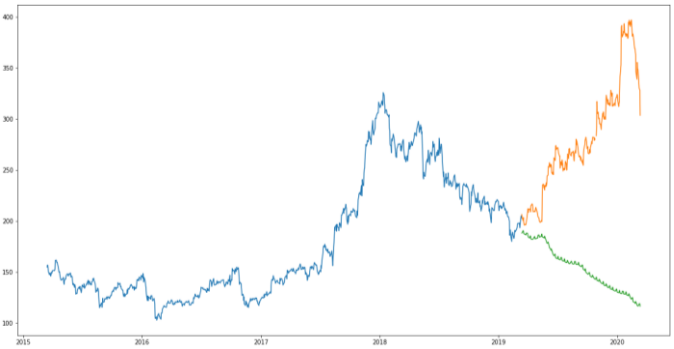


Fig 6: stock price prediction using prophet model for the script TATAGLOBAL

The fig .6 shows the prophet model for the script tata global using prophet model and this indicates that the predicted is totally off the actual close price and thus fails to be an appropriate model to be used for technical analysis.

2.6 Long short-term memory (LSTM) model

Long short-term memory network (LSTM) is a specific form of recurrent neural network (RNN), which is the general term for a series of neural networks that is capable of processing sequential data. LSTM is a special network structure with three "gate" structures. The three gates placed in an LSTM unit, are called input gate, forgetting gate and output gate. While information enters the LSTM's network, it can be selected by rules. Only the information conforms to the algorithm will be left, and the information that does not conform will be forgotten through the forgetting gate [20].

A very large architectures can be successfully trained using LSTM thus used in deep learning. LSTM is used instead of RNN in order to overcome the vanishing gradient problem in RNN. In LSTM neurons or memory block which is connected through layers. Each memory block is made of three gates these are

- Forget Gate: depending on the given set of conditions makes decision on throwing certain information out of the block.
- Input Gate: decisions are taken for updating the input values into the next memory state
- Output Gate: conditionally decides what should be the output based on input and the memory of the block

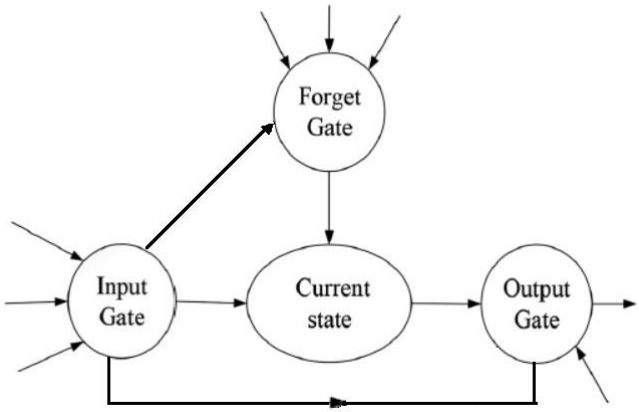


Fig 7: LSTM model

Fig .7 depicts the cycle of a LSTM unit. Each memory block is a micro machine where the gates have weights that are being learned in the training of the process. Fig.8 shows the stock price prediction using LSTM for the script tataglobal

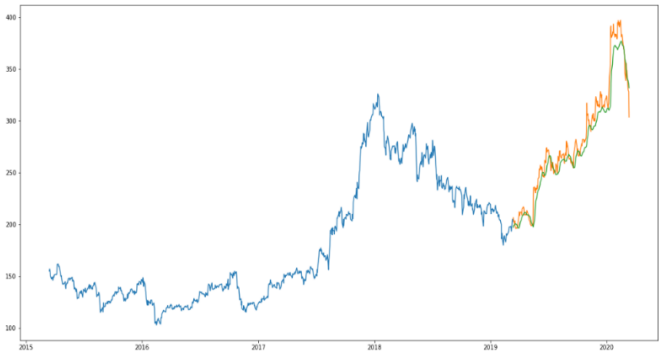


Fig 8: stock price prediction using LSTM model for the script TATAGLOBAL

From the fig. it is evident that when compared to all the other model we have seen in 2 LSTM outperforms all other predictive models for weekly and monthy trading. The error rate at each model by Root Mean Square value (RMSE) value is shown as a histogram in the fig 9.

RMSE is a widely used measure of the differences between values population values and the sample predicted that is the values predicted by the model and the values actually observed.

$$RMSE = \sqrt{(f-o)^2} \quad (3)$$

Where, f = predicted value
o = actual values

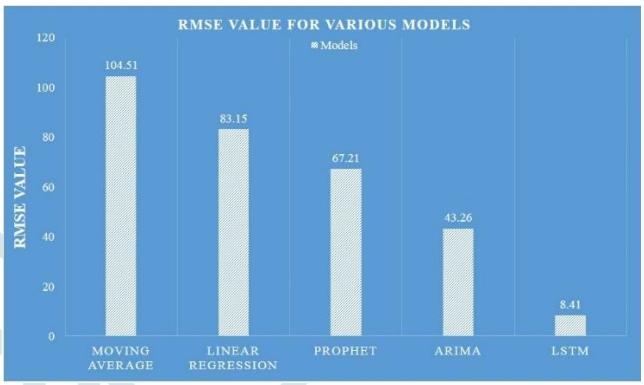


Fig 9: RMSE values for some predictive models

Lower the root mean square higher the accuracy thus LSTM provides a greater accuracy when compared to all the other models thus it is the appropriate model for conducting technical analysis.

3.PROPOSED SYSTEM

The stock market prediction proposed here uses LSTM and sentimental analysis. the system uses sentiment analysis for analyzing the market news based on certain keywords that covers the factors influencing the stocks. These keywords are discussed in introduction part of this paper. A python program has been written and the predicted values where returned as output. The sentimental analysis is performed by natural language processing. Before moving to the results obtained let's discuss how the idea flow is.

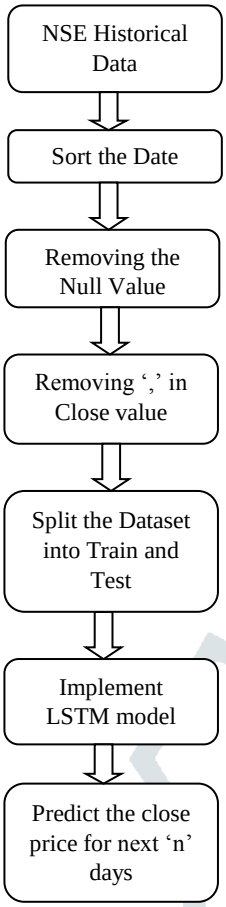


Fig 10: Flow Diagram for technical analysis in proposed system

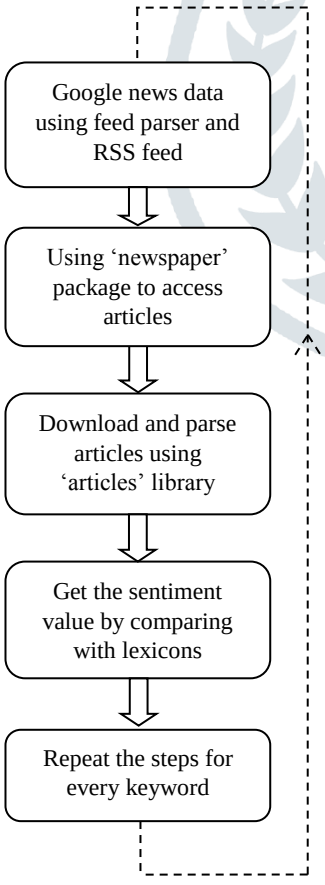


Fig 11: Flow Diagram for sentiment analysis in proposed system

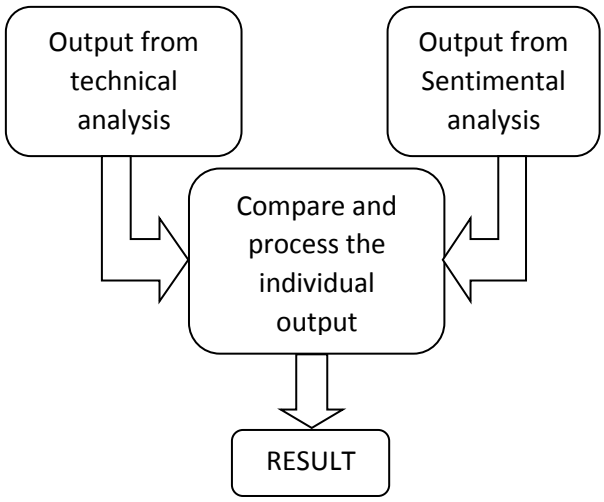


Fig 12: Flow Diagram for producing result in proposed system

The proposed system is a three-phase process that consists of technical analysis, sentimental analysis and a combination of results that is being processed and resulted to improve the accuracy that would be beneficiary for the traders thus improving providing a intelligent trading system. It involves deep learning and sentimental analysis from google news thus providing a real time

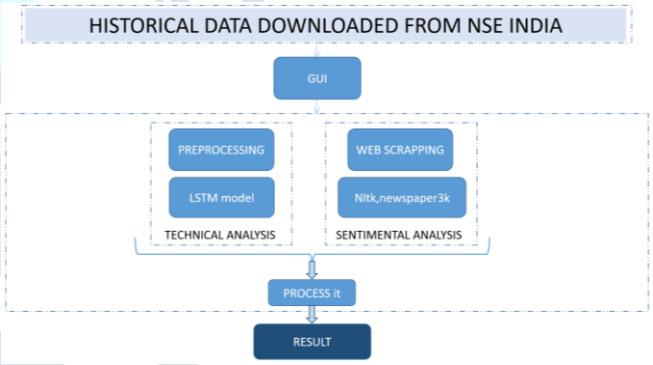


Fig 13: architectural Diagram for proposed system

3.1 DATA COLLECTION

The Historical stock market data is collected from the ‘nseindia’ website in a comma separated variable format and is given as input. Sentiment analysis data is collected through web scrapping using feed parse to get the rss(Really Simple Syndication) feed of google news using specific keyword search and the article is downloaded and parsed using newspaper package in python and article module from the same package and is subjected to natural language processing using natural language processing toolkit (NLPTK), and TextBlob library in python.

	Date	series	OPEN	HIGH	LOW	PREV. CLOltp	close	vwap	52W H	52W L	VOLUME	VALUE	No of trades
2	11-Mar-20	EQ	64,200.00	65,687.30	62,703.00	64,261.85	63,304.10	63,589.85	63,640.54	73,565.70	51,586.30	15621	99,41,28,924.85	9050
3	09-Mar-20	EQ	65,800.00	65,800.00	63,052.45	66,348.15	64,050.35	64,261.85	64,383.43	73,565.70	51,586.30	9596	61,78,23,350.40	6054
4	06-Mar-20	EQ	66,389.00	66,500.20	63,858.55	67,325.15	66,450.00	66,348.15	65,718.23	73,565.70	51,586.30	9851	64,73,90,290.60	6019
5	05-Mar-20	EQ	67,300.05	68,169.90	66,809.50	67,053.35	67,500.00	67,325.15	67,476.49	73,565.70	51,586.30	5745	38,76,52,448.60	3748
6	04-Mar-20	EQ	67,270.00	67,698.80	66,100.00	67,740.70	66,900.00	67,053.35	66,778.30	73,565.70	51,586.30	8187	54,67,13,963.20	5557
7	03-Mar-20	EQ	66,799.00	68,558.35	65,600.00	65,976.30	67,538.35	67,740.70	66,889.93	73,565.70	51,586.30	8202	54,86,31,205.20	5618
8	02-Mar-20	EQ	66,891.00	67,595.90	65,551.00	66,393.20	65,900.00	65,976.30	66,740.93	73,565.70	51,586.30	6419	42,84,10,051.55	4652
9	28-Feb-20	EQ	66,400.00	67,296.65	65,315.00	67,559.45	66,111.00	66,393.20	66,154.10	73,565.70	51,586.30	11768	77,85,01,435.25	7361
10	27-Feb-20	EQ	68,180.00	68,339.95	66,730.70	68,140.00	67,800.00	67,559.45	67,338.71	73,565.70	51,586.30	6856	46,16,74,171.15	4351

TABLE I: Sample dataset for MRF company stock

The data's in dataset are explained as below

- *OPEN:* The price of the first trade for any company or a script in a stock exchange is its opening price for that day.
- *CLOSE:* The volume weighted average of all the trades happened during the last half an of trading price.
- *HIGH:* The highest price for which a stock is traded on a script in a day.
- *LOW:* The lowest price for which a stock is traded on a script in a day
- *PREV.CLOSE:* The close price of the previous active day.
- LTP: Last Traded Price(ltp) is the last traded price of the script in the day.
- *VWAP*: Volume Weighted Average Price(vwap) is the average of high, close, low.

  VWAP = (High + Low + Close) / 3
- 52WL: The least traded price of a script in a year
- 52WH: The highest traded price of a script in a year
- *VOLUME*: No. of stocks traded in a day
- *VALUE:* Total price for which the stocks are traded in a day.
- *No of trades:* Total trade happened in a day for a script

3.2 DATA PREPROCESSING

The data gathered for sentiment analysis from google news forum is processed using Natural Language processing (NLP). The data is an article thus in order to preprocess it article library from newspaper package is being used to download the article and parse it in order to put the article to get the sentiment of the article. After the data is being downloaded it is in the text format, which is then split into words. These solitary words are put in comparison with the dictionary of words containing human-defined lexicons, which is identified as either positive, neutral or negative. Other general emotion symbols and special symbols are also extracted from the data and are further used to classify the particular text from the data as positive, neutral or negative.

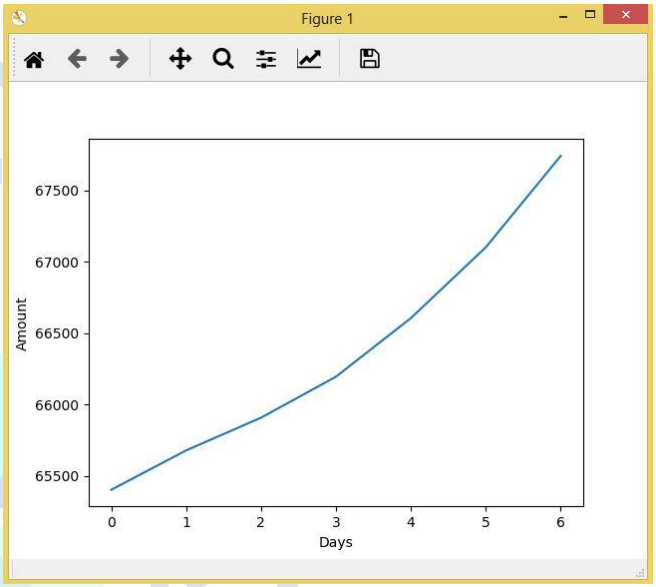
The TextBlob library from Natural Language Processing Tool Kit (NLPTK) are used to determine the subjectivity and polarity of the particular texts. After the subjectivity and polarity are determined these values are stored in a dataframe which is a multi-dimensional array.

The data gathered from the historical dataset is processed by removing the null values that would affect the model to an greater extent. Data visualization is done as and then needed by plotting graphs and plots for the given datasets using Matplotlib library. The rows of data with null values are removed. the rest of the data is stored in a separate data frame and this data frame is further used for fitting it into the LSTM model.

Once the stock price is predicted using LSTM model then the value resulted from the sentimental analysis is processed by taking the value predicted from LSTM into account. If the predicted value is lesser than 100 then the sentiment value is rounded of to one digit or if it is greater than 100 and lesser than 1000 then the sentimental value is rounded of to two digits. Or if the predicted value is in ten thousands then the sentiment value is rounded of to three digits and is subtracted from the LSTM predicted value if the sentiment is negative or vice versa.

4.EXPECTED OUTPUT

The proposed system was analysed with a sample stock. And the predicted price for the stock is as shown below:



	A
1	65401.742
2	65678.11
3	65907.15
4	66194.37
5	66604.76
6	67100.875
7	67741.555

5. CONCLUSION

The system used for stock price prediction combining the long short-term memory for technical analysis and sentimental analysis for fundamental analysis successfully gave good accuracy. Sentimental analysis was done based on the keyword narrowed down. This system would be useful for the user who are unaware of stock market. People with minimum or optimum trading experience can also use this system for stock price prediction in future. In future the system would enable intra-day prediction, improve sentimental analysis to get sarcasm free impact factor for fundamental analysis.

6. REFERENCES

[1] shashank Singh, maaz Ahmad, aditya Bhattacharya, m. Azhagiri ; Predicting Stock Market Trends using Hybrid SVM Model and LSTM with Sentiment Determination using Natural Language Processing. International Journal of Engineering and Advanced Technology (IJEAT)

[2] Aparna Nayak, M. M. Manohara Pai∗ and Radhika M. Pai; Prediction Models for Indian Stock Market, Twelfth International Multi-Conference on Information Processing-2016 (IMCIP-2016)

[3] E. F. Fama, Random Walks in Stock Market Prices, Financial Analysts Journal, vol. 51, no. 1, pp. 75–80, (1995).

[4] S. J. Grossman and R. J. Shiller, The Determinants of the Variability of StockMarket Prices, National Bureau of Economic Research,Working

[5] Mohd Naved ; Technical Analysis of Indian Financial Market with the Help of Technical Indicators, International Journal of Science and Research (IJSR,2013)

[6] Ramadoss, G. & Muthuvel, G. (2013), „Application- of Moving Average as Technical Indicator to Predict Stock Market Direction", International Journal of Management, IT and Engineering,

[7] Jan Ivar Larsen, Predicting Stock Prices Using Technical Analysis and Machine Learning,2013.

[8] Luckyson Khaidem, Snehanshu Saha, Sudeepa Roy Dey ; Predicting the direction of stock market prices using random forest, Applied Mathematical Finance

[9] Shou-Hsiung Cheng, Predicting Stock Returns by Decision Tree ,Combining Neural Network, Department of Information Management, Chienkuo Technology University, Changhua 500, Taiwan

[10] Bruno Miranda Henrique, Vinicius Amorim Sobreiro*, Herbert Kimura, Stock price prediction using support vector regression on daily and up to the minute prices, The Journal of Finance and Data Science 4 (2018),

[11] K. Hiba Sadia, Aditya Sharma, Adarrsh Paul, SarmisthaPadhi, Saurav Sanyal; Stock Market Prediction Using Machine Learning Algorithms, International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-8 Issue-4, April 2019

[12] Sheikh Mohammad Idrees, M. Afshar Alam, and Parul Agarwal; A Prediction Approach for Stock Market Volatility Based on Time Series Data , 2019

[13] Sneh Kalra , Jay Shankar Prasad, Efficacy of News Sentiment for Stock Market Prediction, 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (Com-IT-Con), India, 14th -16th Feb 2019

[14] Mohd Naved ; Technical Analysis of Indian Financial Market with the Help of Technical Indicators , International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

[15] Arize, A., Kallianotis, L. N., Liu, S., Malindretos, J. & Maruffi, B. L. (2014), „The Preponderance of Stock Picking Techniques: The Practice of Applied Money Managers", Accounting and Finance Research, 3(2): 87.

[16] Witten, I. H., Frank, E., & Hall, M. A. (2011). Data Mining – Practical Machine Learning Tools and Techniques (3rd Ed). Burlington: Morgan Kaufmann, pp. 124-125

[17] Yahya Eru Cakra, Bayu Distiawan Trisedya, Stock Price Prediction using Linear Regression based on Sentiment Analysis, ICACSIS 2015

[18] A. J. Conejo, M. A. Plazas, R. Espinola, and A. B. Molina, ``Day-ahead electricity price forecasting using the wavelet transform and ARIMA models," IEEE Trans. Power Syst., vol. 20, no. 2, pp. 1035,1042, May 2005.

[19] K. K. Suresh and S. R. K. Priya, ``Forecasting sugarcane yield of Tamilnadu using ARIMA models," Sugar Tech, vol. 13, no. 1, pp. 2326, 2011.

[20] Guangyu Ding1 · Liangxi Qin1 ; Study on the prediction of stock price based on the associated network model of LSTM, International Journal of Machine Learning and Cybernetics

[21] Twitter Sentiment Classification Using Machine Learning Techniques for Stock Markets Mohammed Qasem, RuppaThulasiram, ParimalaThulasiram Department of Computer Science University of Manitoba Winnipeg, Canada.

[22] Vivek Rajput1, Sarika Bobde , STOCK MARKET FORECASTING TECHNIQUES: LITERATURE SURVEY, International Journal of Computer Science and Mobile Computing (2016)