

# PROGNOSIS OF INDIAN STOCK PRICE THROUGH MACHINE LEARNING ALGORITHM AND SENTIMENT ANALYSIS

Harshwardhan Patil and Rahul Patil

Department of Computer Engineering,  
Pimpri Chinchwad College of Engineering, Nigdi, Pune, India

## **ABSTRACT**

*Unpredictable stock price forecasting is a difficult task due to the markets' flexible and unconditionally volatile nature. views into the machine learning meadow with the impending emotive and quantitative strategy. Increasing computational capabilities, software-based statistical medium of prognosis, and inventive method of prognosticating the model are all combined. In this study, the next due day-closing prices of Indian equities SBIN and Tatamotors are used to compare the long short term memory, Random Forest, and linear regression algorithms. Utilizing the RMSE standard layout indication, the prototypes are assessed. The low value of this indicator results in the most accurate closing price prediction model when compared to others.*

## **KEYWORDS**

*Random Forest Regression; Long-Short Term Memory; Stock Price Prognosis.*

## **1. INTRODUCTION**

The character of the National Stock Exchange may be distinguished as unpredictable, fluid, and non-linear. However, the two main approaches have been proposed for predicting the stock[1] momentum of [1]particular organization. Technical analysis uses[1] historcial price of[1] the stocks like lables such as close and open price of the stock, volume traded, adjacent close[1] value of the particular stock for predicting the future price of the stock.[1] The second type is qualitative analysis, this analysis method is permormed on the basis of external factors[1] such as portfolio of company, current situation trending in market, economical and political circumstances, text based information in the form of stock based and generalized news and social media articiles even blogs by economist , CA , financial advisors, researchers. Modern clever tactics based on either technical or fundamental analysis are [1] utilised to forecast the movement of stock prices in these days of technical trends. In particular, the data amount is enormous and fictitious for stock price forecasting. An effective prototype that can spot hazy patterns and complex linkages in a vast data collection is required to handle varied data. In comparison to the prior approaches, machine learning techniques have shown to increase competency by 55–87% in this field. The majority of tasks in this sector make use of algorithms including linear regression, Random Forest(RF)[6], MAC/ MACD, as well as certain models for anticipating NSE goods prices, such as Autoregressive Moving Average(ARMA)[4] and Autoregressive Integrated Moving Average(ARIMA).[4] Recent studies have shown that the application of machine learning techniques[5] like Support vector machines and Random forests may improve stock price predictions (10). Some neural network-based techniques, such as Artificial Neural

Networks (ANN)[1], Convolutional Neural Networks (CNN)[1], Recurrent Neural Networks[1] (RNN), and Deep Neural Networks (DNN) as Long Short Term Memory (LSTM), have also produced encouraging results[1]. Moreover, sentiment prognosis prototypes records from sentiment dimensions. The sentiment of news articles have been studied in finance sphere. Neiderhoffer (2nd pap-27) analyses NYT headlines over the past 20 years and divides them into 19 predefined semantic categories, ranging from negative to extremely positive. He also examines how exchanges respond to various sets of events and discovers that the market has a propensity to overreact to negative news. The impact of optimistic and pessimistic language used in news on future performance of businesses is examined by Davis et al. (2nd Pap). They suggested two things: (1) there is a conflict between readers' expectations and writers' intentions; and (2) readers significantly respond to both the emotional and content aspects of reports that deviate from their expectations. According to research, investors first forecast event articles and convert them into market feelings. Investors then base their decisions on the market sentiment interpretations. Finally, market prices aggregate all of the investor activities and represent them in the ultimate price movements. As a result, it could be harder to incorporate sentiment prognosis into the forecasting blend. With this the quantitative strategy like LSTM is capable for finding through long term dependability in data. This is gained through module of recurring model which has a blending of 4 layered interacting with each other. RF regression a supervised machine learning algorithm uses ensemble method of learning in regression has also used by researchers' for forecasting the stock price. Linear regression another supervised machine learning algorithm is also used widely for forecasting the stock prices. In this work, we have divided our experimentation in two methodologies on broad level of predicting the stock price momentum. The first method is prognosis of day end price of National Stock Exchange (NSE) listed stock price by LSTM, Random Forest and linear regression. The model uses group of new variable construct by the live National Stock Exchange(NSE) dataset with Low, Close, Open and High of selected NSE listed firm. In terms of the models' improvised forecast, which is assessed by the RMSE performance metric, the models trained using these three techniques will play a key role. The influence of sentiment that has been identified by news flashing in the market and categorised as good, negative, or neutral is being examined as the second technique in this paper.

The following is a full summary of this: Section 2 describes the methodology used. The experimentation is covered in Section 3, and the task experimentation is covered in Section 4.

## **2. METHODOLOGY**

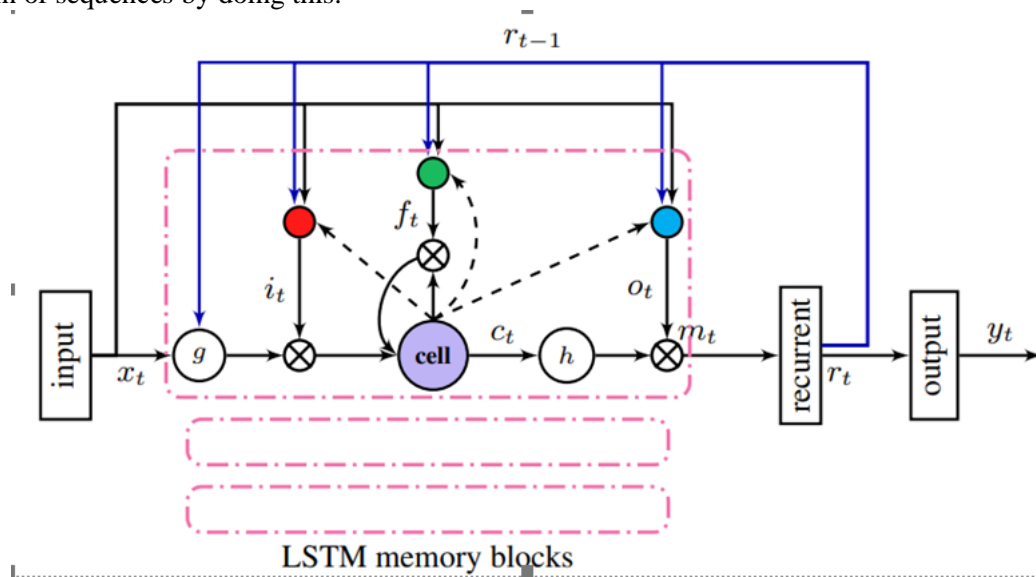
### **2.1. Explanation of Information**

Information data for two firm has gathered from YFG Finance. The dataset consists of 46 days' worth of quantitative data for Tata Motors and State Bank of India from 19 May 2022 to 21 July 2022. The information includes exchange-related price points such as Low, Close, High, Open, AdjacentClose, and Volume. For predicting, the price of the exchange at day's end, taken in 24-hour intervals, has been taken into consideration. The following seven-day closing price is predicted.

### **2.2. LSTM**

Improvised factor after RNN is LongShort-Term Memory. RNN mimickize the learning like humans. As human, start our thinking from scratch each second. LSTMs are a type of RNN that remember information over long periods of time. This mentioned behaviour of LSTM make them better suited for predicting stock prices. The internal gates of LSTM have the ability to control the information flow. The gates are capable of learning which data in a sequence should be kept

or ignored. In order to create predictions, it can convey pertinent information down the extensive chain of sequences by doing this.



### 2.3. Random Forest

RF that is Random Forest it is machine ensemble learnable strategy. It is bearable of computizing regressable & classifiable work. Instead of depending solely on one decision tree, it is proposed that many decision trees be combined in order to identify the desired conclusion, hence reducing model variation. All decision trees that show the choice made at the nodes of the tree may be trained using newly produced variables. The lags in exchange point data usually is presize high reasonably of its large spaced & reasoned that trees to evolve in a finished variant manner comparable to expectable extent . It focus at lowering forecasting error

### 2.4. The Linear Regresson

Machine Learning's supervised another type is linear regression algorithm use in work .LR carries task to prognose a variable that depends variable indicator (y) relave on given free variable x.Regression strategy of this kind explores a L-relation between x,y In this experiment x test datagroup & y prognostable day end price for next seven days.

### 2.5. Sentiment Analysis

In this work for predicting the sentiment intensity of the stock related news we have used sentiment intensity analyzer method from Vader sentiment. To fetcch the news and article related to the stock we made use of Google News and Newspaper package of python. The news articles are current day and previous day mentioned as now and yesterday.

The news is categorized in keywords to intensify positive, negative and neutral sentiments. Positive sentiement will increase if negative count is less than positive count on the other side Negative sentiment increase by +1 if the positive count in news is less than negative count. In the condition where positive count is equal to the negative count then it will be categorized to neutral sentiment.

We implement module by sentiment PO[polarity] , indicator for distinguish the article range of sentiment

gnt = positive when  $(f(p)-f(n))/f(n) \geq th$   
 gnt = negative when  $(f(n)-f(p))/f(p) \geq th$

### 3. CONCLUDING AND EXPERIMENTED RESULT

To digout effect trained systems, comparision is done with RMSE performance score of LSTM, Random forest and linear regression. With the data date starting from 19-05-2022 till 21-07-2022. The stock like SBI, Tata Motors and Cipla are choosen for experimentation purpose. Prognosted next end day prices permormed through (RMSE). The RMSE equat by below formation.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (O_i - F_i)^2}{n}}$$

here ‘The  $O_i$ ’ point to the original day end price, ‘ $f_i$ ’ point to the prognosted day-end price ,n point overall observation space.

Price action chart below shows the trend in SBI stock price for last 46 days. The trend line for LSTM , Linear regression and Random forest shows the positive momentum in SBI . Moreover calculating the RMSE results in choosing best predictive model among LSTM, Random Forest and Linear Regression.

#### 3.1. State Bank of India Prognosis

The figure 2 shows the forecasting of LSTM against SBI stock price from 19 May 2022 to 21 July 2022.

The figure 3 shows the forecasting of Random Forest against SBI stock price from 19 May 2022 to 21 July 2022.

The figure 3 shows the forecasting of Linear Regression against SBI stock price from 19 May 2022 to 21 July 2022.

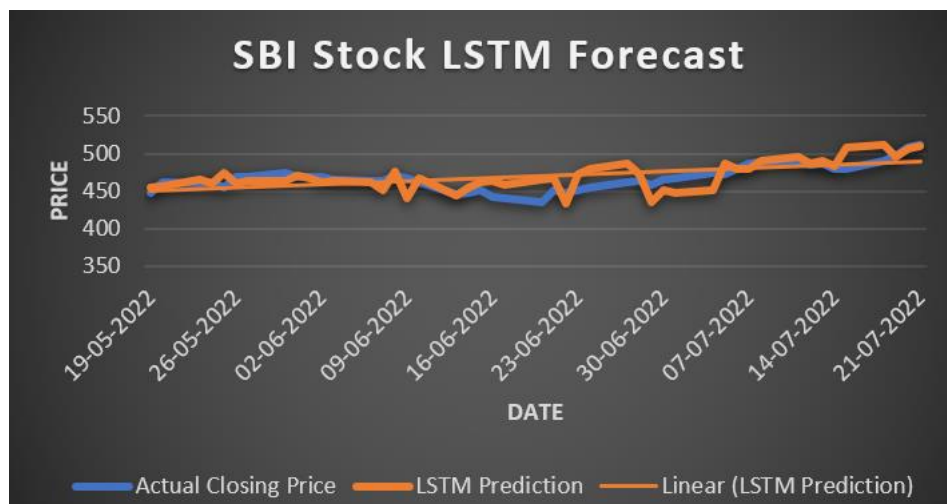


Figure 2. SBI Actual Price Vs LSTM Forecast Price

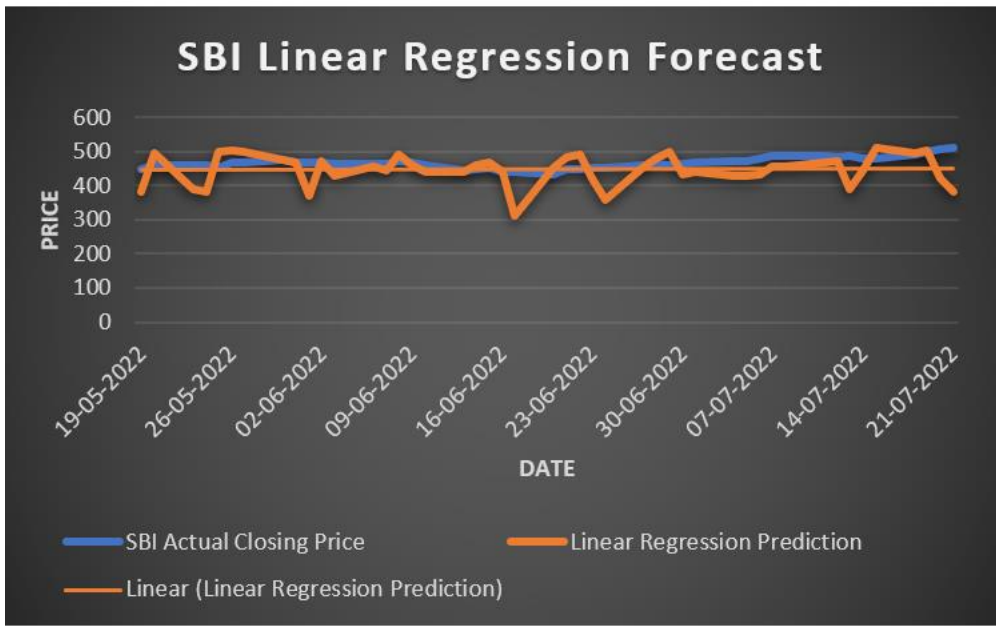


Figure 2. SBI Actual Price Vs Linear Regression Forecast Price

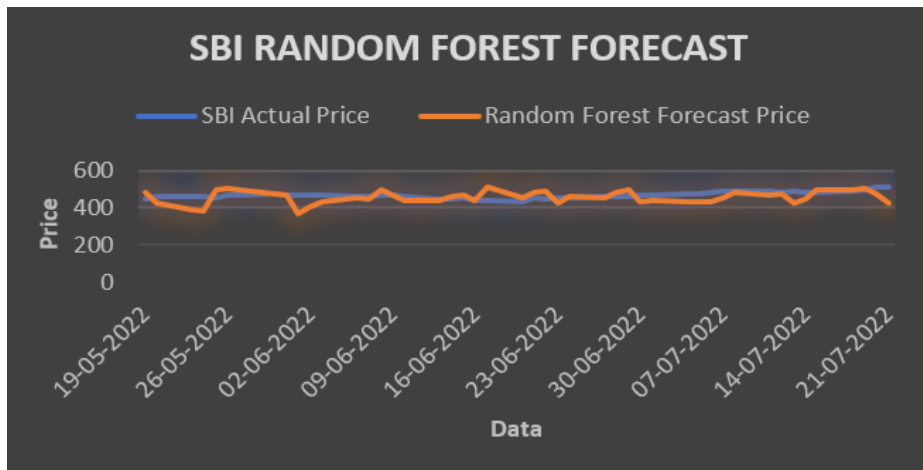


Figure 3. SBI Actual Price Vs Random Forest Forecast Price

Table 1. Root Mean Squared Error of SBI Stock for 46 days.

Algorithm	Root Mean Squared Error
Long Short Term Memory	14.2
Linear Regression	36.5
Random Forest	37.2

The sentiment analysis shows the positive sentiments are 85% whereas negative sentiments has coverage of 10% and Neutral sentiment has coverage of 5%.

Stock related news article believed to have the impact on stock related news article with the momentum of trend. Here SBI stock 46 days average polarity for positive sentiment is 85%. That has witness the positive trend. Similarly tatamotors polarity in figure 6

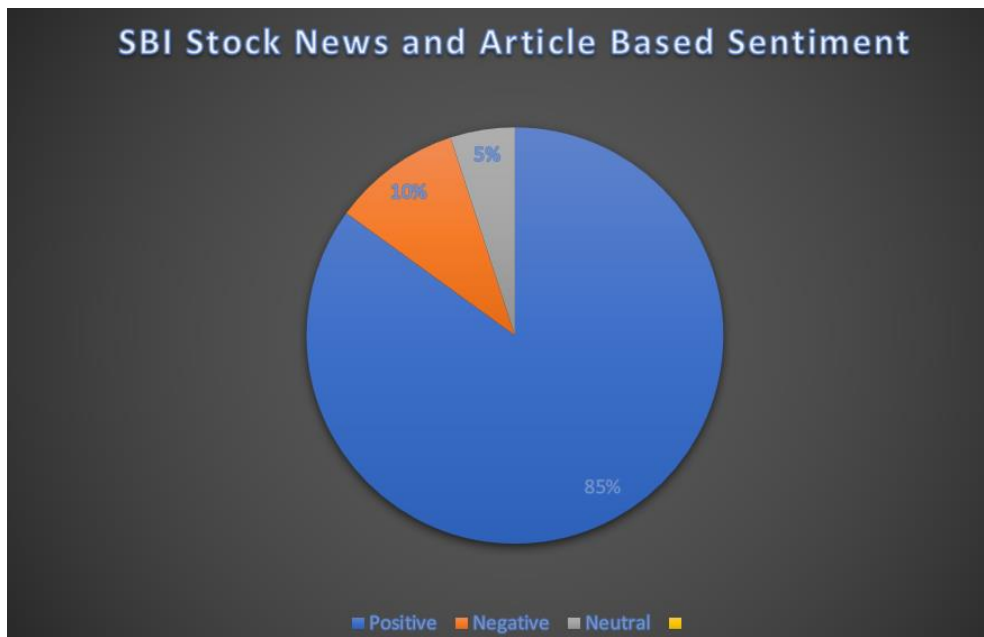


Figure 5. Sentiment Analysis of SBI News

### 3.2. Tata Motors Prognosis

Figure 6 shows the forecasting of LSTM against SBI stock price from 19 May 2022 to 21 July 2022.

Figure 7 shows the forecasting of Linear Regression against SBI stock price from 19 May 2022 to 21 July 2022.

Figure 8 shows the forecasting of Random Forest against SBI stock price from 19 May 2022 to 21 July 2022.

Comparative trend in tata motors has predicted the momentum in the same direction as per actual price. Research proved that for all three algorithms Price is exactly accurate on each day of the market. There observed the variation in the actual and predicted prices. But the significant conclusion here can be stated that the trend momentum of stock price can be predicted i.e in upcoming days price action will be at up-trend or down-trend. Tata motors here shows the positive momentum in futuristic cycle.

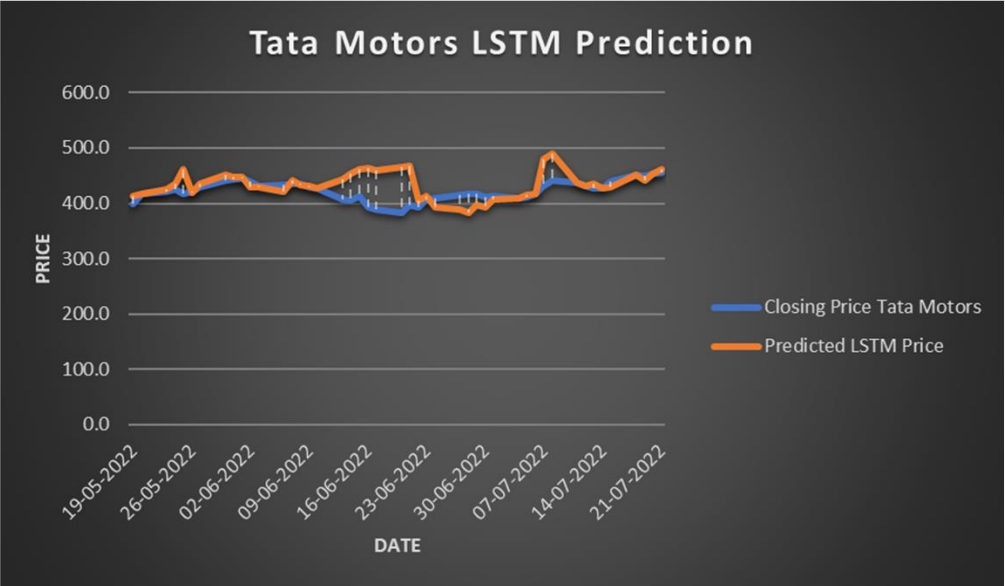


Figure 5. Tata Motors Actual Price vs LSTM forecasted price

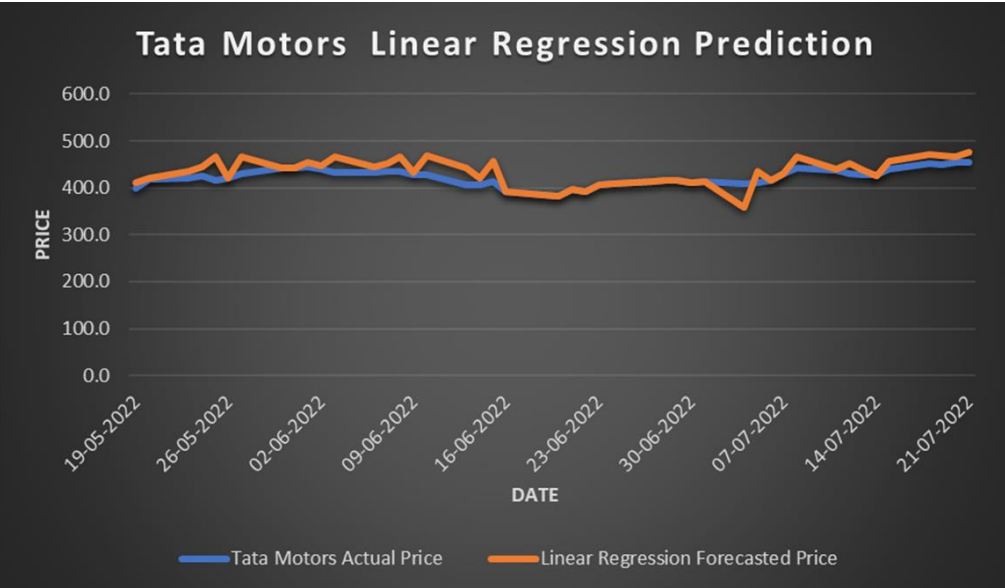


Figure 6. Tata Motors Actual Price vs Linear Regression forecasted price

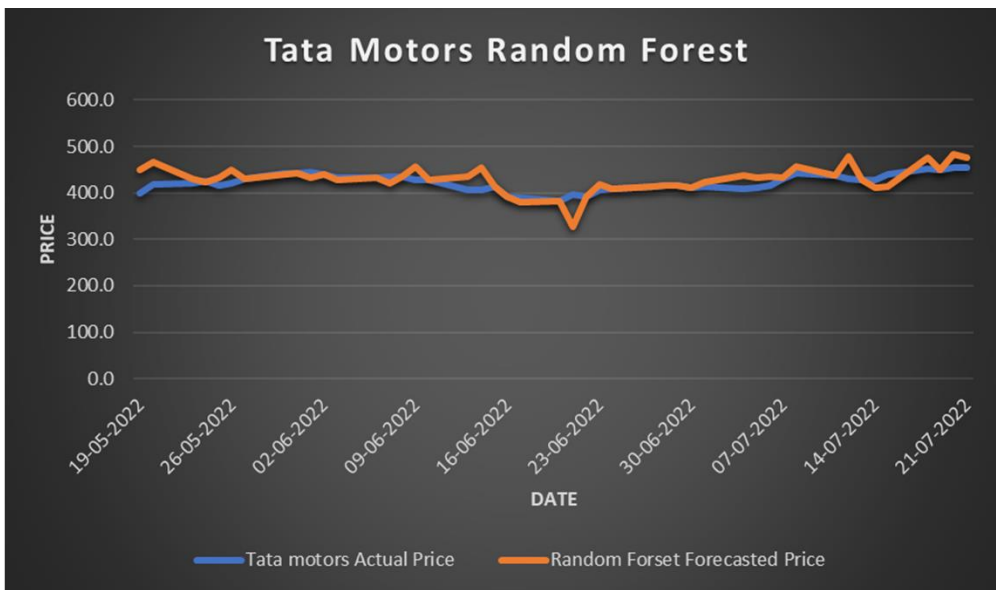


Figure 7. Tata Motors Actual Price vs Linear Regression forecasted price

Table 2. Root Mean Squared Error Tata Motors Stock for 46 days.

Algorithm	Root Mean Squared Error
Long Short Term Memory	11.94
Linear Regression	19.94
Random Forest	22.44

Table 2 shows the root mean squared error value for Tata Motors stock of LSTM, Random Forest and Linear Regression. From table two the result shows tata motors prediction price performance through RMSE value Long-Short Term memory has least root mean squared error with 11.94 whereas it is followed by Linear regression with 19.94 and Linear regression is followed by Random forest with 22.44 RMSE respectively. The 46 days average of tata motors sentiment analysis results the positive sentiments as 78% whereas negative sentiments has coverage of 18% and Neutral sentiment has coverage of 4%. Based on 96 news article in past 46 days the polarity score has been generated. The table below informs the total number of news article that has used for polarity prediction.



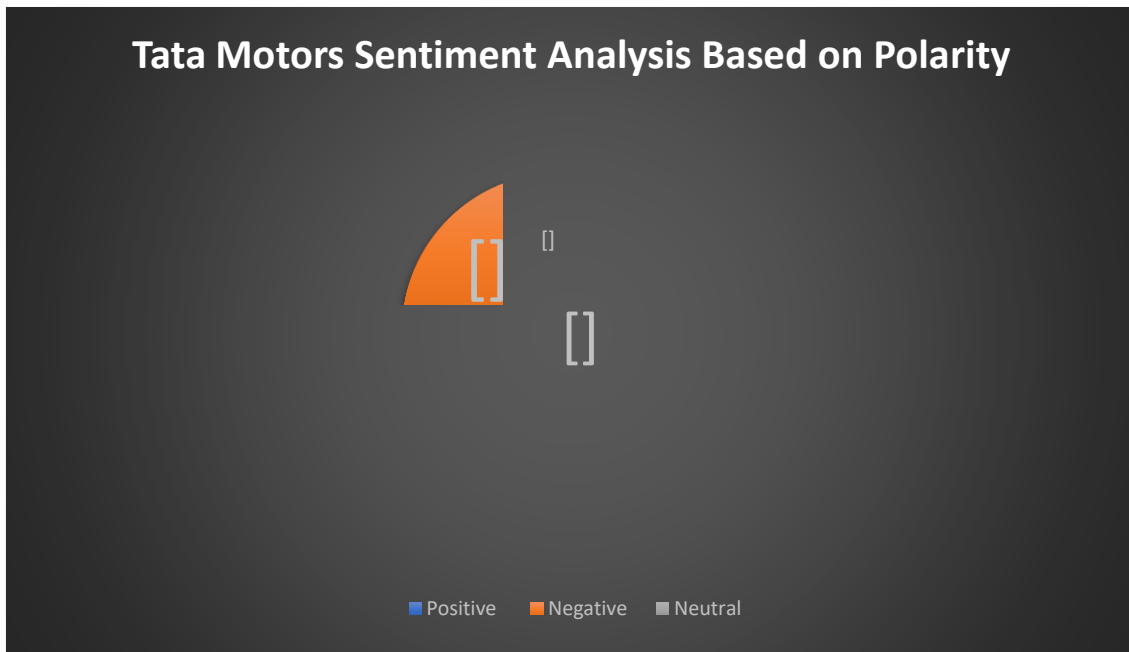


Figure 8. Sentiment Analysis of SBI News

Table 4. Comparative analysis of Tata Motors and SBI RMSE values obtained using LSTM, Random Forest and Linear Regression

Company	LSTM	Linear Regression	Random Forest
Tata Motors	11.94	19.94	22.44
SBI	14.2	36.5	37.2

The values above T4. depict, for SBI & Tatamotors company, LSTM results to be finer technique, giving least rmseRootMeanSquaredError, as shown in the Table 3.

The result of sentiment analysis about SBI and tatamotors predict the positive polarity about both the stock with positive polarity factor of SBI as 85% and that of tatamotors is 78%. This shows bull side will be stronger for these stocks in upcoming days.

#### 4. CONCLUSIONS

Forecasting exchange price outcomes a stimulating work anticipated through routinely fluctuating Indian National exchange price that depend on multiple facts which form compound patterns. Results show that the best values obtained by LSTM model gives RMSE (11.94) for past 46 days of tata mototrs and 14.2(RMSE) for SBI.

The result of sentiment analysis about SBI and tatamotors predict the positive polarity about both the stock with positive polarity factor of SBI as 85% and that of tatamotors is 78%. This shows bull side will be stronger for these stocks in upcoming days

**REFERENCES**

- [1] Mehar Vijn, Deeksha Chandola, Vinay Anand Tikkiwal, Arun Kumar. "Stock Closing Price Prediction using Machine Learning Techniques", Procedia Computer Science, 2020
- [2] <https://www.mdpi.com/ Stock Market>
- [3] [pure.iiasa.ac.at](http://pure.iiasa.ac.at)
- [4] <https://www.science.gov/>
- [5] Stock Market Forecasting using Machine Learning: Today and Tomorrow 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICT)
- [6] Indian stock market prediction using artificial neural networks on tick data. DharmarajaSelvamuthu, etl <https://jfinswufe.springeropen.com/articles/10.1186/s40854-019-0131-7>
- [7] Stock Prediction with Random and Long Short-term Memory. Shangxua
- [8] Research on the text sentiment classification about the social hot events on Weibo Fulian Yin etl
- [9] Analysis of various machine learning algorithm and hybrid model for stock market prediction using python. Sahil Vazirani; Abhishek Sharma; Pavika Sharma
- [10] Predicting Stock Movement Using Sentiment Analysis of Twitter Feed with Neural Networks. Sai Vikram Kolasani1
- [11] Stock Market Analysis using Supervised Machine Learning. Kunal Pahwa; Neha Agarwal
- [12] Word Embeddings and Their Use In Sentence Classification Tasks
- [13] Adi Shalev etl Stock Market Forecasting using Machine Learning: Today and Tomorrow
- [14] Sukhman Singh; Tarun KumarMadan; Jitendra Kumar; Ashutosh Kumar Singh

**AUTHORS**

**Harshwardhan Patil**, MTech Student at Department Of Computer Engineering, Pimpri Chinchwad College Of Engineering For Year 2021-22.



**Prof Rahul Patil**, Professor at Department Of Computer Engineering, Pimpri Chinchwad College of Engineering

