



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

# IJRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 10, Issue 8, August 2021

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

**Impact Factor: 7.569**

9940 572 462

6381 907 438

[ijrset@gmail.com](mailto:ijrset@gmail.com)

[www.ijrset.com](http://www.ijrset.com)



# Stock Price Prediction Using Machine Learning Techniques

Anusha J Adhikar<sup>1</sup>, Apeksha K Jadhav<sup>2</sup>, Charitha G<sup>3</sup>, Karishma K H<sup>4</sup>, Mrs. Supriya H S<sup>5</sup>

B.E. Student, Department of Computer Science and Engineering, Sir M Visvesvaraya Institute of Technology,  
Bangalore, Karnataka, India<sup>1</sup>

B.E. Student, Department of Computer Science and Engineering, Sir M Visvesvaraya Institute of Technology,  
Bangalore, Karnataka, India<sup>2</sup>

B.E. Student, Department of Computer Science and Engineering, Sir M Visvesvaraya Institute of Technology,  
Bangalore, Karnataka, India<sup>3</sup>

B.E. Student, Department of Computer Science and Engineering, Sir M Visvesvaraya Institute of Technology,  
Bangalore, Karnataka, India<sup>4</sup>

Assistant Professor, Department of Computer Science and Engineering, Sir M Visvesvaraya Institute of Technology,  
Bangalore, Karnataka, India<sup>5</sup>

**ABSTRACT:** Stock markets today, have become an extremely important factor in determining the world's economy. Having said that, the nature of the stock market is extensively unstable owing to the presence of such a dynamic environment. Its very impressive in the fact that it attracts various individuals and financial corporations from all different backgrounds across the world. Its extended complexity is undoubtedly influenced by new age technology be it the search engine queries, news broadcast or even the social networking platforms. The spontaneous response of particular news is very regulatory in determining the decision made by an investor as rightly showcased by behavioral finance. In this paper we have put in a sincere effort towards the prediction of the future of stock prices that predominantly makes use of three machine learning algorithms LSTM along with recurrent neural networks, the linear regression and the ARIMA time series forecasting model. Apart from the already mentioned algorithms we have made a considerable use of Sentiment Analysis that adopts the historical data and the different aspects of Twitter news as its input. The sentimental category of that particular company is analyzed with the help of the filtered tweets. The results from the machine learning algorithms are precisely combined with those resulting from the sentiment analysis to make a pretty accurate forecast about the future prices of the stock and recommend it to the user whether to buy or sell the stock.

**KEYWORDS:** ARIMA, LSTM, RNN, Linear Regression and Twitter Analysis.

## I. INTRODUCTION

Forecasting the future of stock prices has become exceptionally important in numerous fields such as trading exchanges, economic and commercial sectors, business stats and computer science and information technology. The stock market shareholders make use of both the fundamental(primary) as well as technical analysis of the stock prices which helps them make critical decisions on whether or not to invest in that specific stock on the basis of the forecasts made by the model used by them. Widespread research and investigation is being done in the current times on the use of machine learning practices so as to obtain more precise results which consequently helps the stock investors in making spontaneous decisions. The stock prices being extremely vulnerable to very minute changes require developers to work towards building effective algorithms and eventually improvising them to obtain the best results possible. Its true that algorithms cannot completely replace the expert's knowledge and experience but a good accurate algorithmic model could guide us to expecting higher profits. We, to the best of our knowledge, have worked towards developing a model that adopts machine learning algorithms such as the linear algorithm, ARIMA time series forecasting model, the Long-Short-Term Memory and the RNN's. Linear Regression is a modest algorithm that predominantly works on determining the relationship between one dependent variable and another independent variable. Arima model tests all possible combinations of p, d and q where p correlates to the log values, d correlates to the difference count and q



correlates to the moving average also referred to as the order and chooses the best combinations to make the best forecast. LSTM makes use of its recursively recalling memory to store the previously obtained results and in turn obtain the future results. The motivation behind our research in this field is attributed to the ever challenging and emerging technologies.

## II. RELATEDWORK

### 1. "Stock market predication using a linear regression" (20-22 April 2021) Dinesh Bhuriya, Girish Kaushal, Ashish Sharma , UpendraSingh

Forecasting everyday performances of the stock prices has inevitably been a daunting task for the investors and business shareholders but overcoming these challenges will certainly assist them in investing with additional confidence. This paper adopts the linear regression algorithm that is used to predict the performance of the Tata Consulting Services data set. As a consequence this paper concludes that linear regression algorithm performs the best when compared to various other regression techniques out there.

### 2. "Stock Market Prediction Using Linear Regression and SVM 4-5"( March 2021) Bhawna Panwar; Gaurav Dhuriya; Prashant Johri; Sudeept Singh Yadav; NitinGaur

Ever since it came to be known the stock market has not left a single stone unturned in proving the fact that it is the quintessence of all kinds of business and financial aspects and how large an impact its highs and lows can create. The best example for the above said is when the stock market crashed in 2008 leaving behind its disastrous effect. This paper adopts the linear regression as well as the support vector machine algorithm. It begins by the use of web scraping for the purpose of collecting the datasets from the share market, eventually analyzing highs and lows of the stock prices and concluding that the linear regression algorithm performs much better than the support vector machine algorithm in this aspect.

### 3. "Twitter Sentiment Analysis Approaches: A Survey" (August 2020) Omar Y. Adwan ( ) , Marwan Al-Tawil, Ammar M. Huneiti, Rawan A. Shahin, Abeer A. Abu Zayed, Razan H. Al-Dibsi, University of Jordan, Amman, Jordan ,International Journal of Emerging Technologies in Learning(iJET)

Social media has grown enormously over the years. People avail this platform to show their emotions and judgement. One of the most sought-after microblogging and social platform is Twitter where millions of brief notes and messages are posted. Analysis of people's view point, interests is done by implementation of twitter sentiment analysis. This paper provides information on the several approaches and algorithms exerted for twitter sentiment analysis.

### 4. "Deep Learning-Based Stock Price Prediction Using LSTM and Bi-Directional LSTM Model" (24-26 Oct. 2020)Md.ArifIstiakeSunny;MirzaMohdShahriarMaswood;AbdullahG.Alharbi

It's a well known fact that the prediction of stock prices has been a much fascinated and desired aspect as far as the financial world is concerned. It also very much holds true that an accurate prediction of the future of stock prices gives a huge fruitful return. This paper makes use of both the above mentioned algorithms which have turned out to be very effective in giving out accurate results. RMSE was used as a factor for measuring the accuracy of the model internally by varying the number of times it was tested on (epochs) and the number of concealed neural network layers . The evaluations were carried out by making use of a readily accessible dataset .

### 5. "Prediction Stock Price Based on Different Index Factors Using LSTM" (7-10 July 2019) Chun Yuan Lai; Rung-Ching Chen; Rezyzy EkoCaraka

In recent times the neural networks have been a predominant means in terms of forecasting of stock prices. Although there are various kinds of well known neural networks such as the CNN , the RNN's and the GRU's here this paper adopts the use of LSTM algorithm. The basic idea behind developing this model is that it uses the average information of the five previous day stock prices in order to be able to predict the average of the next five day stock prices. Apart from this it makes use of technical analysis which can help one decide whether he has to buy, sell or retain the stocks. The dataset being tested on is taken from the National Stock exchange of Taiwan.

### 6. "An Empirical Analysis of Stock Market Price Prediction using ARIMA and SVM" (13-15 March 2019) Lokesh Kumar Shrivastav; RavinderKumar

ARIMA is one of the most widely applied algorithm when it comes to time series forecasting of stock prices. But it has its own limitations of not being able to work on non linear patterns. Many reviewed papers that are available have some limits because they concentrate only on one kind of technique to predict or only one dataset. This paper deals with



comparative study of different tools used to forecast stock prices. It does a comparative study on ARIMA model and SVMmodel.

**7. "Apple Stock Price Prediction using ARIMA Model"( December 2019) URANG Awajionyi S. , KABARI Ledisi G. Department of Computer Science, Ignatius Ajuru University of Education, Port Harcourt, Rivers State, Nigeria. Department of Computer Science, Ignatius Ajuru University of Education, Port Harcourt, Rivers State, Nigeria.**

The authors of this paper have conducted a analysis on Time series data using the stock prices of Apple company from NYSE on a monthly basis. They found that the plot which was plotted showed an upward positive trend. The algorithm used to predict apple stock prices here is ARIMA time series algorithm.Their choice was ARIMA because it was widely suited for the prediction. R software has been used. the results showed an accuracy of 97.2 percent.

**8. Twitter Sentiment Analysis(November 18 – 20, 2014) Aliza Sarlan1 , Chayanit Nadam2 , Shuib Basri3 Computer Information Science UniversitiTeknologiPETRONAS Perak,Malaysia**

Socialnetworksiteshaveconnectedmillionsofpeople.Thepublicutilizesthisplatformtoconveytheirviewpointon a diverse set of fields which eventually manipulates vast section of society. Twitter is one of the social media giants. Sentiment analysis pave way for Businesses and Politics use this platform to analyze where they stand in terms of public’ point of view and rapid measures can be taken based on the same.This paper also throws light on various techniques used for sentiment analysis.

### III. METHODOLOGY

#### ARIMA:

This is a part of time series forecasting algorithms quite effective in predicting the prices of the stocks ahead of its time. The previously obtained values are called lags and these lagged errors are used to forecast the future of the stock prices.ARIMA is a part of the statistical models . Time series that belong to the non seasonal category and that usually do not contain any random noises or inflations are the ones that are chosen to be evaluated by the ARIMA model. The most important terms that needs to be considered. While building an effective arima model are ‘p’ which is the order of AR,’ d’ which is the order of differencing and ‘q’ is the MA. The term AR in the ARIMA that stands for Auto regressive indicates that it is based on regression analysis that makes use of its own lags to predict the future of stock prices. A model based purely on AR is such that its lags are its only dependent factors whichprecisely means that Yt behaves as a function of the lags.

This model has been used in the current project to predict future stock prices given the data set of previous stock prices.

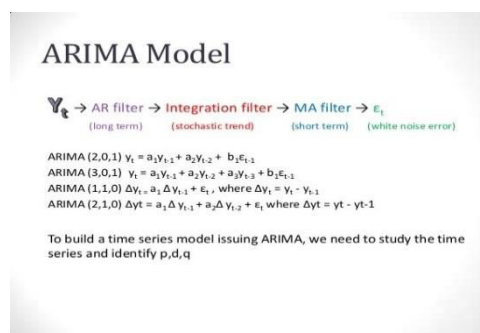


Figure 1. ARIMA model equations

#### NEURAL NETWORKS

The idea for an artificial neural network system evolved from the existing human biological nervous system which consists of networked layers of perceptrons which acts as a replacement to the layers of neurons in the human nervous system. It is a combination of different algorithms that makes it possible for us to perform advanced operations on its data.

#### RECURRENT NEURAL NETWORK

RNN’s belong to a special kind of ANN’s in which the relationships among its nodes form a graph that is directed towards a progressive sequence. They are derived from the feedforward behaviour of the neural networks that makes



use of its extensive memory to be able to process the upcoming varying input sizes. Some of its applications include recognition of handwriting, recognition of speech etc. A finite impulse and an infinite impulse are predominantly the two kinds of networks that make up these recurrent neural networks. Both these kinds of impulses can have extra states that can be stored and controlled directly by the parent neural network. These storage states could in turn be replaced by another network or by a graph in case it consists of any kind of time delays of any kind of result loops. These can also be referred to as the feedback neural networks which is mainly used in dealing with time-based data.

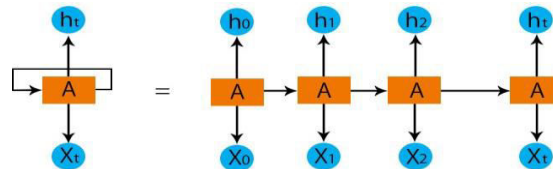


Figure 2. RNN example

The neural network system of the RNN's has a cell state or memory and the input that is being handled depends on its internal state. It consists of repeating modules of tanh layers in RNN's that helps them in storing information but since it doesn't stay for a very long time we are required to use the memory models.

**LONG SHORT TERM MEMORY**

A neural network that can learn data that has dependencies that last long is known as a long and short term memory architecture. The picture shows the four neural networks represented in yellow color, operators represented in green color and input circles in yellow circles and the cell state circles in blue color. It consists of a state of the cell and three gates that helps them to learn selectively, forget the same or even reserve the particulars provided by each of the units. The present condition of the cell helps in the information flow among the units without being changed and making way for minimal interactions. Each and every unit consists of three gates that are input, output and forget which are used respectively to add or remove or forget the data. The information that has to be passed down from the previous state to the next state is determined mainly by the forget gate with the help of something called a sigmoid function. These sigmoid and the tanh functions are multiplication operations. Lastly, the information that has to be transferred to the upcoming hidden state is controlled by the output gates.

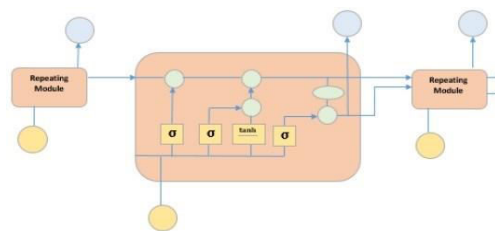


Figure 3. LSTM working

**LINEAR REGRESSION**

Linear regression algorithm is one of the many popularly used supervised learning methods. It works on the basis of regression which makes a target forecast of the dependent variables depending on the independent variables that is being considered and also the count of the independent variables.

Hypothesis function:

$$y = \theta_1 + \theta_2 \cdot x$$

The following are the parameters:

**X:** the input data **y:** labels of the data  **$\theta_1$ :** the intercept

**$\theta_2$ :** the coefficient of variable x



To predict the line that adjusts the best :

$$\text{minimize } \frac{1}{n} \sum_{i=1}^n (\text{pred}_i - y_i)^2 \quad J = \frac{1}{n} \sum_{i=1}^n (\text{pred}_i - y_i)^2$$

Gradient Descent:

In order to minimize the RMSE values we need to upgrade both the theta values for which we make use of the gradient descent algorithm. We begin by taking random values of  $\theta_1$  and  $\theta_2$  values and then we update them repeatedly until we reach a minimum cost value.

### TWITTER SENTIMENT ANALYSIS

Twitter as most of us know of is a hugely popular and a widely used blogging and a social media networking application which acts as a platform for a huge number of messages or posts every single second. Sentiment analysis is a computerized process of acknowledging the various sentiments such as opinions, approaches, perspectives and emotional values from either speech or from texts. The three major categories in which they can be divided into are positive sentiments, negative opinions and neutral ideas. It finds its extensive usage in fields of research study and corporate related political issues allowing them to become aware of the various critics of their products which in turn needs them a helping hand in understanding the customers expectations. It's proven to be very helpful in terms of maneuvering for people working in the marketing fields, helping them detect the faults in their products as a result of the not so positive comments even before these comments or flaws take over their pros. Any kind of contradictions in the political statements at the government level are also very well kept track of. Here in our model we have made use of this module in python to work with regular expression, tweepy, the API which gives access to the twitter rest api techniques, a python library textblob to process textual data and a sentence tokenizer punkt. Fresh Tweets are extracted from the twitter through tweepy. Punkt further splits up a larger body of text into smaller lines, words using unsupervised algorithm. In built functions provided by text blob is utilized to realize the polarity of the instant messages.

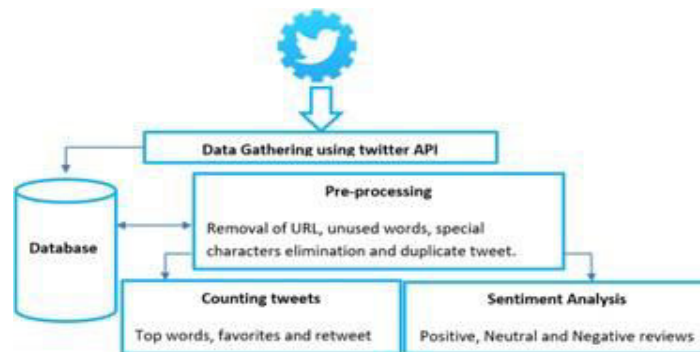


Figure 4. Twitter Sentiment Analysis

### IV. EXPERIMENTAL RESULTS

Firstly here in this paper we have made use of the live yahoo financial dataset which is a python module which takes into account the dataset containing the stock prices for the last two years. The first algorithm that is ARIMA used in this paper is that kind of an algorithm that can be used only with the stationary dataset. In case the data doesn't exist in the stationary state by itself it is our responsibility to convert it into one before implementing the algorithm on it. The most decisive factors in this particular algorithm are the prior value terms, the differenced count and the corresponding moving average term. The second algorithm that we've used here is the Lstm algorithm which is implemented in such a way that it stores the previous seven day average values in its memory cells with help of the multiple recurrent-neural-network layers in order to be able to predict the value for the next day. The third algorithm used here is the linear regression algorithm which is implemented so as to be able to predict the values for the next seven days. Lastly we have implemented the sentiment analysis from Twitter which takes into account the latest 300 tweets to predict the overall sentiment of the company dataset being tested on and consequently tells us if we should buy or sell the stock. The two parameters that we have used for determining the accuracy of this model are as follows.



$$MAPE = \frac{\sum \frac{|A-F|}{A} \times 100}{N} \quad RMSE = \sqrt{\frac{\sum_{i=1}^N (Predicted_i - Actual_i)^2}{N}}$$

The results that were obtained have been tabulated as follows:

Company	Today’s actual closing price	Closing price by ARIMA	Closing price by LSTM and RNN	Closing price by Linear regression
Apple	123.54	125.17	127.09	133.16
Google	2393.57	2375.38	2376.44	2446.07
Amazon	3206.21	3235.43	3272.25	3406.67
TCS	3183.19	3140.083	3130.50	3320.18
ITC	211.45	208.97	210.79	220.32

Table 1. Actual and predicted closing prices by various algorithms

Company	RMSE			MAPE		
	ARIMA	LSTM, RNN	Linear Regression	ARIMA	LSTM, RNN	Linear Regression
Apple	2.37	5.08	8.89	5.09	3.37	6.13
Google	37.84	73.05	89.66	9.57	3.03	3.51
Amazon	49.66	116.55	182.97	4.19	3.05	4.79
TCS	44.00	85.06	208.38	3.13	2.18	5.89
ITC	3.8	7.06	11.86	3.68	2.64	4.85

Table 2. Root Mean Square Error(RMSE) and Mean Absolute percentage Error(MAPE) of algorithms

### V. CONCLUSION

We propose machine learning techniques to predict stock prices. In this paper we adopted yahoo finance as the prime source of input. We then have implemented three significant methodologies that is LSTM -RNN, ARIMA and Linear



Regression. Also twitter sentiment analysis serves as an add on. Union of all these approaches bestow the user with an advice to either buy or sell a stock.

#### REFERENCES

1. Bikrant Bikram P. Maurya B.Tech, Ayush Ray B.Tech, CSE, Aman Upadhyay Bhupesh Gour Professor, Asif Ullah Khan Professor, CSE, LNCTS “Recursive Stock Price Prediction With Machine Learning And Web Scraping For Specified Time Period”, IEEE Conference Publication, 13 February 2020.
2. Lokesh Kumar Shrivastav; Ravinder Kumar. “An Empirical Analysis of Stock Market Price Prediction using ARIMA and SVM”, IEEE Conference Publication, 2020
3. B Jeevan, E Naresh, B P Vijaya kumar, Prashanth Kambli Department of Information Science and Engineering. “Share Price Prediction using Machine Learning Technique”, IEEE Conference Publication, 20 June 2019.
4. M Umer Ghania , M Awaisa and Muhammad Muzammula a Department of Software Engineering, Government College University Faisalabad. “Stock Market Prediction Using Machine Learning(MI) Algorithms”, IEEE Conference Publication, 2019
5. Sumeet Sarode, Harsha G.Tolani, Prateek Kak, C S Lifna-Vivekanand Education Society's Institute of Technology, Mumbai, India. “Stock Price Prediction Using Machine Learning Techniques” , IEEE Conference Publication, 21 November 2019.
6. Muhammad Aiman Asraf Roslan; Mohd Hezri Fazalul Rahiman(August 2018). “Stock Prediction Using Sentiment Analysis in Twitter for Day Trader” , IEEE Conference Publication, August 2018.
7. Yulin Du. “Application and analysis of forecasting stock price index based on combination of ARIMA model and BP neural network”, IEEE Conference Publication, 2018.





**Impact Factor:  
7.569**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details