

# Forecasting Time Series Market Data Using ARIMA

Pooja Ahirwar<sup>1</sup>, Prof. Vijay Anand Sullare<sup>2</sup>

<sup>1</sup>M.tech Scholar <sup>2</sup>Research Guide

Department of Computer Science and Engineering, TIT, Bhopal

*Abstract-Data mining is a subfield of software engineering. It is the computational procedure of finding designs in enormous data sets including strategies at the convergence of AI, machine learning, measurements, and database frameworks. During the time spent data investigation, machine learning is a technique which respects expectation for business use. So it is additionally alluded to as prescient examination. Such examination is helpful to specialists, architects and data experts to come to dependable choice and results. It brings out concealed bits of knowledge through learning from patterns in the data and from chronicled connections. In machine learning, design acknowledgment is one of the significant undertakings. Henceforth arrangement is significant in example acknowledgment. The trouble in machine learning with respect to characterization is apportioning to which of a lot of classifications the new perception has a place. By and large, it depends on the preparation data set which contains perceptions whose relationship is watched. At this crossroads, the scientist owes a clarification with respect to time arrangement which is the center of this study.*

*Time arrangement examination includes strategies for dissecting time arrangement data so as to extricate significant measurements and different attributes of the data. Time arrangement anticipating is the utilization of a model to foresee future qualities dependent on recently watched qualities. Data mining strategies are utilized for the investigation of time arrangement in this work. In this examination work we gather securities exchange data (SBI Bank.)From UCI storehouse after this applying Markov chain technique for foreseeing time arrangement data.*

**Keywords:** Stock Market Prediction, Time Series.

## I. INTRODUCTION

A stock market is an open market for companies or for individuals to raise money. Stock market helps companies to purchase or sell their shares. The cost of shares relies on the interest and supplies of shares. This procedure of purchasing and selling of shares is called trading/exchanging; just the Listed Companies are permitted to do exchanging. Recently huge amounts of investment are exchanged via stock market over the world. National economies are firmly connected and intensely affected the execution of their Stock Markets. Therefore they are identified with macroeconomic parameters, as well as they impact ordinary life in a more straightforward

manner. Hence they constitute a mechanism which has vital and immediate social effects. The nature of stock market in common is unpredictable which may depend on the long and short term future state. This is unpleasant and also unavoidable for the speculator when Stock Market is chosen as an investment tool. The main aim is to reduce this unpredictability and the Stock Market Anticipation (or Forecasting) is used in this process.

Stock market forecast is the procedure of attempting to focus the future stock estimation of an organization. The effective forecast of a stock's future cost could return huge benefit. Stock value movements are administered by the speculations random walk hypothesis and efficient- market hypothesis. The predictors of stock market concentrate on creating methodologies which effectively foresee stock costs utilizing overall characterized trading methodologies. A fruitful anticipation model is the particular case that works with best exactness having least input requirements and minimum complex model. Speculators and government associations depend on anticipating techniques to safeguard against threats and to monitor market changes. For analysts, these serve as a source of perspective for investigations of financial issues like estimating money related subsidiaries and portfolio determination.

It is practically impossible to precisely anticipate the future prediction of future events comes with a margin of error. The margin of error augments particularly when anticipating in deep into the future, or when foreseeing. Variables and their normal impact may change (with social, monetary and political change) and new variables may develop. These mistakes emerge as a consequence of the level of error of the base information utilized and the strategy used to estimate into future. This makes the selection of anticipation method significant when foreseeing into future. As a rule prediction uses quantitative information instead of subjective information which rely on upon the judgment of specialists. Anyhow prediction that uses quantitative information is more exact than that of subjective information.

Securities exchanges have been considered again and again to remove valuable examples and foresee their movements. Securities exchange forecast has dependably had a specific interest for analysts and money related financial specialists. The reason is that who can beat the market, can increase overabundance benefit. Cash related specialists who place assets into securities trades standard speaking don't think about the stock trade direct. They are going up against the issue of stock trading as they don't realize which stocks to buy and which to offer with a particular true objective to obtain benefits. If they can envision the future direct of stock costs, they can act speedily upon it and make advantage.

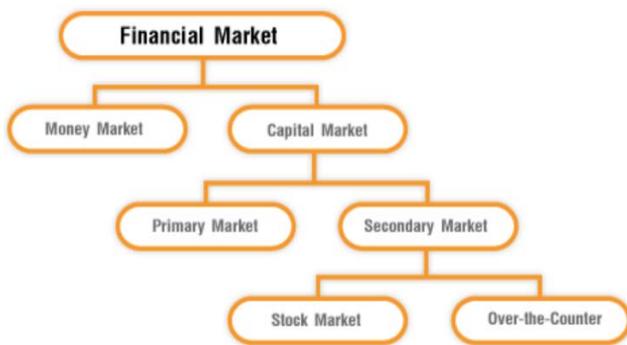


Figure: 3 Types of Stock market

The more precise the framework predicts the stock value movement, the more benefit one can pick up from the expectation demonstrates. Stock value incline forecasting construct exclusively in light of the specialized and basic information examination appreciates extraordinary prevalence. Be that as it may, numeric time series information just contains the occasion and not the reason why it occurred. Printed information, for example, news articles have more extravagant data, consequently abusing literary data particularly notwithstanding numeric time series information builds the nature of the info and enhanced forecasts are normal from this sort of information instead of just numerical information.

## II. STOCK MARKET

Securities exchanges have been considered again and again to remove helpful examples and anticipate their developments. Stock market forecast has dependably had a certain intrigue for scientists. While various logical endeavors have been made, no technique has been found to precisely anticipate stock value development. There are different methodologies in predicting the development of securities exchange and an assortment of forecast systems has been utilized by financial exchange examiners. In the following segments, we quickly explain the two most significant hypotheses in securities exchange expectation. In light of these hypotheses two regular ways to deal with

financial market forecast have developed: Technical and Fundamental investigation (trading methods of insight). The distinction between these two methodologies will be likewise expressed.

From EMH and Random Walk speculations, two distinct trading methods of insight have been risen. These two traditional ways to deal with financial market forecast are specialized examination and basic investigation. In the following segments the distinction between these two methodologies will be expressed.

### A. Technicians Trading Approach

The term specialized examination signifies a fundamental way to deal with stock investing where the past costs are considered, using diagrams as the essential apparatus. It depends on mining standards and examples from the past costs of stocks which is called mining of financial time arrangement. The fundamental principles include ideas, for example, the trending idea of costs, affirmation and difference, and the impact of exchanged volume. A large number of strategies at expectation of stock costs have been created are as yet being created on the grounds of these essential principles.

### B. Fundamentalist Trading Approach

Essential examination investigates the components that influence free market activity. The objective is to accumulate and interpret this information and act before the information is incorporated in the stock cost. The slack time between an occasion and its resulting market reaction exhibits a trading opportunity. Basic examination depends on monetary data of organizations and endeavors to estimate markets using financial data that organizations need to distribute normally, for example yearly and quarterly reports, inspector's reports, monetary records, income proclamations, and so forth. News has significance for investors using principal examination since news depicts factors that may influence free market activity.

## III. PROPOSED METHODOLOGY

As the name recommends, TS (Time Series) is an accumulation of data points gathered at consistent time intervals. These are broke down to determine the long haul pattern in order to conjecture the future or play out some other type of examination. In any case, what makes a TS unique in relation to state a customary regression issue? There are 2 things:

- It is time dependent. So the basic assumption of a linear regression model that the observations are independent doesn't hold in this case.

- Along with an increasing or decreasing trend, most TS have some form of seasonality trends, i.e. variations specific to a particular time frame. For example, if you see the sales of a woolen jacket over time, you will invariably find higher sales in winter seasons.

A period arrangement is a grouping of numerical data points in progressive request. In investing, a period arrangement tracks the development of the picked data points, for example, a security's cost, over a predefined timeframe with data points recorded at standard intervals. There is no minimum or most extreme measure of time that must be included, allowing the data to be assembled in a manner that gives the information being looked for by the investor or expert examining the action. Time arrangement investigation can be valuable to perceive how a given resource, security or financial variable changes after some time. It can likewise be utilized to examine how the progressions related with the picked data point contrast with movements in different factors over a similar timeframe.

This is basic case and as often as possible saw practically speaking. In economy part, the greater part of the indicators are estimated occasionally with the particular time frames, along these lines monetary indicators speak to a proper case of discrete time arrangement. The second choice is, that data are estimated and recorded continuously along the time intervals. Electrical sign from sensors, earth shakings, different indicators from medicine, similar to ECG, or numerous other logical sensors, they all speak to a continuous estimation of corresponding physical amount. This kind of procedures creates a continuous time arrangement. Figure 3.1 shows a seismogram from station HAWA (Hanford, Washington, USA), case of continuous time arrangement.

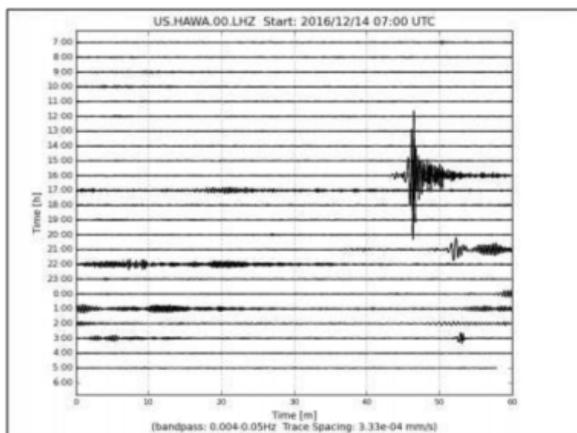


Figure 3.1: Seismogram from HAWA station (Source: Hanford, Washington, USA)

### Time Series Preprocessing

Time-series data is frequently encountered in real world applications. Stock price and economic growth are the typical examples. A time-series represents a set of consecutive observations or measurements taken at certain time intervals. The observations can be real numerical values, for instance, a stock price, or categories, for instance, medicines taken by a patient over a treatment period. These continuous values of a time-series can be converted to categories if needed. Since all stock time series contains a high level of noise, high level time series segmentation is necessary for recognizing the significant movements. Discovery of interesting patterns from a large number of time-series is required in many applications including economic and science.

### Time Series Data Mining

As of late, there has been a great deal of interest within the exploration network in the mining of time arrangement data, which emerge in business just as logical choice help applications; models include stock costs or cash trade rates gathered after some time. Finding intermittent examples in time arrangement databases is a significant data-mining task with numerous applications. In time arrangement data mining, intermittent example help to find the ascent and drop of stock qualities. Have created numerous techniques for searching periodicity designs in enormous data sets. Be that as it may, most past techniques on periodicity hunt are on mining full occasional examples, where each point in time contributes (correctly or around) to the cyclic conduct of the time arrangement.

One of the errands for time-arrangement data mining is scan for intermittent examples. address three various types of example to be specific repeating design, pattern example and occasional example. A repeating example exists when the data display rises and falls that are not of a fixed period. A pattern example exists when there is a long haul increase or lessening in the data. Furthermore, an occasional example exists when an arrangement is influenced via regular components. Many time-arrangements in this present reality have patterns and the stock estimation of a value reserve could be a case of pattern design. have defined the significant undertakings in time arrangement data mining as followings:

- Indexing: Given a query time series  $Q$ , and some similarity/dissimilarity measure  $D(Q,C)$ , find the nearest matching time series in database DB.
- Clustering: Find natural groupings of the time series in database DB under some similarity/dissimilarity measure  $D(Q, C)$ .

- Classification: Given an unlabeled time series Q, assign it to one of two or more predefined classes.
- Segmentation: Given a time series Q containing n data points, construct a model from K piecewise segments ( $K \ll n$ ) such that Q closely approximates Q.

The technique utilized in this investigation to create ARIMA (Autoregressive integrated moving average) model for stock value forecasting is explained in detail in subsections beneath. The instrument utilized for usage is Orange programming variant 3. Stock data utilized in this exploration work are authentic every day stock costs obtained from SBI Bank stock traded.

The data made out of four components, specifically: open value, low value, high cost and close cost separately. In this exploration the closing cost is picked to speak to the cost of the index to be anticipated. Closing cost is picked on the grounds that it mirrors every one of the exercises of the index in a trading day.

#### IV. RESULTS ANALYSIS

Stock value forecast will dependably remain an interesting region for the investors, merchants and analysts to investigate. It has pulled in the interest of analysts to grow better prescient models once in a while. From the subject point of view, stock costs expectations are a significant theme of the subjects, for example, finance and financial matters. The financial institutions and individual investors need a successful procedure to figure and take choice dependent on these conjectures on everyday schedule. The stock value expectation is viewed as a standout amongst the most troublesome errand to accomplish because of complex nature of financial exchanges and investment danger of the securities exchange. This remains an inspiring angle for research researchers to advance with new prescient models or improve the existing ones till date, different models and systems had been utilized to foresee stock costs. Among these ARIMA models are the most famous factual calculations. ARIMA models are known to be effective for forecasting especially for momentary expectation in the field of financial time arrangement and even the more mainstream than ANNs methods in numerous cases.

#### ARIMA Method

The components of this model are consisting of three parts: an Autoregressive (AR) part, a moving Average (MA) part and 'I' the differencing part. This model is particularly referred to as the ARIMA (p, d, q) model (Box-Jenkins). Here 'p' is the order of the Autoregressive part, 'd' is the order of differencing and 'q' is the order of the moving average part. For instance, an ARIMA (2, 1, 2)

model means, that it contains 2 Autoregressive (p) parameters and 2 moving Average (q) parameters and it is differenced once to attain the stationarity of the time series data.

#### General process of forecasting using ARIMA models

ARIMA forecasting processes consists of four steps/stages as mentioned below.

- Model Identification
- Model Parameter Estimation
- Diagnostic Checking
- Forecasting

#### Implementation

The stock market is comprised of various participants with different risk capacity and return characteristics, different perceptions and expectations about stocks and the economy. Investors interpret and react on the information/news in different ways. They focus on diverse pieces of information and reach different conclusions. The level of impact of such information on the stock prices Investment decisions are made by the investors mainly on fundamental and technical analysis of the stocks. Fundamentals of the stocks include study of financial statements of the company involving balance sheet analysis, profit and loss account analysis, etc. for the particular company under consideration. Promoters and management details are taken from annual report and their reputation is also considered for the investment in any stock

ARIMA model is applied on historical data of stock prices. This data has been obtained over duration of 2 years starting from 24.04.2009 to 24.05.2019 for SBI bank.

#### Implementation done in Orange version 3.20

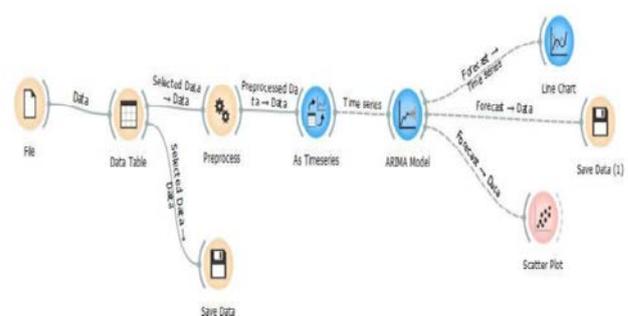


Figure 4.1: Proposed ARIMA model.

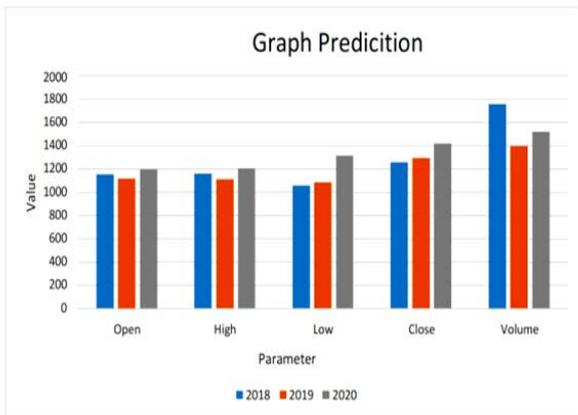


Figure 4.2 Graph Stock data Comparison in different years

## V. CONCLUSION AND FUTURE SCOPE

In this research work, we have tried to forecast the stock prices of SBI bank. The most used statistical models are AR, MA, ARMA and ARIMA. Based the trend and seasonality anyone of them can be applied in order to have better outcome. To apply the time series techniques, we have collected stock price data of 10 years (2009-2019) in CSV format of the above-mentioned stock. The stationary of these stock prices has been moved using decomposing and differencing technique. After that using AUTO ARIMA p, d, q values have been determined to use the appropriate model and finally using that model stock prices for next thirty days are forecast which has been visualized using graph. So, forecasting a stock price is not an easy task to do as it requires removing the dependencies from a time series which lead to complex calculation. As a result, prediction can sometimes be erroneous which can be bad for a business. Therefore, these calculations should be carried out carefully in order to get a better result.

In future works, we would like to concentrate more on the public sentiments of a company to forecast the stock prices. Sentiments can be analyzed and manipulated in order to determine people’s opinions regarding a company. These public opinions can be analyzed from Twitter or facebook and financial news. A hybrid model which combines historical data and sentiments can be developed in order to predict stock prices more accurately. Other factors like environmental factors such as flood, storm etc. can analyzed and combine with sentiment to predict the stock prices and increase the accuracy. As stock market is a huge and random sector more work can be done to find better and accurate predictive times models.

## REFERENCES

[1] Tay, F.E.H., Shen, L., and Cao, L., 2003. “Ordinary Shares, Exotic Methods: Financial Forecasting Using Data Mining Techniques.” River Edge, New Jersey (NJ): World Scientific Publishing Co., Inc.

[2] Fama, E.F., 1970. “Efficient Capital Markets: A Review of Theory and Empirical Work. Papers and Proceedings of the Twenty-Eighth Annual Meeting” (New York, December 28-30, 1969) of American Finance Association, *Journal of Finance*, 25(2), pp.383-417.

[3] Frawley, W.J., Piatetsky-Shapiro, G., and Matheus, C.J. 1991. “Knowledge Discovery in Databases: An Overview.” In G. Piatetsky-Shapiro and W.J. Frawley eds. *Knowledge*

[4] Fayyad U.M., Piatetsky-Shapiro, G., and Smyth, P., 1996a. “From Data Mining to Knowledge Discovery: An Overview”. In U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth and R. Uthurusamy eds. *Advances in Knowledge Discovery and Data Mining*. Cambridge, Massachusetts: AAAI / MIT Press, 1996, pp.1-34. Also in *AI Magazine*, 17(3), pp.37-54.

[5] Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P., 1996b. “The KDD Process for Extracting Useful Knowledge from Volumes of Data”. *Communication of ACM*. New York: ACM Press, 39(11), pp.27-34.

[6] Parker, J., Sloan, T.M., and Yau, H., 1998. “Data Mining. EPCC Technology Watch Report. Edinburgh Parallel Computing Center (EPCC)”, University of Edinburgh. [Online] Available from: <http://www.epcc.ed.ac.uk/>

[7] Karanikas, H., and Theodoulidis, B., 2002. “Knowledge Discovery in Text and Text Mining Software”. Centre for Research in Information Management (CRIM), UMIST University, UK. [Online], Available from: [www.crim.co.umist.ac.uk/parmenides/internal/docs/Karanikas\\_NLDB2002%20.pdf](http://www.crim.co.umist.ac.uk/parmenides/internal/docs/Karanikas_NLDB2002%20.pdf)

[8] Hearst, M.A., 1997. “Text Data Mining: Issues, Techniques, and the Relationship to Information Access”. Presentation Notes for UW/MS Workshop on Data Mining. [Online]. Available from: [www.ischool.berkeley.edu/~hearst/talks/dm-talk](http://www.ischool.berkeley.edu/~hearst/talks/dm-talk).

[9] Sullivan, D., 2000. “The Need for Text Mining in Business Intelligence”. *DM Review Magazine*. December 2000 Issue. [Online] Available from: [http://www.dmreview.com/article\\_sub.cfm?articleId=2791](http://www.dmreview.com/article_sub.cfm?articleId=2791)

[10] Hearst, M.A., 1999. “Untangle Text Data Mining”. In *Proceedings of the 37th conference on Association for Computational Linguistics on Computational Linguistics (Annual Meeting of ACL, College Park, Maryland, June 20-26, 1999)*. Morristown, New Jersey (NJ): Association for Computational Linguistics, 1999, pp.3-10.