A Brief Survey on Stock Price Prediction Using the ARIMA Model

Pooja Ahirwar[1], Prof. Vijay Anand Sullare[2]

[1]M.tech Scholar [2]Research Guide

Department of Computer Science and Engineering, TIT, Bhopal

Abstract-Since the objective of any endeavor is to surrender use to eat up more and better in future, people add to secure increasingly unmistakable wealth and focal points. Today, prediction has been considered as a champion among the most huge pieces of science in economic and business fields and has been developed rapidly. Looking with countless influencing factors, the economic and business boss is ceaselessly scanning for a framework that will help them in choosing. Stock regard prediction is a boss among the most generally broke down issue, drawing in specialists from different fields. The temperamental idea of the offer trading system makes it inconceivably hard to apply basic time-course of action or backslide procedures.

Keywords-Stock price prediction, text mining, stock market, short term prediction Efficient Market Hypothesis.

I. INTRODUCTION

Modeling and Forecasting of the financial market has been an attractive topic to scholars and researchers from various academic fields. Financial market is an abstract concept where financial commodities such as stocks, bonds and precious metals transactions happen between buyers and sellers. In the present scenario of financial market world, especially in the stock market, forecasting the trend or the price of stocks using machine learning techniques and artificial neural networks are the most attractive issue to be investigated.

Financial forecasting is an instance of signal processing problem which is difficult because of high noise, small sample size, non-stationary, and non-linearity. The noisy characteristics mean the incomplete information gap between past stock trading price and volume with future price. Stock market is sensitive with political and macroeconomic environment. However, these two kinds of information are too complex and unstable to gather. The above information that cannot be included in features are considered as noise. The sample size of financial data is determined by real world transaction records. On one hand, a larger sample size refers a longer period of transaction records; on the other hand, large sample size increases the uncertainty of financial environment during the sample period.

Money related series forecasting has been tended to since the 1980s. The goal is to beat budgetary markets and win much benefit. Up to this point, budgetary estimating is still viewed as a standout amongst the most difficult utilizations of present day time arrangement determining. Money related time arrangement have extremely complex conduct, coming about because of countless which could be monetary, political, or mental. They are intrinsically uproarious, non-stationary, and deterministically chaotic.

Because of the multifaceted nature of money related time series, there is some incredulity about the consistency of monetary time series. This is reflected in the outstanding efficient market hypothesis theory (EMH) presented by. As per the EMH hypothesis, the present cost is the best forecast for the following day, and purchase hold is the best exchanging methodology. Be that as it may, there are solid confirmations which reject the proficient market speculation. Hence, the undertaking isn't to question whether money related time arrangement are unsurprising, yet to find a decent model that is fit for depicting the elements of budgetary time series.

One of the uses of content mining is finding and abusing the connection between the archive content and an outer wellspring of data, for example, time stamped floods of information to be specific stock market quotes. Anticipating the developments of stock costs in view of the substance of news articles is one of numerous utilizations of content mining strategies. Data about organization's report or breaking news stories can significantly influence the offer cost of a security.

At present, the stock brokers rely on upon Intelligent Trading Systems which support them in anticipating costs focused around different circumstances and conditions, which also help them in immediate investment decision. Stock market price are thought to be exceptionally dynamic and capable of fast changes since the basic nature of the financial space and partially due to the mix of known parameters (P/E Ratio, Previous Day's Closing Price etc.) and some different components. A trained trader would foresee the stock value and purchase a stock before the cost of stock climbs, or sell it before its esteem

decreases. It is difficult to supplant the expertise that an experienced trader has picked up from his experience but an exact forecast algorithm can straightforwardly come into high benefits, individual experts, which shows an immediate relationship between the precision of the prediction algorithm and the benefit produced using utilizing the algorithm. The key benefit of prediction is that the provision of different stakeholders with profitable data that can be utilized to take decision regarding the future.

Text Mining

The span of data is expanding at exponential rates step by step. Practically all kind of foundations, associations, and business enterprises are putting away their data electronically. A tremendous measure of content is streaming over the web as advanced libraries, archives, and other literary data, for example, websites, web-based social networking system and messages. It is moving assignment to decide suitable examples and patterns to separate important learning from this huge volume of data. Customary data mining instruments are unfit to deal with literary data since it requires time and exertion to extricate information.Figure:1 demonstrates the movement of content mining.

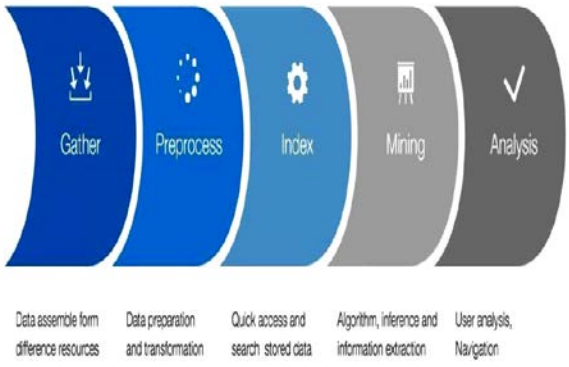

Figure: 1 the Activity of Text Mining.

Content mining is a procedure to concentrate fascinating and critical examples to investigate learning from literary data sources. Content mining is a multi-disciplinary field dependent on data recovery, data mining, machine learning, measurements, and computational phonetics. A few content mining methods like rundown, arrangement, grouping and so on., can be connected to extricate information. Content mining manages regular language content which is put away in semi-organized and unstructured organization. Content mining procedures are consistently connected in industry, the scholarly world, web applications, web and different fields. Application zones like web crawlers, client relationship the board framework, channel messages, item proposal examination, misrepresentation recognition, and web based life investigation use content mining for conclusion mining, include extraction, notion, prescient, and pattern examination. Figure: 2 represent the contrast between Data mining and Text mining.

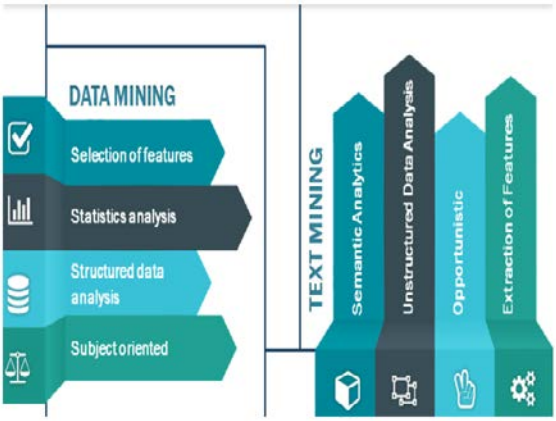

Figure: 2 Difference between Data Mining and Text Mining.

Generic process of text mining performs the following steps:

• Collecting unstructured data from different sources available in different file formats such as plain text, web pages, pdf files etc.

• Pre-processing and cleansing operations are performed to detect and remove anomalies. Cleansing process make sure to capture the real essence of text available and is performed to remove stop words stemming (process of identifying the root of certain word) and indexing the data.

• Processing and controlling operations are applied to audit and further clean the data set by automatic processing.

• Pattern analysis is implemented by Management Information System (MIS).

• Information processed in the above steps are used to extract valuable and relevant information for effective and timely decision making and trend analysis.

- Stock Market Prediction

When predicting the future costs of financial exchange securities, there are two significant speculations accessible. The first is Efficient Market Hypothesis (EMH) introduced by Fama (1964) and the second one is Random Walk Theory. The following segments give the distinction between these two normal hypotheses.

A. Efficient Market Hypothesis (EMH)

Fama's commitment in proficient market speculation is noteworthy. The Efficient Market Hypothesis (EMH) states that the present market cost mirrors the osmosis of all the information accessible. This implies given the information, no forecast of future changes in the cost can be made. As new information enters the framework the unequal state is promptly found and immediately eliminated by the right change in the cost. (Fama, 1970) Fama's hypothesis breaks EMH into three structures: Weak, Semi-Strong, and Strong.

B. Random Walk Theory

An alternate point of view on forecast originates from Random Walk Theory. In this hypothesis, financial exchange expectation is accepted to be unthinkable where costs are determined haphazardly and outperforming the market is infeasible. Irregular Walk Theory has comparative hypothetical underpinning to Semi-String EMH where all open information is thought to be accessible to everybody. Notwithstanding, Random Walk Theory announces that even with such information, future forecast is ineffective.

II. LITERATURE SURVEY

Sr. no.	Title	Author	Year	Approach
1	Predicting SET50 stock prices using CARIMA (Cross Correlation ARIMA)	S. Wichaidit and S. Kittitornkun	2015	This paper aims proposed method is called CARIMA (Cross Correlation Autoregressive Integrated Moving Average. The basic idea of CARIMA is to find the most highly correlated s tock t o predict the target one in addition to ARIMA predicted price.
2	Stock Price Prediction Using the ARIMA Model	A. A. Ariyo, A. O. Adewumi and C. K. Ayo	2014	This paper presents extensive process of building stock price predictive model using the ARIMA model. Published stock data obtained from New York Stock Exchange (NYSE) and Nigeria Stock Exchange (NSE) are used with stock price predictive model developed.
3	ARIMA modeling of wind speed for wind farm reliability analysis	A. K. Rajeevan, P. V. Shouri and U. Nair	2014	This paper is intended to develop an autoregressive integrated moving average (ARIMA) time series model for wind speed. In ARIMA time series modeling, it is possible to change a nonstationary time series to a stationary time series by taking differences.
4	Forecasting household natural gas consumption with ARIMA model: A case study of removing cycle	M. Akpinar and N. Yumusak	2013	In this study autoregressive integrated moving average (ARIMA) method is used and main idea in this study is removing cycling component in time series.
5	Predictive data mining on Average Global Temperature using variants of ARIMA models	C. Narendra Babu and B. Eswara Reddy	2012	This paper analyzes and predicts the Average Global Temperature time series data. Three different variants of ARIMA models: Basic ARIMA, Trend based ARIMA and Wavelet based ARIMA have been used to predict the average global temperature.
6	Sichuan Province service industry development forecast - based on ARIMA model	Zhu Haoyun and Chen Xu	2010	In order to better grasp the development trend and study the development path of service industry, this paper built ARIMA forecasting mode based on related statistical data from 1978 to 2008 to predict the service industry development of Sichuan Province, and analyze the forecast results.
7	Stock Price Prediction: Comparison of Arima and Artificial Neural Network Methods - An Indonesia Stock's Case	Y. B. Wijaya, S. Kom and T. A. Napitupulu,	2010	In this paper, we compared the stock forecasting result of ANTM (PT Aneka Tam bang) using Artificial Neural Network and ARIMA. ARIMA is a technique of time-series forecasting, which means forecast based on the existing pattern.

S. Wichaidit and S. Kittitornkun [1] Investing in stocks is one of the most popular approaches for money investment. This paper aims to predict short-term stock prices of SET50 of Stock Exchange of Thailand (SET). The

proposed method is called CARIMA (Cross Correlation Autoregressive Integrated Moving Average. The basic idea of CARIMA is to find the most highly correlated s tock t o predict the target one in addition to ARIMA predicted price. The results of CARIMA model yield better price trends (measured by 10-day correlation coefficient) while % MAEs (Mean Absolute Errors) are quite similar with those of ARIMA.

A. A. Ariyo, A. O. Adewumi and C. K. Ayo [2] Stock price prediction is an important topic in finance and economics which has spurred the interest of researchers over the years to develop better predictive models. The autoregressive integrated moving average (ARIMA) models have been explored in literature for time series prediction. This paper presents extensive process of building stock price predictive model using the ARIMA model. Published stock data obtained from New York Stock Exchange (NYSE) and Nigeria Stock Exchange (NSE) are used with stock price predictive model developed. Results obtained revealed that the ARIMA model has a strong potential for short-term prediction and can compete favourably with existing techniques for stock price prediction.

A. K. Rajeevan, P. V. Shouri and U. Nair [3] The output power of a wind energy conversion system (WECS) fluctuates with wind speed variations. Accurate wind speed modeling is essential to forecast wind power changes in a site. This paper is intended to develop an autoregressive integrated moving average (ARIMA) time series model for wind speed. In ARIMA time series modeling, it is possible to change a nonstationary time series to a stationary time series by taking differences. A wind speed collection of one year from a 99MW wind farm situated in Theni, Tamil Nadu, India is used in this modeling. The developed model is used to evaluate annual reliability indices like loss of load probability (LOLP), loss of load expectation (LOLE), and loss of energy expectation (LOEE) by convolving wind farm generation model with load model of the grid. Furthermore, variations of LOLE and reliability with changes in peak load are carried out. The study illustrates that system risk index LOLE improves with decrease in peak load and WECS has high reliability to meet the changes in peak load.

M. Akpinar and N. Yumusak [4] Forecasting natural gas consumption in Turkey is very important at energy sector. For this purpose kindly prediction methods are used. In this study autoregressive integrated moving average (ARIMA) method is used and main idea in this study is removing cycling component in time series. For removing cycling, time series divided monthly data and merged co-exhibiting behavior months. Same months and different years data is merged and called as “Model” and 6 Models are prepared. Last model; Model 7 is a general model that includes all consumption data. ARIMA models are applied and mean absolute percent errors (MAPE) are found. Selected minimum MAPE and values of (p, d, q) predictions for Models. For 2012, predictive values of models and Model 7 are compared with actual consumptions. Model that removed cycling (Merged Model) 2.2% better than Model 7.

C. Narendra Babu and B. Eswara Reddy [5] This paper analyzes and predicts the Average Global Temperature time series data. Three different variants of ARIMA models: Basic ARIMA, Trend based ARIMA and Wavelet based ARIMA have been used to predict the average global temperature. Out of all the three linear models, it has been observed that Trend based ARIMA method outperforms basic ARIMA method and Wavelet based ARIMA method outperforms Trend based ARIMA method. MAPE (Mean Absolute Percentage Error), MaxAPE (Maximum Absolute Percentage Error) and MAE (Mean Absolute Error) have been used as a performance measures to compare between the models.

Zhu Haoyun and Chen Xu [6] In order to better grasp the development trend and study the development path of service industry, this paper built ARIMA forecasting mode based on related statistical data from 1978 to 2008 to predict the service industry development of Sichuan Province, and analyze the forecast results.

Y. B. Wijaya, S. Kom and T. A. Napitupulu [7] Neural Network is a network that resembles a human brain tissue, which may infer a result based on the facts or experience that happened. Many applications have implemented neural network. In this thesis, we compared the stock forecasting result of ANTM (PT Aneka Tam bang) using Artificial Neural Network and ARIMA. ARIMA is a technique of time-series forecasting, which means forecast based on the existing pattern. The results of the study showed that forecasting using Artificial Neural Network method has higher accuracy value than the results with ARIMA method.

III. PROBLEM FORMULATION

Cash related specialists who place assets into stock trades ordinarily don't think about the offer exchanging framework 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 ultimate objective to secure advantages. All of these customers understand that the progression of the offer exchanging framework depends a lot on critical news and they have to deal step by step with enormous proportion of data. They need to examine all the

news that shows up on daily papers, magazines and other printed assets. In any case, investigation of such measure of money related news and articles so as to extricate helpful learning surpasses human abilities. Content mining procedures can help them consequently removing the valuable information out of printed assets

IV. CONCLUSION

The current condition in the stock trading system is finding the opportunity to be one of the rule subjects of vitality for specific individuals. The inspirations are specific – from a theoretical need to raise capital or adventure resources, to the craving to envision on its reason the lead of the whole economy. The stock trade gives brisk and devious businesses, in a short range makes fortunes, and in the interim foils decades-long endeavors of past ages. It is a spot for starting and pushed players. In any case, both have a near objective to broaden benefits.

As present day science and technology is getting rich well ordered, more data are being associated in the field of cash to envision its future conditions to restrict adversity and intensify advantage. To address these deciding issues various algorithms have been made and associated with test the effect. Regardless, sadly no such model confirmations to envision the stock market viably as different components and conditions expect noteworthy occupation in choosing future stock expenses.

REFERENCES

[1] S. Wichaidit and S. Kittitornkun, "Predicting SET50 stock prices using CARIMA (Cross Correlation ARIMA)," 2015 International Computer Science and Engineering Conference (ICSEC), Chiang Mai, 2015, pp. 1-4.

[2] A. A. Ariyo, A. O. Adewumi and C. K. Ayo, "Stock Price Prediction Using the ARIMA Model," 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, Cambridge, 2014, pp. 106-112.

[3] K. Rajeevan, P. V. Shouri and U. Nair, "ARIMA modeling of wind speed for wind farm reliability analysis," 2014 Annual International Conference on Emerging Research Areas: Magnetics, Machines and Drives (AICERA/iCMMD), Kottayam, 2014, pp. 1-5.

[4] M. Akpinar and N. Yumusak, "Forecasting household natural gas consumption with ARIMA model: A case study of removing cycle," 2013 7th International Conference on Application of Information and Communication Technologies, Baku, 2013, pp. 1-6.

[5] Narendra Babu and B. Eswara Reddy, "Predictive data mining on Average Global Temperature using variants of ARIMA models," IEEE-International Conference On Advances In Engineering, Science And Management (ICAESM -2012), Nagapattinam, Tamil Nadu, 2012, pp. 256-260.

[6] Zhu Haoyun and Chen Xu, "Sichuan Province service industry development forecast - based on ARIMA model," 2010 International Conference on Computer Application and System Modeling (ICCASM 2010), Taiyuan, 2010, pp. V4-592-V4-595.

[7] Y. B. Wijaya, S. Kom and T. A. Napitupulu, "Stock Price Prediction: Comparison of Arima and Artificial Neural Network Methods - An Indonesia Stock's Case," 2010 Second International Conference on Advances in Computing, Control, and Telecommunication Technologies, Jakarta, 2010, pp. 176-179.