Stock Market Prediction and Analysis-A Review

Anita H. Kenchannavar* and S.F. Rodd**

ABSTRACT

Investing in stock market has been one of the most profitable options for people since long time as it yields high returns with less investment. Stock market investment has become a popular choice rather than buying gold or making deposits in bank where the returns may not be that promising. But it is associated with a greater risk because stock price movements of a particular company depend on various factors such as market position, world economy, country's economy, company's growth performance, etc. Hence, investing in the stock market requires a profound knowledge and a complete understanding of the company stocks, its performance, the global scenario, etc. for making safe investment decisions. In this context, many algorithms have been proposed and tools have been developed using which the stock market data has been analyzed and stock price movements have been predicted. This paper presents an exhaustive survey of the algorithms proposed for stock price analysis and predictions. The algorithms proposed include machine learning techniques, data mining methods, statistical analysis techniques, sentiment analysis of the user tweets available from Twitter sites, financial discussion forums and blogs.

Index Terms: data mining, machine learning, sentiment analysis, stock price movement prediction, time series analysis.

1. INTRODUCTION

Stock market investment is both an interesting and challenging task for the investors. It is interesting because, one develops expertise to predict and derives joy when predictions come true. It is challenging because, there are so many diverse parameters that decide the stock price and it also involves the risk of losing real money. Stock analysis is an active area of research and several authors have proposed models to analyze stocks and predict their price in future to help users in making safe decisions about investing in the shares. Stock price prediction can be done using different approaches such as fundamental analysis and technical analysis. Fundamental analysis for predicting the future price of a company's stock depends on the different parameters. An individual must gain knowledge about the company before investing and its performance such as its businesses, what is the current demand for its products and how the demand will be in next 3 to 5 years. A company's position in the market can be very well understood by looking at some of the performance indicators such as company's earnings, current valuation of company's shares, future earnings growth, debt status, etc. These indicators can be evaluated by calculating related parameters. To evaluate a company the investors need to know the company's earnings, its current valuation in the market, future earnings growth, debt status, etc. The earnings of a company can be known, by looking at its last 5 to 10 years performance, profits earned and losses incurred. There are various financial indicators listed in the balance sheet indicating a company's financial health and investment opportunities. In this paper, the research carried out using different techniques and methods such as sentiment analysis, artificial neural network, data mining, etc. for future stock price prediction has been presented.

* Department of Computer Science, Research Scholar, K.L.S's Gogte Institute of Technology, Belagavi, Karnataka, India, *Email: anitaken24@gmail.com*

** Department of Computer Science, Research Scholar, K.L.S's Gogte Institute of Technology, Belagavi, Karnataka, India, *Email: sfroddgit@git.edu*

The paper is organized as follows: Section 2 discusses the related work followed by the important conclusions drawn from these techniques for giving an insight into the future prediction of the stock market.

2. RELATED WORK

The following section gives an overview about the work carried out in this domain using different approaches such as sentiment analysis, data mining, statistical analysis, etc. addressing varied factors which affect the stock price movement.

The authors Desheng Dash et. al. in their research work[1], propose an approach for analyzing the online stock forums which integrates the sentiment analysis and Support Vector Machine (SVM) and Generalized AutoRegressive Conditional Heteroskedasticity (GARCH) modeling. The result analysis depict that the investor sentiments have a powerful effect particularly on the value stocks compared to growth stocks. Jasmina Smalovic et. al. [2] have performed sentiment analysis on discussions related to specific companies on Twitter sites and tried to correlate how they affect the company stock market price. To address this issue they have designed an active learning approach considering static Twitter data streams for training the SVM sentiment classifier to find out the best Twitter specific preprocessing setting. The authors have used Granger's causality test for static data analysis. The SVM classifier in this approach categorizes sentiments into 3 types – negative, positive and neutral. The results show that probability of positive sentiments in tweet streams about a specific company can be used as an indicator for predicting the changes in stock closing prices. In order to tackle the problem of growing financial markets [3] and availability of huge amount of complex data, the authors in their research work, have introduced a novel method called CIMAWA- Concept for the Imitation of the Mental Ability of Word Association,which integrates word associations and lexical resources for carrying out sentiment analysis for financial markets in German language.

The research paper [4] presented by authors Vaanchitha Kalyanaraman, et. al. discusses about analyzing the sentiments from news articles by creating a specialized dictionary of sentiments to predict their effect on the stock prices. The authors have obtained the dataset by using the Bing API to get hold of the news articles about an individual company. The dataset has been tested by applying two different machine learning algorithms namely the Gradient Descent and Normal Equation for predicting the stock price movement. By comparing the accuracy of both the algorithms, the results show that Gradient Descent provided 81.82% accuracy on comparison with actual stock prices as compared to the Normal Equation. In the research article [5], an approach has been proposed by the authors Masoud Makrehchi and group, which labels the social media text as big losses or gains based on events providing meaningful automated labels which can be easily quantified on the basis of returns from individual stocks or indices. They obtained important stock price movements and gathered appropriate contemporaneous, pre and post text from social media sources like twitter. For every tweet, they assigned a negative or a positive label and then by training the model on the data set they have predicted appropriate labels for future tweets. Finally the authors have considered the average of the net sentiments for each day and have shown that it has more predictive power to forecast the future price movement of the stock. The authors have come up with a successful trading strategy based on this system which resulted in significant returns as compared to other baseline methods. The trading strategy developed using this approach was able to beat S&P 500 by about 20% returns within the duration of 4 months.

An effort [6] has been made by the authors to help the investors deal with rising amount of information to make profitable investment resolutions. The authors have combined text mining and sentiments of the investors in a novel two-stage approach to predict the impact of the specific news about a company on its future stock price. They have stated that this particular combination of data mining and investor sentiments improves the forecasting results and also results in a higher accuracy using the financial related word lists

for sentiment analysis as compared to generic dictionary. The results obtained using the current approach indicate that better improvement in the prediction of the stock price can be attained with the articles expressing positive sentiments ("boom") as compared to articles expressing negative sentiments ("ruin"). In the research paper [7], the authors have recommended a model for predicting the companies' future stock price by extracting the characteristic features from the time series data and social networking sites and assess its performance. Experimental results indicate that the proposed approach outperforms the other baseline methods in terms of magnitude prediction measures such as Mean-Average Percentage Error (MAPE), Root Mean Squared Error (RMSE) and Moving Average Error (MAE) considered for three popular Japanese companies' stocks in US market.

From behavioral economics [8] it is seen that emotions play a major role in an individual's behavior and decision making. In this context, Johan Bollen, et. al. have tried to anayze as to how the public emotions or public mood can be an indicator for predicting the economical changes of a company. The authors have analyzed the textual content of daily Twitter feeds and have tried to correlate the public mood with the value of Dow Jones Industrial Average over a period of time. The analysis of the Twitter feeds and the prediction of the company's economic status were done using two available mood tracking tools – OpinionFinder and Google Profile of Mood States (GPOMS). The OpinionFinder analyzed the public sentiments by measuring positive vs negative while the GPOMS analyzed the Twitter feeds based on the 6 states of human mood - Calm, Alert, Sure, Vital, Kind, and Happy. The resulting mood time series from the tools was cross verified by comparing their ability to detect the people's response to Thanksgiving Day and Presidential elections in 2008. A Self-Organizing Fuzzy Neural Network and a Granger Causality test were then used to inspect the hypothesis that the public mood analysis as carried out by OpinionFinder and GPOMS time series can predict the changes in DJIA stock price closing values. From the experimental results it was found that by including specific mood dimensions the prediction accuracy of daily DJIA stock price movements could be improved to 86.7% and also the Mean-Average Percentage Error was reduced by more than 6%.

Jasmina Smailoviæ, et. al. [9] has tried to explore whether the public opinion data about the companies and their products available on Twitter feeds are a suitable source for predicting the closing stock price movements. The authors have used Granger causality test to show that polarity of sentiments (positive or negative) can be used for providing indications about the stock prices a few days in advance. Later the authors have adopted SVM approach to classify the Twitter feeds into positive, negative and neutral which resulted in improved prediction of the stock price movements. A method of mining text [10] opinions has been introduced by authors Yoosin Kim, et. al. to analyze Korean language news to predict rises and falls on the KOSPI (Korea Composite Price Index) by performing NLP on the textual content. The experiments show that this method can be used to understand the unstructured big-data and also reveal that news sentiments can be used for predicting the stock price fluctuations.Financial applications such as trading systems use common approaches like heuristic methods or evolutionary algorithms [11]. The authors Yao-Hsin Chou et. al. have presented a method to help the traders to determine the best time to sell or buy stocks to maximize the profit with low risks by applying a trading system using the quantum inspired Tabu search algorithm to find the optimal composition and combination of strategies. Experimental results show much better performance in terms of earning money than other approaches and outperforms the buy and hold method.

In the article [12], a new method for fuzzy forecasting based on two-factors second-order fuzzy-trend logical relationship groups and the probabilities of trends of fuzzy-trend logical relationships has been discussed. This method has been used to predict the TAIEX and NTD/USD exchange rates. The experimental results clearly indicate that the presented method outperforms the existing methods. The authors in their research work [13], propose a technique to predict the Taiwan Stock Exchange Capitalization Weighted Stock Index (TAIEX) using fuzzy time series and automatically generated weights of multiple factors. This

method generates the predicted value of the next trading day depending on the closing index and the final forecasted variation of TAIEX on the trading day. A new complex neurofuzzy AutoRegressive Integrated Moving Average (ARIMA) computing approach is presented in [14] for the problem of time-series forecasting the future stock price. This method integrates a Complex NeuroFuzzy System (CNFS) using Complex Fuzzy Sets (CFSs) and ARIMA models to form the proposed computing model, CNFS-ARIMA. A basic structure [15] for integrating the news automatically in stock trading strategies has been presented in 2014 for predicting stock market prices.

According to [16], the stock market traders deal with statistics, company analysis, news etc. that influence economy in real time. The purpose of this paper is to review the studies about the impact of stress on the decision making process and on biological stress parameters applied in the sensor design and also to compare different ways to measure stress using sensors and to suggest new directions in the use of sensor technology in stock markets. The stock market prediction without sentiment analysis has been done in [17] based on the web traffic extracted from financial online communities. The authors have collected financial dataset of SP500 stocks for about 8 years and the classifier has been trained using the features extracted from the web traffic for prediction purposes. A multi-agent system [18] that works on numerical, financial and economical data has been described in order to evaluate the company's position in the market, profitability, performance, future expectations during the company's evolution. The authors have developed a prototype for result validation.

3. CONCLUSION

Research in this direction so far has been concentrated on stock prediction using the parameters like opening and closing stock price for the day, high and low values of the index. The techniques employed in the prediction research are sentiment analysis, data mining, statistical analysis, time series analysis, different machine learning algorithms such as SVM, Back Propagation algorithm, Artificial Neural Network, etc. Not much research on the stock prediction has been carried out using the Fundamental Analysis of financial parameters like Sales growth, Return On Capital Expenditure (ROCE), topline and bottom line margins. No empirical formula in the form of equation relating the company's financial performance parameters exists for predicting the direction of stock price movement for Indian stock market across different business verticals. Financial parameter ranking of major business verticals in the Indian stock market has not been done to help investors in predicting the stock price of a particular company in a given sector. These are some of the issues which need to be addressed.

REFERENCES

[1] Desheng Dash Wu, Lijuan Zheng, and David L. Olson: Decision Support Approach for Online Stock Forum Sentiment Analysis, IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 44, No. 8, pp. 1077-1087, (2014).

[2] Jasmina Smailovic´, Miha Grc¡ar, Nada Lavrac¡, Martin Z¡nidaršic¡: Stream-based active learning for sentiment analysis in the financial domain, Journal on Information Sciences, pp. 181-203, (2014).

[3] Patrick Uhr, Johannes Zenkert and Madjid Fathi: Sentiment analysis in Financial Markets – A Framework to utilize the human ability of Word Association for analyzing Stock market news reports, IEEE International Conference on Systems, Man and Cybernetics, pp. 912-917, (October 5-8 2014), San Diego, CA, USA.

[4] Vaanchitha Kalyanaraman, Sarah Kazi, Rohan Tondulalkar and Sangeeta Oswal: Sentiment analysis on News Articles for Stocks, 8th Asia Modelling Symposium 2014, IEEE Computer Society, pp. 10-15, (2014).

[5] Masoud Makrehchi, Sameena Shah and Wenhui Lao: Stock Prediction using Event-based Sentiment Analysis, IEEE/WC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology, IEEE Computer Society, pp. 337-342, (2013).

[6] Michael Siering: Boom or Ruin – Does it make a difference? Using Text mining and Sentiment Analysis to Support Intraday Investment Decisions, 45th Hawaii International Conference on System Sciences, IEEE Computer Society, pp. 1050-1059, (2012).

[7] Shangkun DENG, Takashi MITSUBUCHI, Kei SHIODA, Tatsuro SHIMADA and Akito SAKURAI: Combining Technical Analysis with Sentiment Analysis for Stock Prediction, IEEE Ninth International Conference on Dependable Autonomic and Secure Computing, IEEE Computer Society, pp. 800-807, (2011).

[8] Johan Bollena, Huina Maoa and Xiaojun Zeng: Twitter mood predicts the stock market, Journal of Computational Science, pp. 1-8, (Feb 2014).

[9] Jasmina Smailoviæ, Miha Grèar, Nada Lavraè, and Martin Žnidaršiè: Predictive Sentiment Analysis of Tweets: A Stock Market Application.

[10] Yoosin Kim, Seung Ryul Jeong and Imran Ghani: Text Opinion Mining to Analyze News for Stock market Prediction, International Journal Advance Soft Comput. Appl., vol. 6, No. 1, ISSN 2074-8523 (March 2014).

[11] Yao-Hsin Chou, Shu-Yu Kuo, Chi-Yuan Chen and Han-Chieh Chao: A Rule-Based Dynamic Decision-Making Stock Trading System Based on Quantum-Inspired Tabu Search Algorithm, IEEE Access, pp. 883-896, (2014).

[12] Shyi-Ming Chen and Shen-Wen Chen: Fuzzy Forecasting Based on Two-Factors Second-Order Fuzzy-Trend Logical Relationship Groups and the Probabilities of Trends of Fuzzy Logical Relationships, IEEE Transactions on Cybernetics, Vol. 45, No. 3, pp. 405-416, (March 2015).

[13] Shyi-Ming Chen, Fellow, IEEE, Huai-Ping Chu, and Tian-Wei Sheu; TAIEX Forecasting Using Fuzzy Time Series and Automatically Generated Weights of Multiple Factors, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART A: SYSTEMS AND HUMANS, vol. 42, NO. 6, pp. 1485-1495, (NOVEMBER 2012) .

[14] Chunshien Li and Tai-Wei Chiang: Complex Neurofuzzy ARIMA Forecasting—A New Approach Using Complex Fuzzy Sets, IEEE TRANSACTIONS ON FUZZY SYSTEMS, vol. 21, NO. 3, pp. 567-584, (JUNE 2013).

[15] Wijnand Nuij, Viorel Milea, Frederik Hogenboom, Flavius Frasincar, and Uzay Kaymak: An Automated Framework for Incorporating News into Stock Trading Strategies, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, vol. 26, NO. 4, pp. 823-835, (APRIL 2014).

[16] Javier Mart´ýnez Fern´andez, Juan Carlos Augusto, Member, IEEE and Ralf Seep: A Sensor Technology Survey for a Stress-Aware Trading Process, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART C: APPLICATIONS AND REVIEWS, vol. 42, NO. 6, pp. 809-824, (NOVEMBER 2012).

[17] Pierpaolo Dondio: Stock Market Prediction without Sentiment analysis: Using a Web-traffic based classifier and User-level Analysis, 46th Hawaii International Conference on System Sciences, IEEE Computer Society, pp. 3137-3146, (2013).

[18] Monica Tirea and Viorel Negru: Intelligent Stock Market Analysis System- A Fundamental and Macro-economical Analysis Approach, 16th International Symposium and Numeric Algorithms for Scientific Computing, pp. 519-526, (2014).