

Time Series Approach for Trade Market Prediction

¹Ms. Savani Kadam, ²Dr. S. V. Kedar

¹Student, Department of Computer Engineering,
JSPM's RSCOE, Tathawade, Pune

²Head of Department, Department of Computer Engineering,
JSPM's RSCOE, Tathawade, Pune

¹savani.kadam@gmail.com, ²seema_kedar@yahoo.com

Abstract

Stock market and the changes in price of money dependent on progress of the country has always been a probability of rise and falls which not predicted by humans with great accuracy. This needs a solution so as the risks of investing into stocks can be reduced by the user. In time along with moving technology and increase of artificial intelligence in day to day life, we have proposed a whole system for prediction of stocks prices for a user based on the news and tweets tweeted on the twitter platform. The increasing use of social network has let the world receive updates about a stock or a company in every minutes of refresh. These updates from a trusted user will be matched and hence will be further segregated via sentiment analysis of data. Further study will be done using a research-based system consisting of algorithmic approach for updated technology which is Time Series Forecasting. Auto Regressive Integrate Moving Average i.e. ARIMA model from Time Series Forecasting will be implemented ARIMA works progressively artful in prediction problem in terms of value bind to time. In this paper, Time Series will work for trade market prediction on the basis of value of stocks on each date of last 10 years of various stocks and companies. Also, the trusted user tweets will be picked up and will predict the increase or decrease in the rates of stocks in upcoming future.

Keywords: Trade Market, Data Analysis, Natural Language Processing, Time Series

1. Introduction

The exponential growth of the Internet and the WWW (World Wide Web) has increased into each facet of our daily routine has enhanced our skill to access information at all times. In the 21st century, online communities and social media has become an important aspect of the internet. The massive growth of user-generated content in online communities and social media had deeply affected financial markets. It is widely accepted that online community's social media contains invaluable information for financial markets and trading.

Since their inception, both have received a growing attention as an interesting source of market analysis. Stock is a market is an important factor here which increases the financial status of a country [1]. It can be seen that due to continuous fluctuations observed in market only some people invest in the stocks. It can be counted as only 10% of the whole country's population [1]. To increase the investment in these stocks people have started implementing it through computations and different algorithms [2]. Prediction of the stocks for a company has fascinated many scholars in numerous arenas such as statistics, operations research and computer science.

Social networking sites also have people tweeting the same about a stock price [3]. There are influences of people tweeting on the stock price. There are sources of social networking which can enable a user to predict the forecast of the price of upcoming prices. Stock news sentiment depicts the rate of change of company and the profit or loss happening to a company at a larger scale [4]. Many different methods are implemented till date. Some have implemented it using the ANN technique of Elman neural network (ENN). The system has a context layer and added to it is a dynamic system with has a feedback ability. These systems can replicate the alteration faster and has robust computation [2]. While others have used Natural Language processing methods for prediction of stock market data [5]. SVM has also been a great significant way of establishing the classification and prediction [3]. Regression analysis is also performed for predicting the stock data [6]. ARIMA algorithm is also used in various sectors like Health care, genetics etc [7-11].

Here, we try to analyze and exploit the behavior of web users and online communities; not only by analyzing how much they say, but also who says, and when; including user profiles and metadata available from social media. The main goal is analyzing the impact of online communities and social media on financial stock markets time series patterns. We build models to predict corresponding stock market with the use of artificial intelligence and time series algorithm.

The outcomes of our project would provide additional reliance to the likelihood of scheming efficient trading models grounded on the examination of web users, online communities and social media data. Moreover, this project intentions to advance specialists kind of the influence of social networking media into stock markets, providing them with additional tools to increase their investment performance and mitigate their risk.

In our paper, the approach of sentiment investigation and then implementing time series of stock data. This together will lead us to the prediction of stock price. This process involves various steps as follows. First, the data is scraped from Twitter. This includes tweets from a particular stock. It can be tweets given by a user as well as tweets news of the company itself. These helps build a concrete base for information regarding to the company. These tweets scraped are the raw data which needs to be pre-processed. It can be pre-processed using various elements like removal of stop words, removing punctuations etc. The pre-processing is the basic and most important step on data as we require clear root words for sentiment analysis [12].

Then sentiment analysis is performed on the pre-processed data. The tweet is being categorized as positive, negative or neutral [13-16]. These tweets are segregated depending on the words that form a whole sentence. More of positive words then negative words tend to shift the sentence in positive side. In such a way company is analyzed and positive graph will obtain or negative graph will be obtained is found out. These graphs help in understanding a positive or negative growth in the upcoming stocks of the company.

Later stock market data is retrieved of a particular company. This data consists of companies opening prices or closing prices. Time Series ARIMA model approach is implemented on these prices. ARIMA model basically attempts to take the prices of the stock and then predict the future prices itself. These prices are based on the trained prices pattern observed in the training data given to the algorithm.

2. Motivation

In the era of technology and increasing use of machines to reduce human effort, it is necessary to design machine for financial purposes. People should increase their perspective and liking towards investing in stocks as this increases the economy of the country. In intention to many people in investing in stocks, the system will be created for a stock market predictor to reduce the risk of the prices which fluctuate continuously after every interval. These will help people invest in stocks with a smaller number of loss per cent and also ensure profit to them by investing.

3. Related work

Yaojun Wang, Yaoqing Wang [3] have proposed price forecast in trade market is measured to be a bit of tedious factors as the price is dynamic i.e, it fluctuates always. Earlier learnings showed us that trade price unpredictability is carefully associated to the market sentiment, which has a major hand on the small companies progressing to its peak. Authors have specified the social media mining expertise to give measurable assessments to market segment and in concoction with other aspects to expect the stock price trend in short period. Research consequences prove it to us that by using social media mining shared with other material, the stock prices calculation model can estimate more exact (Yaojun Wang & Yaoqing Wang, 2016).

Dev Shah, Haruna Isah et al. [4] presents the dependability of the calculative replicas on prediction for stock prices. It is crucial as it's delicate to the country and can lead to monetary damage. Researchers have recovered, mined, and studied the consequences of news sentiments on the stock data for that particular company. Their main aids are the expansion of "sentiment analysis dictionary" for the economic categories, the expansion of a "dictionary-based sentiment analysis" model, and the "assessment of the model for judging the impacts of news sentiments on stocks" for the medicinal company market (Dev Shah, Haruna Isah, & Farhana Zulkernine, 2018). With newscast sentiments, they attained a correctness of 70.59% in foreseeing the patterns in brief movements of prices of stocks.

Shaminder Singh, Pankaj Bhambari et al. [7] stated that 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 (Shaminder Singh, Pankaj Bhambari, & Jasmeen Gill, 2011). In this paper, authors have given the deep summarization of ramification of over training and under training.

Nabilah Filzah Mohd Radzuan et al. [8] has made reviews about different time series implementations on time series statistics in climate calculation. The purpose of their research was to check that ambiguous data when feed with the information to increase knowledge, fits alongside short dimensional models, and does calculation. "Euclidean distance, Monte Carlo simulation, data mining and particle swarm optimization" are approaches that have been associated to examine the finest ways of envisaging. They have shown the results and performance of all the methods implemented till date from since the 19th century.

Mondher Bouazizi and Tomoaki Ohtsuki [16] gave multi-class classification. Sentiment analysis is done with multi-class classification to observe the accuracy in that aspect and use it for further classification to obtain highest accuracy. The resulting method of "multi-class classification" reaches an accurateness of 60.2% for seven various sentiment classes when compared to an accurateness of 81.3% for binary classification. This highlights the impact of having numerous classes on the organization act. All various

classifications like “[Love vs Hate], [Happiness vs Sadness], and [Fun vs Anger] and neutral are clustered together” (Mondher Bouazizi and Tomoaki Ohtsuki, 2019).

Soheila Mehrmolaei, Mohammad Reza Keyvanpour [17] states that in fresh years, we have seen a blast of attention increasing for the predicting of time series in various functional extents. Forecasting is a key goal of TMs withdrawal of time series. There has been a successful hike in fit decision making in time series estimating. Range of methods has been claimed to get to the vision of calculation and scrutiny of content in this zone in different perspective (Soheila Mehrmolaei & Mohammad Reza Keyvanpour, 2016). In this paper they have proposed an innovative solution to advance by smearing a mean of estimation error in ARIMA model for time series predicting.

Debadrita Banerjee [18] have given the solution to most consistent method to guess the imminent issues and to comprehend the existing and thus they have set some aims to their research conducted. Their research has given the examination of the current state of the Indian Stock Market (NSE and BSE) so as to grasp and shot a better forthcoming scope for investment (Debadrita Banerjee, 2014). On this upbringing of their research thesis, they poised data on the regular closing stock directories of Sensex for six years (2007-2012) and grounded on these they have been anxious to mend a suitable outcome of analysis which would help them to estimate the upcoming disregarded morals of the Indian stock market directories. This thesis offers a submission of ARIMA model grounded by the prediction of the upcoming stock directories will have a robust impact on the functioning of the Indian economy. The stock market of India is the center of curiosity for countless researchers, stockholders and businessmen and thus it is vital for them to keep an eye on the visualization that will happen in the present position (Debadrita Banerjee, 2014). To create the solution, they did smear the authentication method with the data of Sensex 2013.

Shihab Elbagir and Jing Yang [19], proposed a model which has the skill to recognize sentimentality from tweets using various ML procedures. This had bound curiosities in many researchers due to more efficiency induced by the systems provided by these machine learning procedures. Researcher has proposed in the paper which attacks to the bottom of every algorithm and its efficiency while performing sentiment analysis of Twitter data. Author has also briefed about the library sklearn which is also called as Sci-kit learn. Analyses on Twitter datasets is made which was copyrighted and available in NLTK Corpora. While doing this they have also taken features into consideration which is needed for the model. They have trained and tested various machine learning classifiers such as Multinomial NB, Bernoulli NB, Logistic Regression, SVC, SGD classifier, Linear SVC, etc. Investigational fallouts prove that Bernoulli NB, Logistic Regression, and SGD gave accuracy up to 75% (Jing Yang and Shihab Elbagir, 2018).

Sheikh Mohammad Idrees, M. Afshar Alam et al. [20] stated that there is a significance of time series analysis and forecasting because it is increasing its reach around all spheres. It mentions to methodical arrangement just like a clusters of information elements that can give value at regular intervals of time. There has always been a mentality of users to look at the stock market as most complex changing structure. There are various stocks to be dealt here whose price fluctuates by reverence to time. This all comprises recognition of the market patterns by stock market forecasting. Many of the investors and businessmen object to make the most of the earnings over their funds and diminish the jeopardies. Stock markets are always subtle and liable to sudden variations that develop in a company. The main aim of the prediction given by author is to progress new pioneering tactics to forecast consequence in huge returns. Their techniques are implemented to examine the time series data of the Indian stock market. Based on the same they have also

created a statistical model that can easily predict the forthcoming prices of the stocks (Sheikh Mohammad Idrees, M. Afshar Alam, & Parul Agarwal, 2019).

3.1. Existing work disadvantages

1. ARIMA approach was performed on NSE and BSE data which showed huge pattern changes in the prices.
2. Different method to combine both algorithms to not implemented.
3. Data is taken out and predicted only from single ARIMA model.

4. Conclusion

It very helpful to use upgraded technologies for reducing human work and error eventually. Also, to get better output which tend to improve work accuracy. Investment in progressive form which is very important part of financial risk could be get better suggestions with the help of these high technologies. Sentiment Analysis helps in segregating the data into positive, negative or neutral tweets. Thus, prices will be predicted with these patterns. ARIMA would help in better understanding of predicted price and also the effect of social media on the stock market prices.

References

- [1] Eugene F. Fama, “The Behavior of Stock Market Prices”, the Journal of Business, Vol 2, No. 2, pp. 7-26, (1965).
- [2] Ze Zhang, Yongjun Shen, Guidong Zhang, Yongqiang Song, Yan Zhu, “Short-term Prediction for Opening Price of Stock Market Based on Self-adapting Variant PSO-Elman Neural Network” IEEE, 978-1-5386- 0497-7/17 67, (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 (ICBDA), (2016).
- [4] Dev Shah, Haruna Isah, Farhana Zulkernine, “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”. ICECA, (2017).
- [7] Shaminder Singh and Jasmeen Gill, “Time Series based Temperature Prediction using Back Propagation with Genetic Algorithm Technique”. IJCSI, Vol. 8, Issue 5, No 3, (2011).
- [8] Nabilah Filzah Mohd Radzuan, Zalinda Othman, Azuraliza Abu Bakar, “Uncertain Time Series in Weather Prediction”. 4th ICEEI, (2013).
- [9] Ruben Lam, “Forecasting trends in the Healthcare sector”. Procedia Computer Science
- [10] 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.
- [11] Verlly Puspitaand Ermatita, “Time Series Forecasting for Electricity Consumption using Kernel Principal Component Analysis (kPCA) and Support Vector Machine (SVM)”. 10.1088/1742-6596/1196/1/012073
- [12] Neethu M S and Rajasree R, “Sentiment Analysis in Twitter using Machine Learning Techniques”. IEEE-3166, (2013).
- [13] Harpreet Kaur, Veenu Mangat and Nidhi, “Survey of Sentiment Analysis techniques” International Conference on I-SMAC, (2017).
- [14] Biswarup Nandi, Mousumi Ghanti, Souvik Paul, “Text Based Sentiment Analysis”. IEEE International Conference on Inventive Computing and Informatics (2017).

- [15] Harshali Patil, Mohammad Atique, “Sentiment Analysis for Social Media: A Survey”. IEEE 2nd ICISS, **(2015)**.
- [16] Mondher Bouazizi and Tomoaki Ohtsuki, “Multi-Class Sentiment Analysis on Twitter: Classification Performance and Challenges”. BIG DATA MINING AND ANALYTICS ISSN 2096-0654 03/05 pp181-194, **(2019)**.
- [17] Soheila Mehrmolaei and Mohammad Reza Keyvanpour, “Time series forecasting using improved ARIMA”. 978-1-5090- 2169-7/16.
- [18] Debadrita Banerjee, “Time Series Forecasting of Indian Stock Market using Time-series ARIMA Model”. ICBIM, **(2014)**.
- [19] Shihab Elbagir and Jing Yang, “Sentiment Analysis of Twitter Data Using Machine Learning Techniques and Scikit-learn”, ACM New York, NY, USA c 2018, ISBN: 978-1- 4503-6625-0, **(2018)**.
- [20] Sheikh Mohammad Idrees, M. Afshar Alam and Parul Agarwal, “Prediction Approach for Stock Market Volatility Based on Time Series Data”. ACCESS.2019.2895252, **(2019)**.