Research Article

Time Series Approach for Trade Market Prediction

Ms. Savani Kadam and Dr. S. V. Kedar

Department of Computer Engineering Rajarshi Shahu College of Engineering Pune, India

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*

Introduction

The exponential growth of the World Wide Web into every aspect of our lives has improved our ability to access data in real time and in these ages. In the 21st era of changing technology, online communities and social media has become an important aspect of the internet. The increasing growth and usage of user-created data and content in online communities and social media had deeply affected financial and economic markets. It is understood that in online communities social media contains invaluable information for financial trading and related area. Since their start, they both have received a growing and increasing attention as an interesting source of market analysis. Stock 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 stocks. It can be said as only 10% of the whole country's population [1]. To increase the amount of human investment in these stocks, people have started implementing it through computations and different algorithms. Stock price prediction has attracted many researchers in multiple disciplines such as computer science, statistics, economics, finance, and operations research [5].

Social networking sites also have people tweeting the same about a stock price. 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 [3]. Stock news sentiment depict 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. It has a context layer, and it is a dynamic system with feedback ability. This kind of network can reflect the change of dynamic system directly and also has strong computing power [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 [16-20].

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 results of this project would give further faith to the possibility of designing efficient trading models based on the analysis of web users, online communities and social media data. Moreover, this project aims to improve specialists understanding of the influence of social media into stock markets, providing them with additional tools to increase their investment performance and mitigate their risk.

In this paper, we will use the approach of sentiment analysis and then implementing time series of stock data. This together will be led 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 [7].

Then sentiment analysis is performed on the pre-processed data. The tweet is being categorized as positive, negative or neutral [8-11]. 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 analysed 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.

The structure of this paper is designed as follows. In Section II, literature survey is described, followed by Section III, which provides details on the proposed methodology. Section IV presents the partial experimental results. Finally, in Section V we have summarized our conclusions.

Literature Survey

Dev Shah, Haruna Isah [4] presents the reliability of the computational models on stock exchange prediction. it's important because it is extremely sensitive to the economy and may directly cause loss. During this paper, they need retrieved, extracted, and analyzed the consequences of stories sentiments on the stock exchange. Their main contributions include the event of a sentiment analysis dictionary for the financial

sector, the event of a dictionary-based sentiment analysis model, and therefore the evaluation of the model for gauging the consequences of stories sentiments on stocks for the pharmaceutical market. With news sentiments, they achieved a directional accuracy of 70.59% in predicting the trends in short-term stock price movement.

Mohammad Reza Keyvanpour, Soheila Mehrmolaei [12] states that in past some years, there has been an explosion of interest in forecasting databases in several applied areas.

Forecasting is one among the most goal's mining of your time series databases. Statistic forecasting has been shown effective in suitable deciding in various domains. So far, a spread of techniques is proposed to get goal of prediction and analysis of literature this area is in several directions. During this paper they need proposed a completely unique approach to enhance ARIMA model by applying a mean of estimation error for statistic forecasting.

Debadrita Banerjee [13] have proposed that the foremost reliable thanks to forecast the longer term is to undertake to know this and thus, accordingly they've set their prior objective because the analysis of this scenario of the Indian stock exchange so on understand and check out to make a far better future scope for investment. On this context, they have collected data on the monthly closing stock indices of Sensex for 6 years (2007-2012) and supported these we've tried to develop an appropriate model which might help us to forecast the longer-term unobserved values of the Indian stock exchange indices. This study offers an application of ARIMA model supported which we predict the longer-term stock indices which have a robust influence on the performance of the Indian economy. The Indian stock exchange is that the center of interest for several economists, investors and researchers and hence it's quite important for them to possess a transparent understanding of this status of the market. To establish the model, they applied the validation technique with the observed data of Sensex of 2013.

Mondher Bouazizi and Tomoaki Ohtsuki [11] gave multi-class classification. Sentiment analysis is completed with multi-class classification to watch the accuracy therein aspect and use it for further classification to get highest accuracy. The proposed approach of seven different classes on a multi-class classification achieves an efficient accuracy of 60.2% which, compared to an accuracy of 81.3% for binary classification, emphasizes the effect of getting multiple classes on the classification performance. All different classifications like [Love vs Hate], [Happiness vs Sadness], and [Fun vs Anger] and neutral are clustered together.

Proposed Methodology

The proposed system is based on the prediction of each stock price which can help the user to analyze 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 user which is genuine may tweet is related to the upcoming news that a price may increase or decrease. The

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 analyzed company's historical data of stock. Data consists of Dates Opening Price, Closing Price Highest Price, Lowest Price of each day of last 10 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.

The system contains two modules,

1) Scraping data from twitter and making sentiment analysis on data
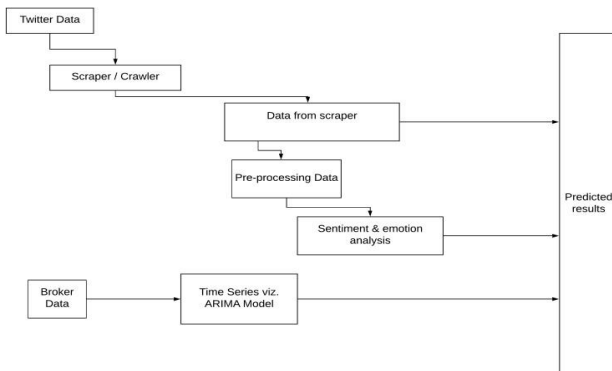2) Using ARIMA model finding the closing prices of stocks and creating an AI to predict further prices

Fig. 1. System Architecture

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 analyzed 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. The basic working of sentiment analysis is:
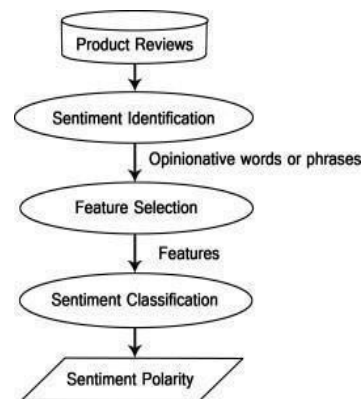

Fig. 2. Sentiment Analysis

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.

Text Classifier — The basic building blocks Sentiment Analysis

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 behavior.

Intent Analysis helps us in analyzing 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 (also known as opinion mining or emotion AI) is the use of (NLP) natural language processing, computational linguistics, text analysis, and biometrics to systematically quantify, identify, extract, and study affective states and subjective 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 analyzed as positive words 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 statistics models for forecasting and analyzing time series data. It explicitly caters to a suite of standard structures in time series data, and as such provides a simple yet powerful method for making skillful time series forecasts.

This acronym is descriptive, capturing the key aspects of the model itself. Briefly, they are:

AR: Auto-regression. A model that uses the dependency relation between an observation and few numbers of lagged observations.
I: Integrated. The use of differencing of raw 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 residual error from a moving average model and is applied to observations lagged.

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
Step 4: Select Auto-regression and Moving Average term.
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 will be known

Results and Discussion

A very basic sentiment analysis model is built on the partial data scraped from the tweets. 1500 tweets were taken from the twitter. The results of partial execution are shown in Fig. 3 and Fig. 4. Fig. 3 shows the pie chart of apple stocks and Fig. 4 shows the pie chart for google stocks.
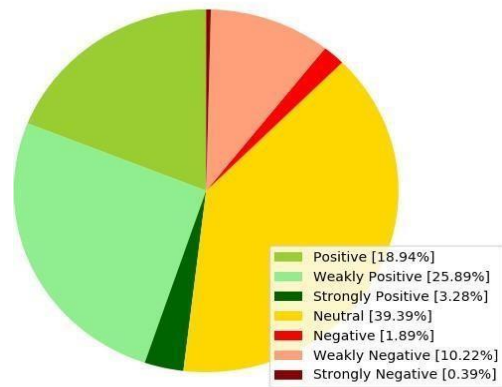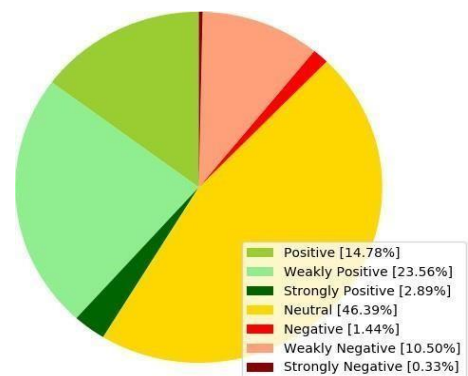

Fig 3. Analysis for AAPL

Fig 4. Analysis for GOOG

Fig 5. shows the prediction using ARIMA model. x axis and y axis show prices and years respectively. Fig. 6 shows forecasting using ARIMA. x axis and y axis show prices and years respectively. Both the graphs indicate the price rise and price fall in consecutive years. Furthermore, the accuracy can be increased to a larger extent.
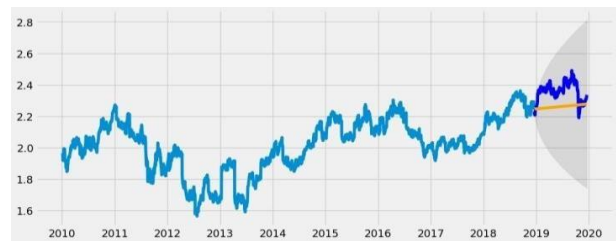

Fig. 5. Prediction using ARIMA

Fig. 6. Forecasting using ARIMA

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.

Acknowledgement

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, 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 (2017), 978-1-5386-0497-7/17
[3]. Yaojun Wang, Yaoqing Wang. "Using Social Media Mining
[4]. Technology to Assist in Price Prediction of Stock Market" 2016 IEEE International Conference on Big Data Analysis (ICBDA)
[5]. Dev Shah, Haruna Isah, Farhana Zulkernine. "Predicting the Effects of News Sentiments on the Stock Market". 2018 IEEE International Conference on Big Data (Big Data)
[6]. 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.
[7]. Ashish Sharma, Dinesh Bhuriya and Upendra Singh. "Survey of Stock Market Prediction Using Machine Learning Approach". International Conference on Electronics, Communication and Aerospace Technology
[8]. ICECA 2017
[9]. Neethu M S and Rajasree R. "Sentiment Analysis in Twitter using Machine Learning Techniques". IEEE – 31661
[10]. Harpreet Kaur, Veenu Mangat and Nidhi. "A Survey of Sentiment Analysis techniques". 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)
[11]. Biswarup Nandi, Mousumi Ghanti, Souvik Paul. "Text Based Sentiment Analysis". IEEE International Conference on Inventive Computing and Informatics (ICICI 2017)
[12]. Harshali Patil, Mohammad Atique. "Sentiment Analysis for Social Media: A Survey". IEEE 2015 2nd International Conference on Information Science and Security (ICISS)
[13]. 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
[14]. Soheila Mehrmolaei and Mohammad Reza Keyvanpour. "Time series forecasting using improved ARIMA". 978-1-5090-2169-7/16
[15]. Debadrita Banerjee. "Forecasting of Indian Stock Market using Time- series ARIMA Model". 2014 2nd International Conference on Business and Information Management (ICBIM)
[16]. Ayodele A. Adebiyi., Aderemi O. Adewumi and 2014 UKSim- AMSS 16th International Conference on Computer Modelling and Simulation Charles K. Ayo. "Stock Price Prediction Using the ARIMA Model".
[17]. Sheikh mohammad idrees, m. Afshar alam, and parul
[18]. Agarwal. "A Prediction Approach for Stock Market Volatility Based on Time Series Data". ACCESS.2019.2895252
[19]. 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
[20]. Nabilah Filzah Mohd Radzuan, Zalinda Othman, Azuraliza Abu Bakar. "Uncertain Time Series in Weather Prediction". 4th International Conference on Electrical Engineering and Informatics (ICEEI 2013)
[21]. Ruben Lam. "Forecasting trends in the Healthcare sector". Procedia Computer Science
[22]. Sultana Rajia, Md. Sabiruzzaman, Md. Kamrul Islam, Md. Golam Hossain,
[23]. Pete E. Lestrel. "Trends and future of maternal and child health in Bangladesh". PLoS ONE 14(3): e0211875.
[24]. 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
[25]. 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