

# Prediction Of Market Trends

Prof.Gajanan Walunjkar

Dr. Rahul Desai

Prof. Yuvraj Gholap

R Sankalp Shukla

Army Institute of Technology, Pune, India

---

Forecasting about Indian market has always been interesting and topic of discussion among analyst and researchers. With the arrival of machine learning and artificial intelligence the race is now becoming the competition with best algorithms to be used and give investors more profit. In past years prediction was only based on experience and daily headlines of business newspapers but now it depends on various international, national and political economic factors and the sentiments and reaction of people over the issues. With the growing power of social media, the game in market over this also changed now with the help of sentiment analysis over social media we can determine the mood of investors over the news. In the present scenarios you can divide two categories for the prediction strategies one is the time series analysis of stocks and the second is artificial intelligence property over the market. AI property contains multi-layer perception, SVM, naive Bayes , back propagation,CNN,LSTM, RNN etc during this we have a tendency to come with plan of combination of each. Within the paper we have conjointly covers the assorted challenges that are encountered while building prediction models. This whole module focuses on use of statistical analysis and conjointly development of the sentimental analysis and to get better results. The LSTM has the advantage of analyzing relationship between time series knowledge through memory functions. The performance of the system is improved by combine efforts of time-series and sentiments with the LSTM prediction model.

Keywords: lstm, time-series , recurrent neural network.

---

## 1. INTRODUCTION

Stock market has provided a platform for investors to deploy their resources in Indian industries in exchange of their equities. Indian equity market development got pace in the initial phase of liberalization. Volatility is one of the characteristic factors in stock market which makes it speculative and provided a room for predicting its future trend. And volatility itself is influenced by trading volume and new information or sentiments regarding that stock. volatility is a key parameter used in many financial applications so, it is important to estimate volatility in stock market. The most important role of any investor in stock market is to analyze its movement and to make an precise prediction to have profit on his investment. The amount of people which invest over stocks increased, especially after the start of bull market and loss of jobs in covid 19. Even the single investors are important part to the development of Indian markets. With the ease of investing through brokers apps and good internet facilities both the things have changed unexpectedly investing and expressing thoughts which affects the stocks. Therefore, we are using digital data and news feed (as they are constant source) of any major indexes like NIFTY50, SENSEX, etc. to construct the outcome model and test the co-relation between media attention(sentiments) and trading volume in predicting their trends. Some important definition for market understanding are as follows:

- (1) **Buy** : To take a position by buying shares of a company.
  - (2) **Sell** : To sell the shares you currently own.
  - (3) **Bid** : When a trader in the market makes an order to buy shares.
  - (4) **Ask** : When a trader orders their shares for sale at a certain price.
  - (5) **Bull Market** : A market condition where stock prices are continually rising.
-

- (6) **Bear Market :** Bear markets occur when the prognosis for a firm, an industry, or the economy as a whole appears to be bleak. Traders and investors are reluctant to purchase equities, and many are looking to sell their holdings. As a result, prices drop.

The interaction impact of media attention and trading volume on stock return is considerable, implying that media attention's influence on stock return grows as trading volume climbs.

## 2. BACKGROUND AND RELATED WORK

- **Statistical Method :** Using various statistical approach, a way to forecast the securities market price is to utilize fbprophet. may be a python library typically want to forecast seasonality, trend, weekly, and yearly. This method is incredibly useful to forecast its behavior. Exponential Smoothing is usually used for time-series analysis. This method smoothest out the historical graph data to raised visualize the long run stock price. Mean Squared Error (MSE) could be a measure to indicate what number points are near the road of best fit. Furthermore, this is able to show how far apart are the information points on the scatter plot. Deploying the above- mentioned statistical tools to the dataset would assist this research in analysis and forecast stock prices. This task will include extracting historical dataset also as studying about the info at timed intervals in years, weekly, seasonality and trend behavior.
- **Sentimental Method :** What is Natural Language Processing (NLP)? You should use NLP while creating applications that are related to: Generate Text, Speech Recognition, or Text Recognition! Essentially, NLP is the ability to analyse words, but not in the "Human Form." In NN Approach, we'll go over this extensively. Assume you're creating a model to find news terms. Before you start making the model, there are a few measures you need take. To begin, you'll want to enter the data into your model once. However, since we're talking about NLP and Neural Networks, we'd like to adapt this data for our NN. We want to recast this word in numbers to enter it into our NN. What's more, why do we wish to alter it? Instead of merely entering the words into the NN, why not convert them to numbers? To answer this question, first take a look at the image below:

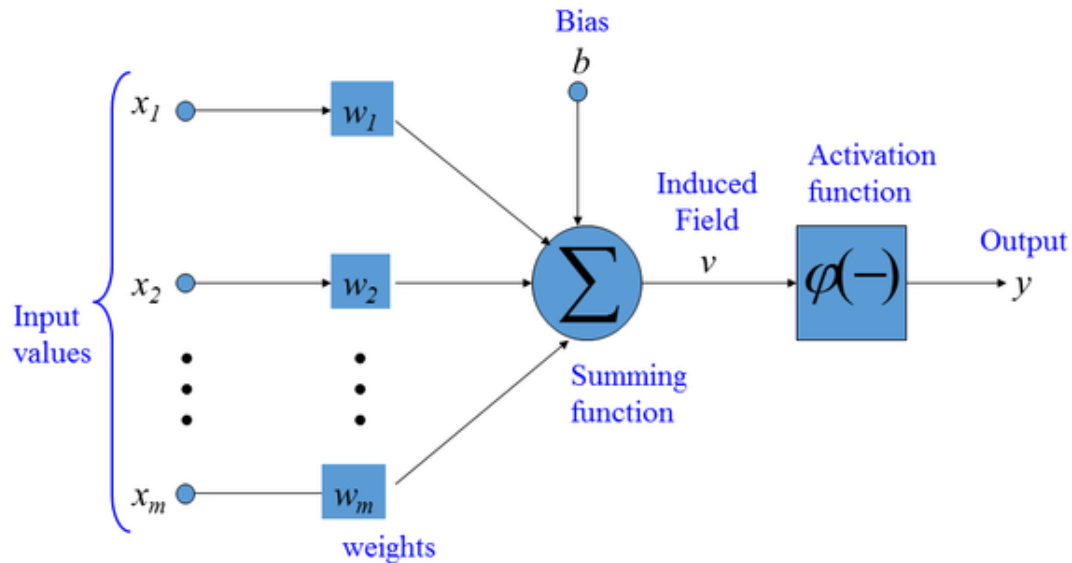


Figure 1. An illustration of an artificial neuron receives n inputs

As you can see, the magic happens in any NN's structure because of the math. Numbers are the universal language of Artificial Intelligence, regardless of domain. Many algorithms, such

as Reinforcement Learning or the inverse, will continue to reign supreme. We are prepared to utilise a method known as tokenization to convert words into integers. Feed Forward NN's don't produce clever outcomes when you're working with consecutive data, where the preceding data is important. In a single word, the decision to use RNNs will be justified: Memory When working with consecutive data, the RNN's unit is capable of storing information from prior phases, which benefits us greatly.

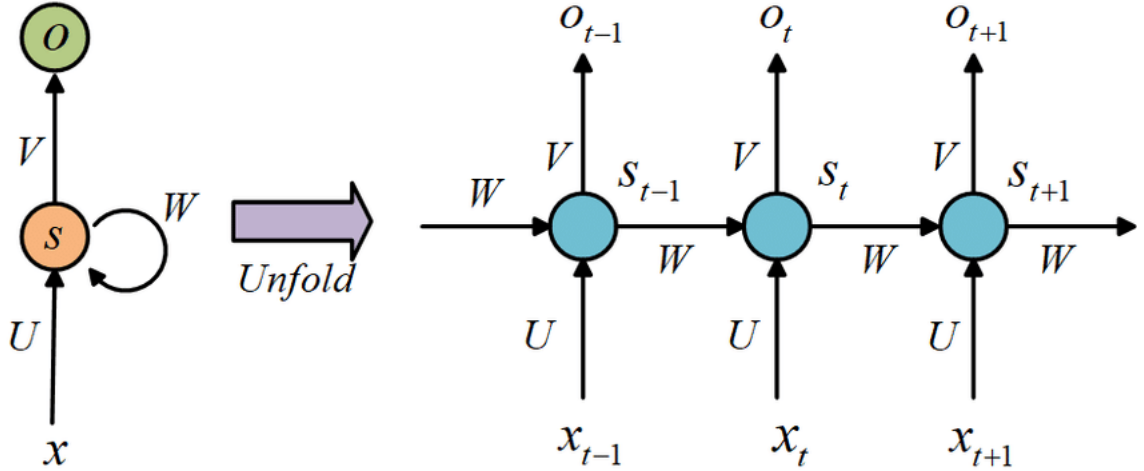


Figure 2. Basic-RNN-structure

2.0.1 *Comparison between Models* ∴ Accuracy of various classifier area unit taken as a basis for recommending the algorithm. An analysis was created on a news data set to spot pretend news accurately. The precision, accuracy and f1 score are often computed with the assistance of confusion matrices. one confusion matrix was created for every model. The values shown area unit the averaged values over ordered trial.

Model	Accuracy
Naïve Bayes	71.84%
Support Vector Machine	87.37%
LSTM	94.27%
Keras Based Neural Network	90.62%

Figure 2. Comparisons between classifiers

### 3. MODEL DESIGN

The way we are approaching toward the problem it reduces the chance of error because we are covering both the aspects of the share market

- (1) Statistical Method
- (2) Sentimental Method

These both methods are the most appropriate combinations to tell you about the real predictions.

### 3.1 Statistical Method

Most of the existing machine learning model (ML) input as samples/examples however there is no time involved in the context. The time-series method made up models can predict future pricing or determined value. Time-series are always used to predict for non-stationary knowledge these are like mean and variance are n't constant over time however instead, these metrics vary over time. And these autos generated computer files (input for model) are known as time-series data. Some useful examples of it are temperature values over the crop and stock price of industry associated, worth of house over time period, price of gold with time series etc. Input can be an information that is observed with reference to-time. Time series analyzers record data points at regular intervals over a predetermined period of time, rather than intermittently or arbitrarily. This type of research includes more than just collecting data over time. Time series data is distinguished from other types of data by its ability to represent how variables change over time. It includes a data source as well as a list of data dependencies in a predetermined sequence. Time series analysis frequently necessitates a large number of data points in order to maintain consistency and dependability. A large data set assures a representative sample size and that your research can cut through noisy data. Time series data can also be used to forecast or predict future data based on existing data.

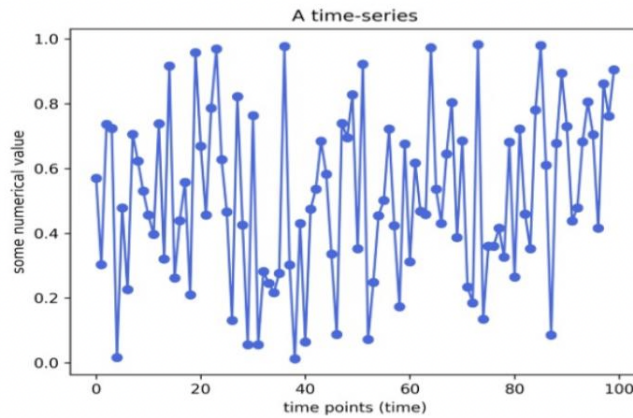


Figure 3. Time series analysis

### 3.2 Sentimental Method

**3.2.1 *Sentiment and Market*:** Social media networks like Twitter, Instagram and Facebook etc. are growing their scope day by day. It is used by each and every category of people regardless of caste, sex, gender to put their opinions over any topic don't matter whether its international, national or political or launch of business products. This information is very much useful to know the sentiment of people and it can be used to make policies, planning's and processes for the development of product or company. Here comes the role of sentiment analysis which is used to extract such opinions and classify them into majorly three types which are as follows – positive, negative and neutral. Although we can define it in many ways it is basically capturing of perspective or views of individual or organization from any platform available (for now only web). There are some serious challenges also associated with it because human are not very straightforward because machines must be taught to evaluate and comprehend emotions in the same way that the human brain does. This is in addition to comprehending the subtleties of other languages. As data science advances, sentiment analysis tools will be better able to address these difficulties. Words like "love" and "hate" have strong positive (+1) and negative (-1) polarity

scores. These are easy to understand. However, in-between conjugations of words like "not so bad" may convey "average" and therefore fall near the middle of the polarity spectrum (-75). The sentiment score is diminished when statements like these are omitted. People use irony and sarcasm in informal social media talks and memes. The use of backhanded compliments to signify negative feeling may make it difficult for sentiment analysis algorithms to understand the true meaning of the remark. As a result, there may be an increase in the number of "positive" but actually unfavourable comments. These are the issues that must be addressed. As in above sections you can see that lstm technique is best to use to all this .

**3.2.2 LSTM architecture :** This is an RNN variant that is extremely useful when you need your network to memorise data for a lengthy period of time. The LSTM relies on gates with functions that determine which data should be passed to the next cell, as well as a few other configurations.

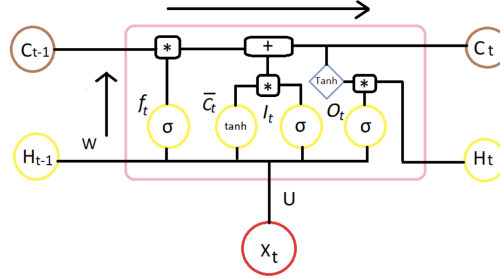


Figure 3. Long Short Term Memory Cell

In Long transient memory the segment are as per the following Forget Gate "f", Candidate layer "C", Input Gate "I", Output Gate "O", Hidden state "H" and Memory state "C". As the contribution to lstm are "X" (present state),"H" (last secret state), C (last memory state) and the yield from it are "H" (current secret state) and C (Current memory express) The outline for this is as per the following at T time step

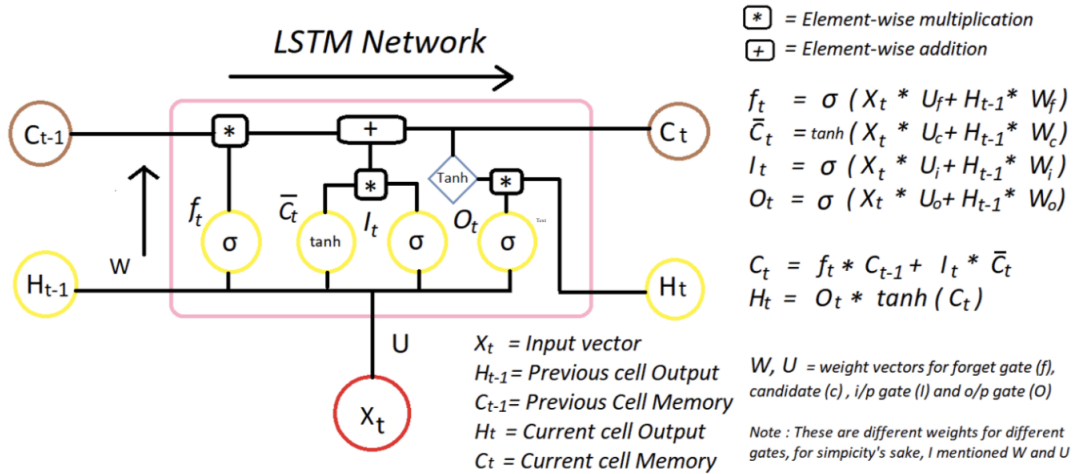


Figure 3. LSTM cell at T time step.

Input and output gates are one-layer neural networks with sigmoid activation functions that are not included in the candidate layer (the input in it is Tanh for activation). Gates here take "U","W" as input, concatenate it, and then apply an activation function, resulting in vectors (valued between 0 and 1 for Sigmoid -1 to for Tanh), so now we have four vectors for each step "F","C","I","O," and also a portion of memory state C with each and every time step. For the sake of illustration, we merely save the (context) or previous state in this state. Payal enters a room, and Shikha enters as well, greeting Payal. Who did Shikha say "hello" to? In order to foresee this, We save "Payal" in memory C since it should be stored somewhere. If you like, you can modify this. We can add and remove information in LSTM cells according to the latest memory state  $C_{t-1}$ , and partwise multiply using the forget gate "F." If the forget gate price is 0 and  $C_t = C_{t-1} * F_t$ , the previous memory state is completely forgotten; if the forget gate price is 1, the previous memory state is completely given to the cell (b ear in mind f gate offers values between zero and one). We now have a path for calculating new memory states from input state and C layer for the current memory state  $C_t$ .  $(I_t * C'_t) = C_t$   $C_t$  denotes the current memory state at the time stamp t to next time step. Here is multidimensional language for  $C_t$ . In last we would like to do calculation aiming for output. This output are going to be supported by Our cell state  $C_t$  however these are going to be filtered version due to this we would force Tanh on  $C_t$  then we proceed for part wise multiple of the output gate O , that can be our current hidden state  $H_t = \text{Tanh}(C_t)$  We then pass these  $C_t$  and  $H_t$  to subsequent time step and repeat an equivalent.

#### 4. WORKING

In on top of section the matter of memory and additionally of fake news is introduced and solved by lstm design and additionally however the design of lstm works is mentioned during this section we have a will see to overcome all while creating model and the way this model after this, we have to process the data and make it suitable to use for that we simply need to do tokenization. The tokenization method divides a sentence word for word, as shown in the graphic below.:



Figure 3. Tokenisation

After the tokenization we need to eliminate anomalies for example acquiring dispose of phenomenonally long or short audits. To damage every little and extremely long audit, we'll do cushion or shortening in the entirety of our surveys to a proper chose length. we will in general blueprint this length by Sequence Length. This arrangement length is same as scope of your time ventures for LSTM layer. For surveys more limited than" length", we'll cushion with 0s. For surveys longer than" length" we'll shorten them to the essential" length" words. Whenever we have our informational index in pleasant structure, we will part it into preparing, approval and in test sets. Here we characterize the method of LSTM working with the picture underneath

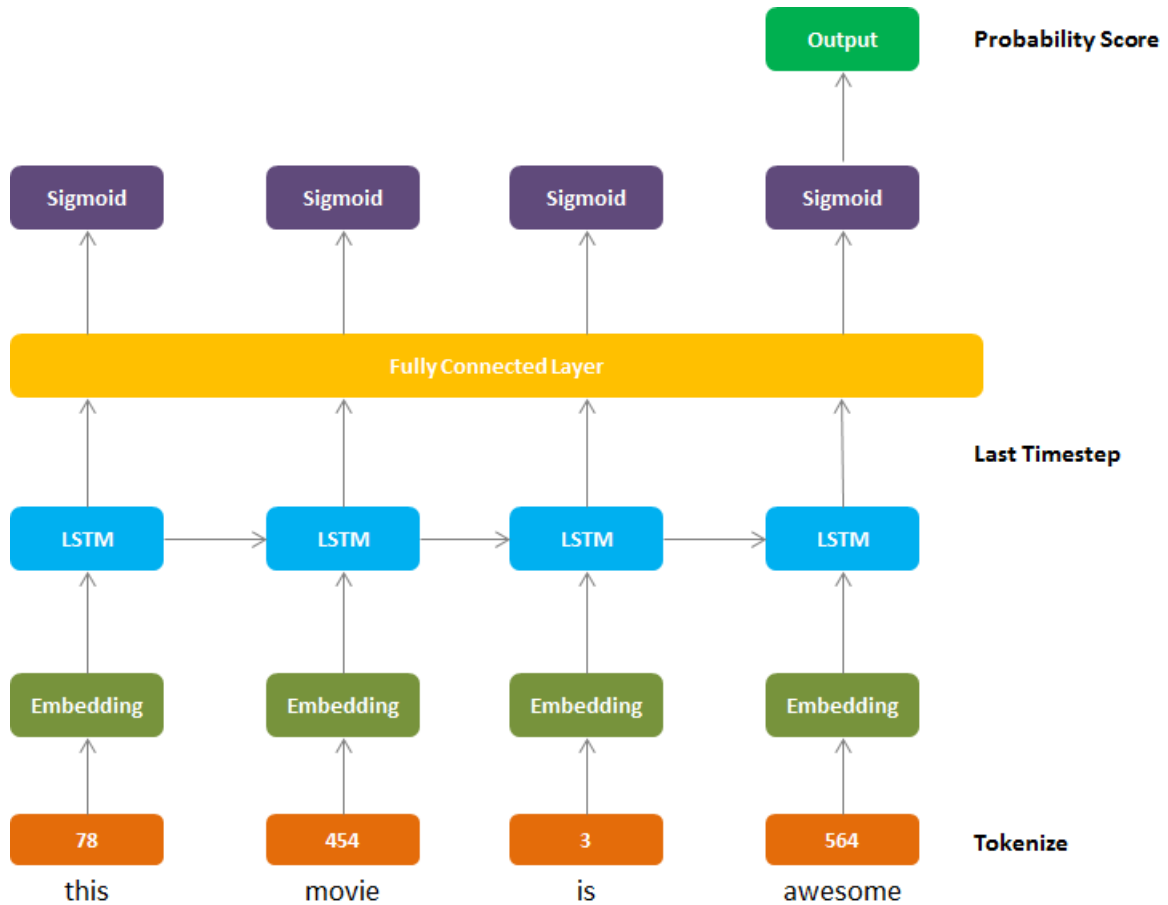


Figure 3. LSTM Architecture.

## 5. CONCLUSION AND FUTURE WORK

As a future work, we will aim to consider some more features of stock market that we can integrate with our current setup to increase the efficiency of the system in making short- term trend prediction. We will also aim to extend the same system for some individual stocks with high daily volume, instead of any index.

## References

- Tijare, Poonam. (2019). A Study on Fake News Detection Using Naïve Bayes, SVM, Neural Networks and LSTM.
- Ding, G., Qin, L. (2020). Study on the prediction of stock price based on the associated network model of LSTM. International Journal of Machine Learning and Cybernetics, 11(6), 1307-1317.
- Pulver, A., Lyu, S. (2017, May). LSTM with working memory. In 2017 International Joint Conference on Neural Networks (IJCNN) (pp. 845-851). IEEE..
- Mohan, S., Mullapudi, S., Sammeta, S., Vijayvergia, P., Anastasiu, D.C. (2019, April). Stock price prediction using news sentiment analysis. In 2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService) (pp. 205-208).
- Khatri, S. K., Srivastava, A. (2016, September). Using sentimental analysis in prediction of stock market investment. In 2016 5th International Conference on Reliability, Infocom Technologies and Optimisation (Trends and Future Directions)(ICRITO) (pp. 566-569). IEEE
- Reddy, N. N., Naresh, E. (2020, July). Predicting Stock Price Using Sentimental Analysis Through Twitter Data In 2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT) (pp. 1-5). IEEE

**Prof Gajanan Walunjkar** Prof. Walunjkar is an Asst. Professor at Army Institute Of Technology, Pune (India).He has wide range of knowledge and experience of 18 years in teaching with major in ad hoc network and ML . Also contribute in the field of software design and modelling .



**Dr. Rahul Desai** Dr. Desai is Asst. Professor at Army Institute Of Technology, Pune (India) earlier to this he studies from Savitribai Phule Pune University and done M Tech in computer and Ph.d. in Electronics and communication besides being a Professor at the Dept. of Information Technology and 21 years experience in teaching he is also doctorate in philosophy. Prior to joining AIT, he worked in a Software Developer and Training Division. With wide area of expertise Dr.Desai holds ML and ubiquitous computing tightly in hand.



**Prof Yuvraj Gholap** Prof. Gholap is an Asst. Professor at Army Institute Of Technology, Pune (India).He has wide range of knowledge and experience of 16 years in teaching with major in Data Base Managment Systme and ML .



**R.Sankalp Shukla** Currently pursuing Bachelor of Engineering in Information and technology from Army institute of Technology .

