

Stock Market Increase and Decrease using Twitter Sentiment Analysis and ARIMA Model

S. V. Kedar¹, Savani Kadam²

¹ Faculty, JSPM's Rajarshi Shahu College of Engineering

² Student, JSPM's Rajarshi Shahu College of Engineering

Article History: Received: 10 November 2020; Revised: 12 January 2021; Accepted: 27 January 2021;

Published online: 05 April 2021

Abstract: With centuries and decades, people started evolving and slowly started entering into technology era. Social networks era came before everyone which connected people from far away countries. Such an example of social network are applications like Twitter, Facebook, Instagram, LinkedIn etc. Every application has its own significance. Such an application is Twitter where people tweet regarding their opinion about a topic, a person anything. The tweets regarding company its performance and people's opinion about the stock is also tweeted. People like to invest in stocks using this data posted of social networking. This data keeps them updated about a company. In this paper, we will be using tweets related to stocks so that we can analyse sentiments of people regarding a particular stock. This sentiment analysis can provide a feedback about the company so that we will be able to understand an increase or decrease with respect to the people or company performance. In later stages, we will be comparing this analysis with ARIMA model which is time series forecasting model. ARIMA takes values of stocks and predicts its future prices based on its algorithm. Using both of these techniques a cumulative result for stock exchange will be obtained. The dataset is the fresh tweets taken from twitter and also the stock data will be imported directly for ARIMA.

Keywords: Stock Market, Twitter, Tweets, sentiment analysis, ARIMA, Time Series

1. Introduction

With the increasing Internet era we are going into a life of limitless connection irrespective of distance differences. We communicated through social media from one country to another which is miles away from each other. Internet has brought this evolution for every sector i.e. medical sector, health care, pharmacy, military, farming, banking, security as well as in the stock market where money revolves. Stock market and its prices depends on the performance of a company and also of the country. Currency prices are observed different for various countries. This difference is the gap of developed and developing countries or the economic stability of the country. Thus, stock market plays a vital role in the economy of the country. Many people try and invest in the stocks. When more people invest in stock the country economy increases while it is vice versa then economy decreases [1].

There is also a lot of fluctuation in the stock market. Hence many people have thought of using some computer programs to solve this issue. With the use of these algorithms, users will have less risk while investing into stocks. So, from 10% population investing in stock will become to 50% population [2]. These have also interested many researchers to come forward and implement different techniques to predict the stocks.

Social networking has brought a tremendous change in the lives of people. They can connect and upload their views, thoughts and information through these online platforms like Facebook, Instagram, Twitter, etc. They have created an environment where people can come together and express their thoughts and views about anything. In similar way, these sites can also hold data for stock prices. Experts in this area generally make a public statement sometimes for some stocks. Also, some news is published regarding a stock or a company which determines performance of a company. This performance can enable a user to understand how the stock will trend in mere future.

One such platform taken into consideration here is Twitter. People can tweet or share company information about a stock for any users to analyse [3]. These things can influence a user in investing into stocks as information becomes a lot handy. Also, performance of a company is understood by this method. They also give us the profit or loss about a company well in advance [4]. Many have implemented different techniques for this purpose. Researchers have implemented ANN technique of Elman neural network. There is a context layer, and there is a dynamic system with feedback ability. They can reflect the dynamic system change directly and also has strong computing power [2]. As well as natural language processing is also preferred by the researchers for this study [5]. Different types of SVM classification are taken into consideration by these scientists so that these algorithms can separate dataset accordingly. Regression analysis is also one of the methods which can be taken

into consideration for the same [6]. Different sentiment analysis with various approaches can be seen for sentiment analysis [7-9].

Health and genetics, medical sector has also seen increased use of ARIMA model to prediction of values [10-14]. ARIMA is also in use in various areas like production, tourism, economy etc [15-18]. We will try to scrape and analyse data from online platform like Twitter and access those tweets about a company which will determine its performance and any major changes about it which can affect the stock price. The main aim of our project will be analysing how the effects of social media are seen on the stock market prediction. Moreover, the project will help people in analysing different aspect through social networking as well as through stock market ARIMA and help in investment for that user.

We would be first scraping data from Twitter. Basically, the data will be tweets, tweeted by the user about the company. These tweets will be pre-processed, cleaned and then sentiment analysis will be performed on this data. Later a result will be obtained regarding the positivity in sentiment of company or neutral or negativity of company's performance [19-23]. Tweet is segregated on the words being used in the tweet being positive, negative or neutral. This forms a foundation of the project.

In later phase, stocks data will be downloaded for a specific period from yahoo finance. This data will be then given to ARIMA model to predict future values for these values. ARIMA is a Time series algorithm to predict future series of anything depend on the data given to it.

The remainder of the paper is organized as follows: Section II briefs the Related work, Section III explains the Methodology, Section IV discusses Results and Discussion, and Section V concludes the paper.

2. Related Work

Price prediction in stock market to be one of the most difficult tasks, as price is dynamic and fluctuates always. Earlier study has found out that stocks price volatility is monitored along to the market sentiment for minor company stocks. Some researchers have used social media withdrawal technology for measurable assessment to the market segment, and other factors to predict the stock price trend in short term. Test results conclude that when the users use social media with drawal collective with other evidences shows us the stocks forecast model will give better exactness of the solution [3].

It is important to have reliable model to predict stock as it is very intricate to the criteria of the Indian economy and can sub tune a user to monetary damage. Researchers have also extracted, retrieved, and analysed the impact how the news sentiments in social media would give on the stock price. Their main goal or aim which researcher can contribute can be found in the growth of a sentiment study for financial sector and the evaluation of the model for judging the whole research would present the effects of news sentiments on stocks for the category of pharmaceutical companies in market [4]. Some have proposed multi-class classification on the sentiment analysis. Sentiment analysis is performed with multi-class classification to witness the accuracy in that aspect and use it further for classification to obtain highest accuracy. All different classifications like [Love vs Hate], [Happiness vs Sadness], and [Fun vs Anger] and neutral are bunched together [23].

In freshages, it has been noticed a detonation of curiosity in estimating time series in variable functional areas. Time series predicting has been exposed to actual trending results in appropriate decision making in numerous spheres. A variety of procedures have been proposed to obtain goal of prophecy and examination of literature. People have proposed ARIMA model by smearing a mean of approximation error for time series forecasting[24]. Researchers have anticipated that the most dependable way to conjecture the forth coming events to try to recognize the current and then according to that we have establish our prior aim as the examination of the Indian Stock Market. The research will help in understanding and trying to create a better imminent scope for speculation by collecting statistics on the monthly closing stock indices of Sensex for six years. Establishing the same hypothesis to progress an apposite model which has helped some authors to estimate the imminentun noticed values of the Indian stock market indices. This study offers an application of ARIMA based on which the prediction of the future stock catalogues is done and have a firm effect on the recital of the Indian economy. To establish the model, they embedded the validation technique with the observed data of Sensex of 2013[25].

They have trained and tested various machine learning classifiers such as Multinomial NB, Bernoulli NB, Logistic Regression, SGD classifier, SVC, Linear SVC, and NuSVC. Experimental results demonstrate that Bernoulli NB, Logistic Regression, and SGD classifier reached accuracy as high as 75% [26]. Temperature can also be predicted using time series approach. Due to numerous changes going on in the climate we need to

predict these for utilization in future use. This can be done by neural networks. They have also shown the dependency of temperature series using integrated back propagation with genetic algorithm technique [10]. The intention of ambiguous time series analysis is to prove that indeterminate data with no firm pattern inside is analysed in order to advance knowledge, fit low dimensional models, does prediction. Particle swarm optimization, Euclidean distance, data mining, and Monte Carlo simulation are some approaches that have been related to examine the best ways of foreseeing [8].

3. Methodology

The proposed system is based on the prediction of each stock price which can help the user to analyse the probability of the price upcoming in future. The influence of the social network like twitter on this type of prediction maybe more.

A. Architecture

The tweets which are genuine about a company or news about the company will flash on twitter on its own performance of company which is related to the upcoming trend that a price may increase or decrease. Such type of data is been scrapped. Later, pre-processing and sentiment analysis is done on the data. Through the sentiment analysis polarity of the tweets are also calculated. The tweet percent through which it classified the tweet as positive, negative and neutral is mentioned. This sentiment analysis gives a rough idea of the amount of negative and positive tweets found in the system. Graphs are plotted to verify the same.

Later, to predict stock price of company, system would use ARIMA model. This model will evaluate and analyse company's historical data of stock. Data consists of Dates Opening Price, Closing Price Highest Price, Lowest Price of each day of last couple of years. ARIMA will mathematically work in terms of predicting closing price of company stock in future time. Graphical representation would show the trade index.

Based on this system will suggest best choice to invest in organization in order to have better profit using live data analysis. Figure 1. shows the system architecture

The system contains two modules,

- 1) Scraping data from twitter and making sentiment analysis on data
- 2) Using ARIMA model finding the closing prices of stocks and to predict future prices

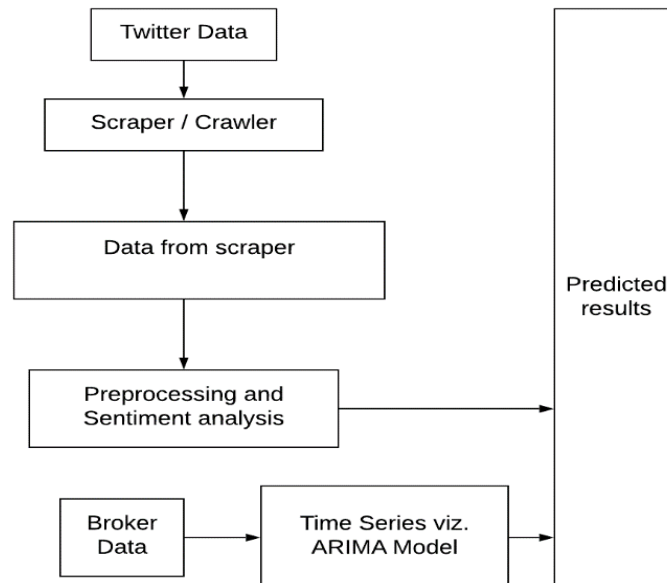


Figure 1. System Architecture

Figure 1. shows that when data is scraped through a scraper how pre-processing is performed on the tweets extracted and later sentiment analysis is done on them. For module 2 we will be executing a time series algorithm of ARIMA to predict a pattern of values

Module 1: Scraping data from twitter and making sentiment analysis on data

For the first module we will be scraping the data from twitter and further perform sentiment analysis on data. This analysis will enable us to check the tweet received will decrease the stock or increase it or will be neutral. This scraping includes extracting all the columns present in the dictionary object. Sentiment analysis also performs pre-processing the data i.e. removing stop words and punctuation. The pre-processed data is then analysed in sentiments. These sentiments enable the system to segregate a tweet as positive, negative or neutral factor.

Module 2: Using ARIMA model finding the closing prices of stocks and predict further prices

This module finds the dataset for whole data and further computes the closing price of the stock. This determines the pattern of the rise and fall that was happening in the company. The instability of the stock is enabled. Also, it predicts in what way further the price will become for the stock. This prediction can be compared with the tweets and produce results in relation to future prices of the stocks.

B. Algorithms

For rise or fall of stock, prediction is done through sentiment analysis and Time series ARIMA model. Sentiment analysis is emotional mining of text which extracts and identifies theoretical information in source input, and helps a business to understand the social sentiment of their product, brand or service while monitoring online conversations.

The basic working of sentiment analysis is:

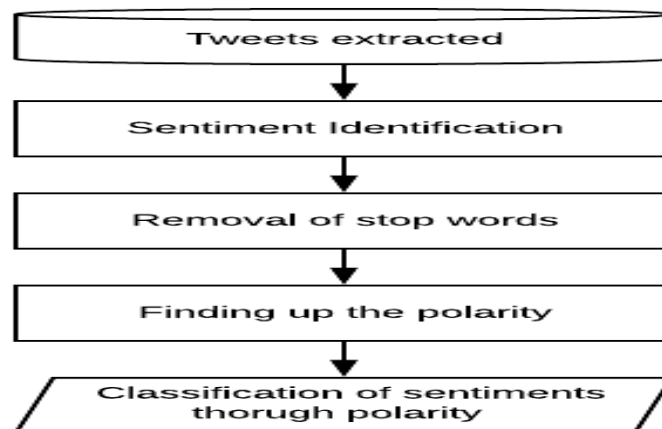


Figure 2. Sentiment Analysis

Figure 2. specifies the algorithm of sentiment analysis and its steps in detail. It starts from **tweets extraction** which is an asset to scrape live data needed for sentiment classification. Next step is **Sentiment identification and removal of stop words** which actuals the root words of any tweets removing all unnecessary words from it so that it can be segregated accordingly as the basic word tends to describe a feeling of positivity, negativity or neutral. Based on the root word, **finding up the polarity** is done and based on that **classification of sentiments** is performed.

Classification of text a basic asset:

Sentiment Analysis is the text classification method that analyses a text given to it and tells whether the underlying sentiment is positive, negative our neutral. You can give your own sentence and give it to the system to verify that it is positive, negative or neutral in behaviour.

Intent Analysis helps us in analysing the user’s intention behind the text he tweeted or gave as an input and identifying whether it relates an opinion, any new news, marketing, complaint, suggestion, appreciation or query.

Sentiment analysis is the interpretation and classification of emotion from within the text to identify that customer is happy sad or has average views regarding a company, product or an event. Practice of utilizing

natural language processing, text analysis to methodically quantify, identify, extract, and learning affective states and independent information.

Algorithm

Sentiment Analysis

Step 1: We give the input as data in .csv or .json format

Step 2: The data obtained is pre-processed by removing punctuations and stop words

Step 3: The data is analysed as positive, neutral or negative words are present

Step 4: Polarity is calculated and values are given

Step 5: Repeat steps 2-4 with all data in the training set.

ARIMA model is a class of numbers for forecasting and analysing time series data. It overtly supplies a suite of typical constructions in time series data. Also, in addition it provides influential technique for skilful time series forecasts.

ARIMA is an acronym to key features being implemented inside it. They are as follows:

AR: Auto-regression. A classical phenomenon that uses the dependency relation between an observation and scarce figures of lagged observations.

I: Integrated. The custom differencing of observations (e.g. subtracting an observation from an observation at the previous time step) in order to make the time series stationary.

MA: Moving Average. A model that uses the dependent phenomenon between an observation and a lingering error from a moving average model and is applied to observations lagged.

Each of the mechanisms specified for ARIMA are overtly specified in the model as a parameter. A standard representation of ARIMA(p,d,q) is where the parameters are replaced with integer values to designate the specific ARIMA which can be taken into consideration.

The variables of the model ARIMA are well-defined as:

p: The number of lag observations included in the model, also referenced the lag order.

d: The number of times the raw observations are differenced, also termed the degree of differencing.

q: The size of the moving average window, also named the order of moving average.

The value of d, is the minutest number of differencing needed to make the sequences stationary. If the time sequence is stationary, then d=0. 'p' is the demand of the 'Auto Regressive' (AR) term. It refers to the number of intervals of Y to be used as forecasters. And the command of the 'Moving Average' (MA) is 'q'. It mentions to the number of lagged prediction errors that should go into the ARIMA Model.

The following are the steps of process:

1. Model Identification. This step involves the data of statistics to identify patterns, seasons and trends to the elements to understand the amount of differencing and the size of the lag that will be required.

2. Parameter Estimation. A suitable technique to find the constants of the regression model.

3. Model Checking. They then use the plots done by the data to the amount and type of chronological construction which won't be captured by the model.

The procedure is recurring until necessary level of fitting is attained on the training or test dataset. Some pre-calculation on the data is done by following processes and formulae.

Rolling mean standard deviation:

To keep track only of n (sample size), $S_1 = \sum_i a_i$ and $S_2 = \sum_i a_i^2$ where a_i is your data. Mean and variance of the data is calculated by:

$$\hat{\mu} = \frac{S_1}{n} \tag{1}$$

$$\hat{\sigma}^2 = \frac{S_2}{n} - \left(\frac{S_1}{n}\right)^2 \tag{2}$$

Here $\hat{\mu}$ is your sample mean, and $\hat{\sigma}^2$ is your sample variance. The hat only means it's an estimation.

The estimation of rolling mean standard deviation is:

$$\hat{\sigma} = \nu \quad (3)$$

Moving Average:

Anuntainted Moving Average model is the one where Y_t depends only on the lagged calculation errors

$$Y_t = \alpha + \epsilon_t + \varphi_1\epsilon_{t-1} + \varphi_2\epsilon_{t-2} + \dots + \varphi_q\epsilon \quad (4)$$

Seasonal values in arima time series: The seasonal ARIMA model includes both non-seasonal and seasonal factors in a multiplicative model. It can easily be given as a formula:

ARIMA(p,d,q)X(P,D,Q)S

with Q = seasonal MA order, q = non-seasonal MA order, P = seasonal AR order, p = non-seasonal AR order, d = non-seasonal differencing, D = seasonal differencing, and S = time span of repeating seasonal pattern.

ARIMA model in words:

Predicted Y_t = Linear Combination of Lagged forecast errors (upto q lags) + Linear combination Lags of Y (upto p lags) + Constant

Algorithm

Time Series ARIMA model

- Step 1:** Input the data of the opening or closing price of the stock.
- Step 2:** The series should be made stationary.
- Step 3:** Filter out validated sample.
- Step4:** Select the terms for the Auto-regression and Moving Average.
- Step 5:** Build the model and the period to forecast the values.
- Step 6:** Compare the predicted values with actual values and find the difference in those values. Stock values will be predicted by this algorithm successfully and range of value of the stock prices for different companies will be known.

4. Results and Discussion

All the research done are conducted using a laptop with Intel® Core™ i5-8265U CPU @ 1.80 GHz, x64-based processor, 8GB RAM.

The dataset is the live data scraped from twitter so that data can be live for stock price as well as social network. The data takes the previous values for testing and gives a series of values of prediction. This in turn helps us to work on the recent data which will be useful for the user in application of risk into stocks. The Python library of text blob was used for the purpose of sentiment analysis. So that as soon as data is put into buffer sentiment analysis is done on that data successfully and results are sorted accordingly. Also, pmdarima and scikit-learn was implemented to build the whole model of ARIMA and its calculations.

When the data was being scraped it was so implemented for 10 companies of the NYSE (New York Stock Exchange) which would result in more tweets from social media. Only such companies are considered again while following the ARIMA model.

As sentiment analysis form the root for analysis of tweets there are polarities which are calculated against the tweets and the pie chart obtained after that prediction is referenced in Figure 3. which represents Google and Figure 4. Shows results for netflix company

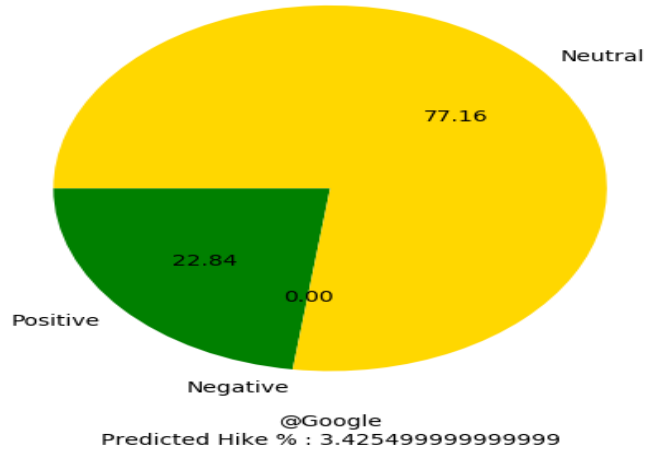


Figure 3. Sentiment Analysis for Google

Hike for a particular stock is then calculated by:

If there are no negative stocks then:

$$ratio = (sum\ of\ positive)/2$$

And for both positive and negative stocks present the value will be:

$$Ratio = (sum\ of\ positive) / (sum\ of\ negative)$$

$$Ratio_final = Ratio / 2$$

Depending on these values Hike will be calculated

The calculations for tweets are done according to what polarity is obtained and can be seen clearly in Table. 1. for google company and in Table. 2. for Netflix company. The words present to depict the sentiment is picked up clearly by the polarity calculations.

Table 1. Sentiment Analyser with polarity(Google)

Tweet	Negative	Positive	Neutral
"@kpgalligan Hi Kevin. We'd like to move this conversation to DM. Please follow us and let us know", 'when you have so we can share next steps.'	0	0.235	0.765
Want to tour Machu Picchu or see the Northern Lights? You can visit these sites on Street View and', 'travel the world in one day. Avoid the jetlag and start exploring → https://t.co/sPPybHTXG6	0.145	0	0.855
@Ignmcrules Hi there. This guide may help: https://t.co/VRPMm6Pb4f . If something looks out-of-line,', 'we suggest adding extra layers of security to your Google Account with these tips:', 'https://t.co/k6X4NgzqKJ.	0	0	1
"@Wayneclift7 Hi there. Let's see what we can do to help. Look out for a DM with next steps.	0	0.137	0.863

Results for 'Netflix Company':

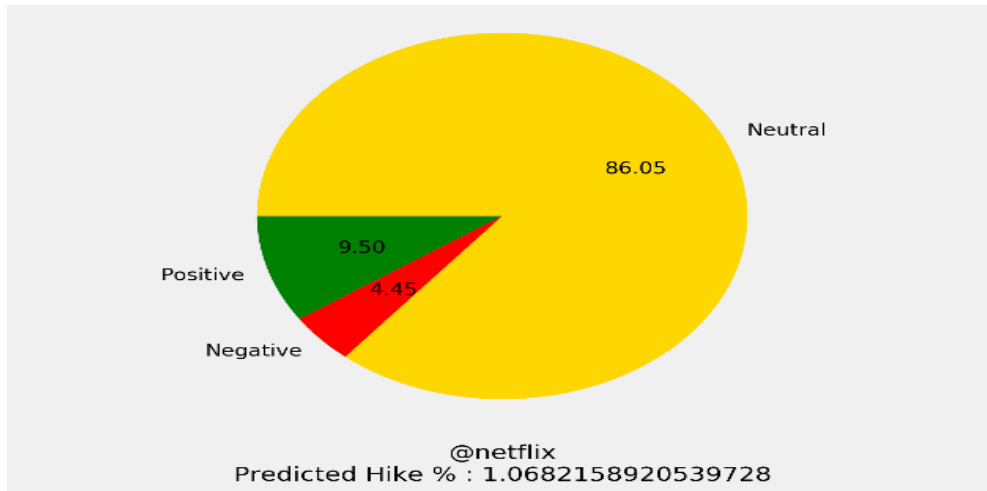


Figure 4. Sentiment Analysis for Netflix

Table 2. Sentiment Analyser with polarity(Netflix)

Tweet	Negative	Positive	Neutral
@BigTechAlert @TheKalenAllen of course we are dont be crazy	0	0.203	0.797
RT @contodonetflix: tfw you hahaha and jajaja #GentifiedLiveRead https://t.co/pa3xDiq1j	0	0.31	0.69
THE LOVEBIRDS Issa Rae and Kumail Nanjiani star as a couple forced to clear their names after they', 'accidentally get caught up in a murder mystery.	0.542	0.458	0

Moving to the ARIMA model, we can take up values for ARIMA is sequential factor and they are a series of values not-stationery. Data is retrieved for the particular stocks which we have analysed for sentiment. Series of the same data is given to train to the ARIMA model created and it further gives us the values of series for that data.

The instinct behind a unit root test is that it controls how sturdily a time series can be defined by a pattern. There are two hypotheses defined, one which verifies that time series has unit root and is non-stationery (Null Hypothesis) and other which verifies that time series does not have a unit root and is stationery (Alternate Hypothesis). P-value will suggest this whether it will be stationary or non-stationery. Results obtained from this dickey fuller test are:

Test Statistics: 0.791510
pvalue: 0.991498

No. of lags used: 11.000000
Number of observations used: 2505.000000
critical value (1%): -3.432963
critical value (5%): -2.862694
critical value (10%): -2.567385

We can conclude here that it comes under the null hypotheses. Also, it means that values have time-dependent structure here.

An ARIMA model is the algorithm where the time series was computed into differentiation at least once to make it immobile and you syndicate the AR and the MA terms.

ARIMA Model Results

=====

Dep. Variable: D.Close
Model: ARIMA (1, 1, 1)
Method: css-mle
Date: Tue, 19 May 2020
Time: 21:36:48
Sample: 1
No. Observations: 2261
Log Likelihood: 6238.758
S.D. of innovations: 0.015
AIC (Akaike information criterion): -12469.517
BIC (Bayesian information criterion): -12446.623
HQIC (Hannan–Quinn information criterion):
-12461.163

MSE: 0.006545229233850982
MAE: 0.07011591910837617
RMSE: 0.0809025910206279
MAPE: 0.009863053459445615
Prediction Accuracy: 90.13694654055439

Table 3. ARIMA predicted values

Date	Values
20-05-2020	6.943274
21-05-2020	6.94382
22-05-2020	6.944363
23-05-2020	6.944907
24-05-2020	6.94545
...	...
12-08-2020	6.988922
13-08-2020	6.989465
14-08-2020	6.990009
15-08-2020	6.990552
16-08-2020	6.991095

Values in Table. 3. are seen for the period of next 90 days from the search for google stocks. Following computations were done on the data: Moving average, Standard deviation and rolling mean was found from the data are closing prices of the stock seen in Figure 5. and Figure 6.

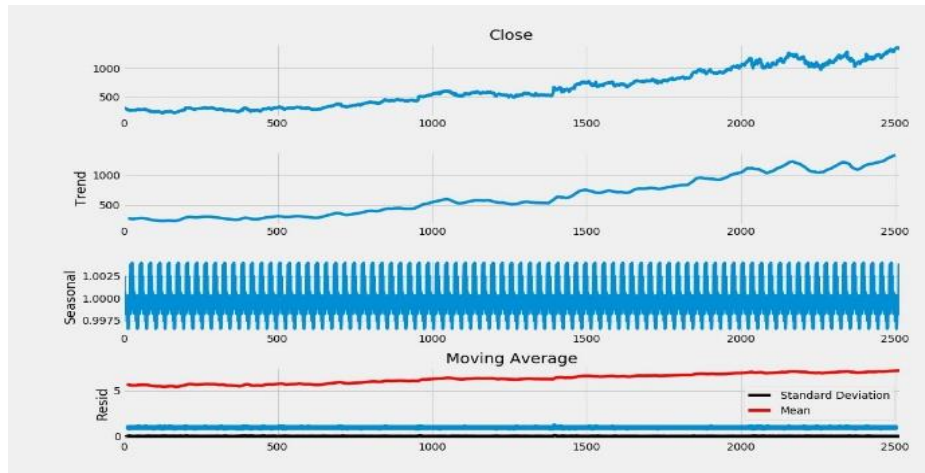


Figure 5. Moving Average Google

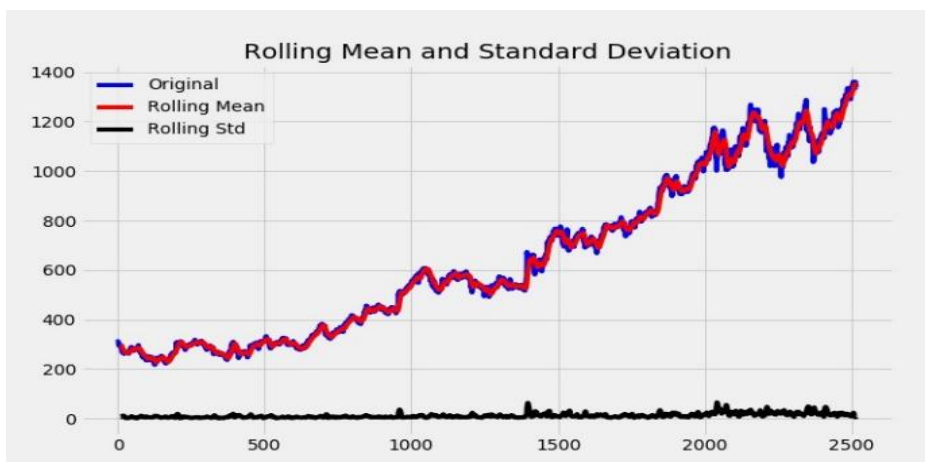


Figure 6. Rolling mean and Standard Deviation of google



Figure 7. Scatter plot of closing price till date of google

The past values of the stock prices of the company are taken into deliberation for plotting the scatter plot for the stock. Figure7. shows a detailed scatter plot of the company. In Figure 8. And Figure 9. The future values of the graph will be seen with respect to the dates. While Figure 10. shows us the performance as well as the prediction of the stock Google.

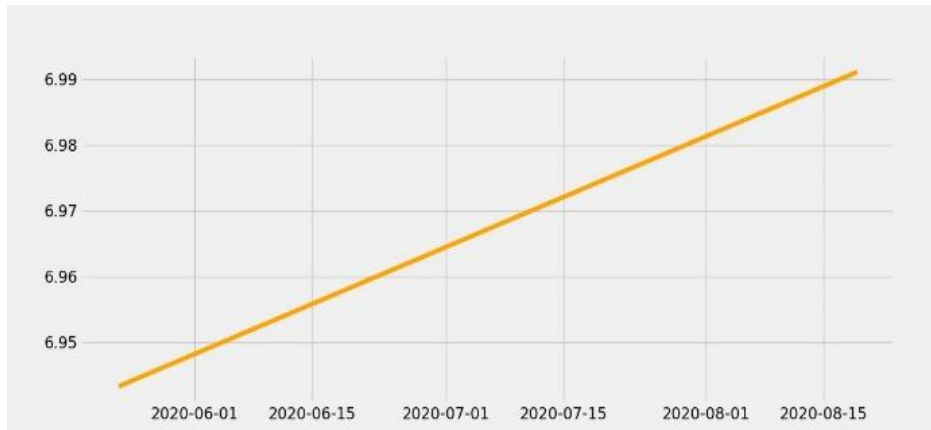


Figure 8. Future values of Google

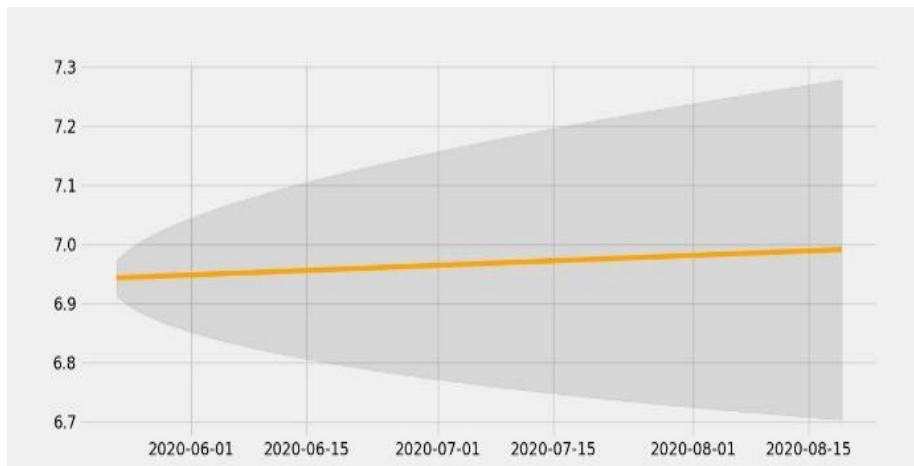


Figure 9. Range of prediction for google

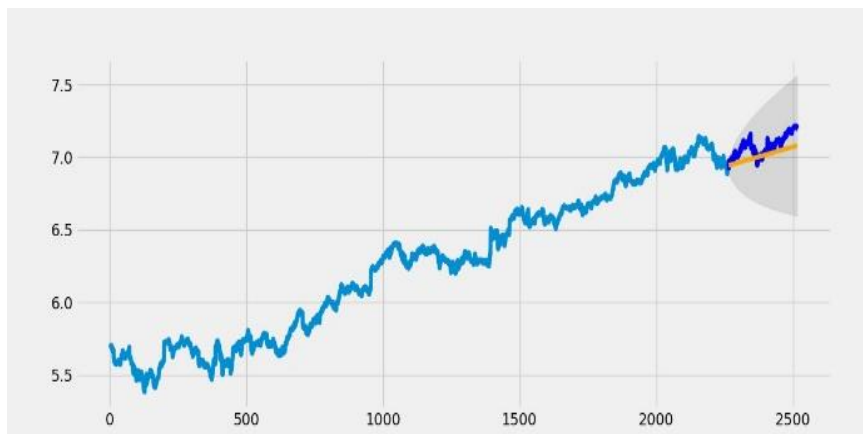


Figure 10. Performance and future curve for google

Same can be seen in the case of different companies implemented using same strategy. Let's take Netflix into consideration for the same. If dickey fuller test results are verified then following can be seen:

Test Statistics: -0.252967
pvalue: 0.931906
No. of lags used: 27.000000
Number of observations used: 2489.000000
critical value (1%): -3.432980
critical value (5%): -2.862702
critical value (10%): -2.567389

We can hence conclude that even these will come under null hypothesis there is a time-dependent structure.

ARIMA Model Results

=====
 Dep. Variable: D.Close
 No. Observations: 2261
 Model: ARIMA(1, 1, 1)
 Log Likelihood: 4463.928
 Method: css-mle
 S.D. of innovations: 0.034
 Date: Mon, 25 May 2020
 Time: 19:58:32
 Sample: 1
 AIC (Akaike information criterion): -8919.856
 BIC (Bayesian information criterion): -8896.961
 HQIC (Hannan–Quinn information criterion):
 -8911.501

MSE: 0.04008166618764743
MAE: 0.18766013208374746
RMSE: 0.2002040613665153
MAPE: 0.03250226637813977
Prediction Accuracy: 67.49773362186022

Table 4. ARIMA predicted values(Netflix)

Date	Values
26-05-2020	6.943274
27-05-2020	6.94382
28-05-2020	6.944363
29-05-2020	6.944907
30-05-2020	6.94545
...	...
18-08-2020	6.988922
19-08-2020	6.989465
20-08-2020	6.990009
21-08-2020	6.990552
22-08-2020	6.991095

Following results are seen for Netflix due to Covid-19 pandemic of year 2020. Results can vary in such cases where stock increase/decrease hampering the growth of industries around the globe.

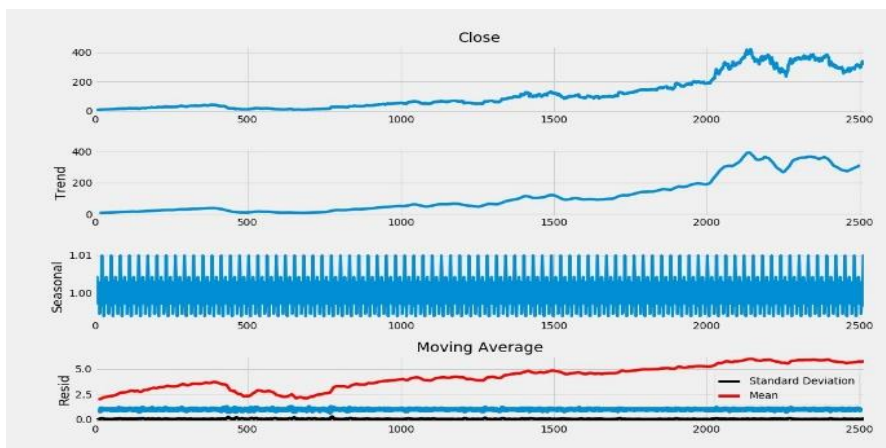


Figure 11. Moving Average Netflix

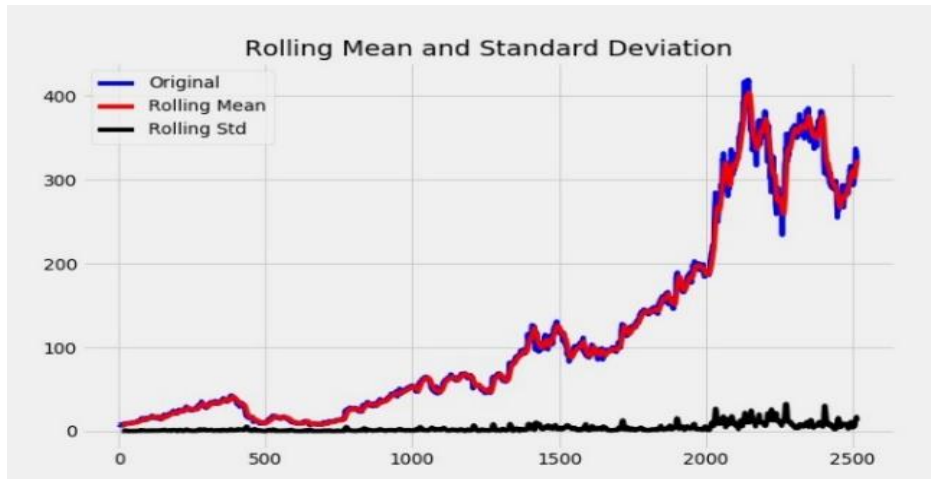


Figure 12. Rolling mean and standard deviation of Netflix



Figure 13. Scatter plot of Netflix

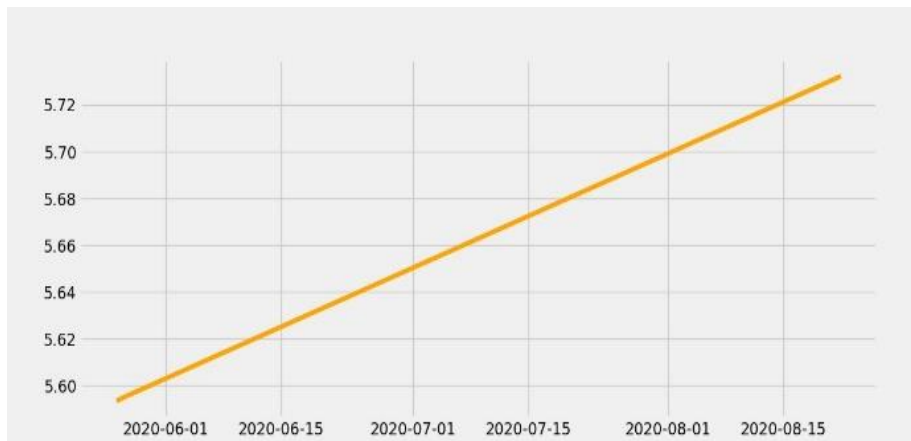


Figure 14. Future values of Netflix

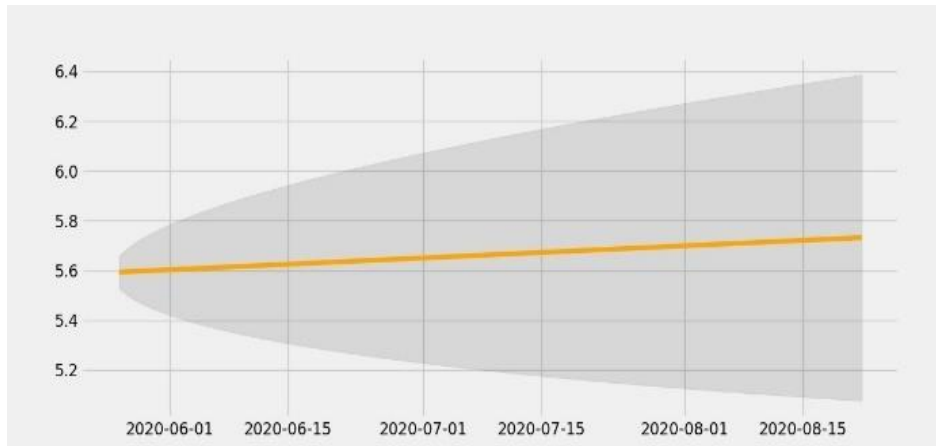


Figure 15. Range of prediction for Netflix

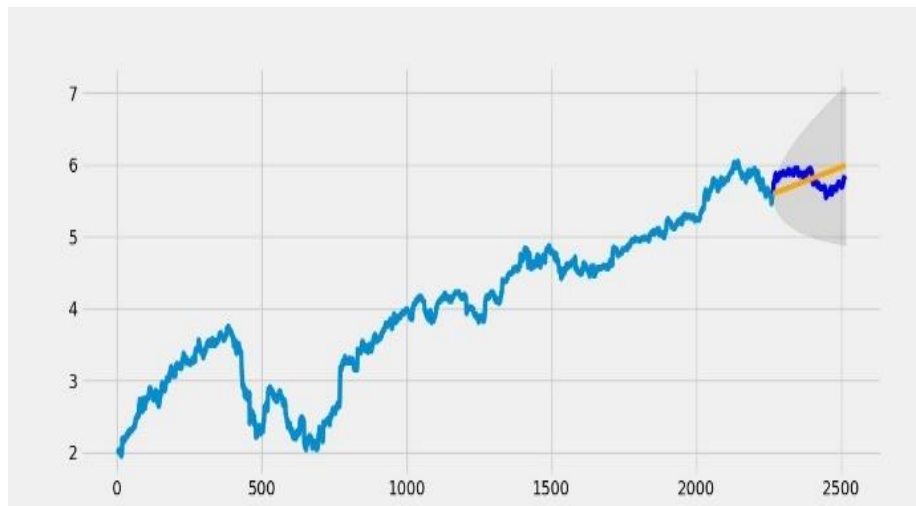


Figure 16. Performance and future curve for netflix

With use of such algorithm’s different applications of stock related usage of data can be covered. We have tried implementing sentiment and ARIMA model for NYSE related companies. Use of dickey fuller test for understanding p-value and implementing recursive methods of calculations like rolling standard deviation, moving average etc. is composed together in this paper which has not termed to be accessed in different researches. All parameters taken from different papers were merged to give results in more descriptive format. Prediction accuracy is found out in this paper for companies of NYSE implemented as a new feature.

Comparison analysis can be referenced in the table below:

Table 5. Comparison Analysis of different Time series algorithm

Author, reference and year	Techniques/algorithm	Values found in papers
SoheilaMehrmolaei and Mohammad Reza Keyvanpour [24] 2016	ARIMA	MSE: 0.09 RMSE: 0.30 MAE: 0.26
Sheikh Mohammad Idrees et. Al. [27] 2019	L-Jung Box	p-value: 0.9099
Shaminder Singh et. al. [10] 2011	Genetic algorithm time series	MAPE: 1.85
KUMAR Manoj et. Al. [15] 2012	Box-Ljung	p-value: 0.6093
Proposed method	ARIMA	MSE:0.0065 MAE: 0.07 RMSE: 0.08 MAPE: 0.0099 Prediction Accuracy: 90.137

		p-value: 0.991
--	--	----------------

With the comparison analysis in Table. 5. we can see that the features proposed algorithm gives us a cumulation of all other approaches. Also, the error found in proposed approach is less than what researchers have tried to obtain. Different measures are added in accordance to other researchers to get a dept regarding the results attained.

4. Conclusion

An approach to understand sentiment analysis and ARIMA is seen here in the paper. Accordingly, both are taken into consideration while taking values for the project. Sentiment analysis helped in segregating the tweets and giving an idea for the trend. While ARIMA has given prediction for those values based on past values. When different companies are taken into consideration the prediction accuracy can show fluctuation due to change in the performance of the company due to COVID-19 pandemic. If data is taken for a company various calculation along with dickey fuller test is seen which is computed together in this paper as an added computation. Observed with the values that if the errors of MSE, MAE etc are less than the accuracy obtained from the method can be increased. Achievement to study the impact of social network on stock market is seen. In the future scope, trying to add more processes or methods to grab hold on changing accuracy can be done.

5. Acknowledgement

The authors of the paper would take this opportunity to acknowledge the publishers as well as the researchers for making their capitals accessible to us and teachers for their direction. We are grateful to the authorities of Savitribai Phule University (SPPU), Pune. We are also appreciative to the reviewers for their treasured suggestions.

References

1. Eugene F. Fama, "The Behavior of Stock Market Prices", the Journal of Business, Vol 2, No. 2, pp. 7–26, January 1965.
2. Ze Zhang, YongjunShen, Guidong Zhang, Yongqiang Song, Yan Zhu, "Short-term Prediction for Opening Price of Stock Market Based on Self-adapting Variant PSO-Elman Neural Network", 978-1-5386-0497-7/17, IEEE, 2017.
3. Yaojun Wang, Yaoqing Wang, "Using Social Media Mining Technology to Assist in Price Prediction of Stock Market" IEEE International Conference on Big Data Analysis, 2016.
4. Dev Shah, HarunaIsah, FarhanaZulkernine, "Predicting the Effects of News Sentiments on the Stock Market", IEEE International Conference on Big Data (Big Data), 2018.
5. J. Zhang, S. Cui, Y. Xu, Q. Li, and T. Li, "A novel data-driven stock price trend prediction system", Expert Systems with Applications, vol. 97, pp. 60-69, 2018.
6. Ashish Sharma, Dinesh Bhuriya and Upendra Singh, "Survey of Stock Market Prediction Using Machine Learning Approach", International Conference on Electronics, Communication and Aerospace Technology, 2017
- A. Brahmananda Reddy, D.N.Vasundhara, P. Subhash, "Sentiment Research on Twitter Data", International Journal of Recent Technology and Engineering, Volume-8, Issue-2S11, September 2019.
7. Hamid Bagheri and MdJohirul Islam, "Sentiment analysis of twitter data".
8. Abdullah Alsaeedi and Mohammad Zubair Khan, "A Study on Sentiment Analysis Techniques of Twitter Data", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 10, No. 2, 2019.
9. Shaminder Singh and Jasmeen Gill, "Time Series based Temperature Prediction using Back Propagation with Genetic Algorithm Technique". IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 5, No 3, September 2011.
10. NabilahFilzahMohdRadzuan, Zalinda Othman, Azuraliza Abu Bakar, "Uncertain Time Series in Weather Prediction". 4th International Conference on Electrical Engineering and Informatics, 2013.
11. Ruben Lam, "Forecasting trends in the Healthcare sector",Procedia Computer Science
12. Sultana Rajia, Md. Sabiruzzaman, Md. Kamrul Islam, Md. Golam Hossain, Pete E. Lestrel, "Trends and future of maternal and child health in Bangladesh", PLoS ONE 14(3): e0211875.

13. VerllyPuspitaandErmata, "Time Series Forecasting for Electricity Consumption using Kernel Principal Component Analysis (kPCA) and Support Vector Machine (SVM)", 10.1088/1742-6596/1196/1/012073.
14. Kumar Manoj and AnandMathu, "An Application of Time Series Arima Forecasting Model for Predicting Sugarcane Production in India", Studies in Business and Economics.
15. PrapannaMondal, Labani Shit and SaptarsiGoswami, "Study of Effectiveness of Time Series Modeling (Arima) In Forecasting Stock Prices", International Journal of Computer Science, Engineering and Applications (IJCSA), Vol.4, No.2, April 2014
16. BiljanaPetrevska, "Predicting tourism demand by A.R.I.M.A. models", Economic Research, VOL. 30, 2017.
17. Jamal Fattah, LatifaEzzine et al., "Forecasting of demand using ARIMA model", International Journal of Engineering Business Management Volume 10: 1–9, 2018.
18. Neethu M S and Rajasree R, "Sentiment Analysis in Twitter using Machine Learning Techniques", IEEE – 3166, 2013.
19. Harpreet Kaur, VeenuMangat and Nidhi, "A Survey of Sentiment Analysis techniques", International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2017.
20. Biswarup Nandi, MousumiGhanti, Souvik Paul, "Text Based Sentiment Analysis", IEEE International Conference on Inventive Computing and Informatics, 2017.
21. HarshaliPatil, Mohammad Atique, "Sentiment Analysis for Social Media: A Survey", IEEE 2nd International Conference on Information Science and Security, 2015.
22. MondherBouazizi and TomoakiOhtsuki, "Multi-Class Sentiment Analysis on Twitter: Classification Performance and Challenges", Big data mining and analytics ISSN 2096-0654 03/05 pp181–194, September 2019.
23. SoheilaMehrmolaei and Mohammad Reza Keyvanpour, "Time series forecasting using improved ARIMA", 978-1-5090-2169-7/16.
24. Debadrita Banerjee, "Forecasting of Indian Stock Market using Time- series ARIMA Model", 2nd International Conference on Business and Information Management, 2014.
25. ShihabElbagir and Jing Yang, "Sentiment Analysis of Twitter Data Using Machine Learning Techniques and Scikit-learn", ACM New York, NY, USA c, ISBN: 978-1-4503-6625-0, 2018.
26. Sheikh Mohammad Idrees, m. AfsharAlam, and ParulAgarwal, "A Prediction Approach for Stock Market Volatility Based on Time Series Data", ACCESS.2019.28952, 2019.