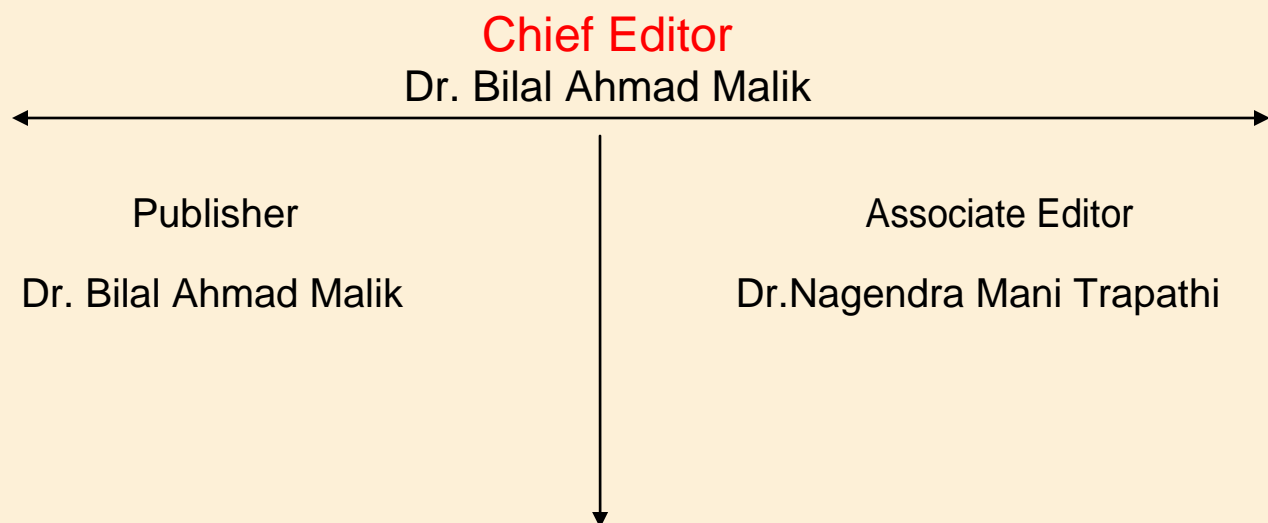


North Asian International Research Journal Consortium

North Asian International Research Journal

Of

Science, Engineering and Information Technology



NAIRJC JOURNAL PUBLICATION

North Asian
International
Research Journal Consortium



Welcome to NAIRJC

ISSN NO: 2454 -7514

North Asian International Research Journal of Science, Engineering & Information Technology is a research journal, published monthly in English, Hindi. All research papers submitted to the journal will be double-blind peer reviewed referred by members of the editorial board. Readers will include investigator in Universities, Research Institutes Government and Industry with research interest in the general subjects

Editorial Board

M.C.P. Singh Head Information Technology Dr C.V. Rama University	S.P. Singh Department of Botany B.H.U. Varanasi.	A. K. M. Abdul Hakim Dept. of Materials and Metallurgical Engineering, BUET, Dhaka
Abdullah Khan Department of Chemical Engineering & Technology University of the Punjab	Vinay Kumar Department of Physics Shri Mata Vaishno Devi University Jammu	Rajpal Choudhary Dept. Govt. Engg. College Bikaner Rajasthan
Zia ur Rehman Department of Pharmacy PCTE Institute of Pharmacy Ludhiana, Punjab	Rani Devi Department of Physics University of Jammu	Moinuddin Khan Dept. of Botany Singhaniya University Rajasthan.
Manish Mishra Dept. of Engg, United College Ald.UPTU Lucknow	Ishfaq Hussain Dept. of Computer Science IUST, Kashmir	Ravi Kumar Pandey Director, H.I.M.T, Allahabad
Tihar Pandit Dept. of Environmental Science, University of Kashmir.	Abd El-Aleem Saad Soliman Desoky Dept of Plant Protection, Faculty of Agriculture, Sohag University, Egypt	M.N. Singh Director School of Science UPRTOU Allahabad
Mushtaq Ahmad Dept.of Mathematics Central University of Kashmir	Nisar Hussain Dept. of Medicine A.I. Medical College (U.P) Kanpur University	M.Abdur Razzak Dept. of Electrical & Electronic Engg. I.U Bangladesh

Address: -North Asian International Research Journal Consortium (NAIRJC) 221 Gangoo, Pulwama, Jammu and Kashmir, India - 192301, Cell: 09086405302, 09906662570, Ph. No: 01933-212815, Email: nairjc5@gmail.com, nairjc@nairjc.com, info@nairjc.com Website: www.nairjc.com

STOCK MARKET CLOSING VALUE PREDICTION USING BUSINESS INTELLIGENCE

PROF. SHYAM INGLE ¹, MR. SURAJ TELI ², MR. AKSHAY KARVANDE ³, MR. TEJAS KADAM ⁴ &
MR. ANIKET JAGTAP ⁵

¹Prof. Department of Computer Engg., Nutan Maharashtra Institute of Engineering & Technology, Pune, India
²³⁴⁵Students, Department of Computer Engg., Nutan Maharashtra Institute of Engineering & Technology,
Pune, India

ABSTRACT

Economy is the backbone of any developing country such as India. Financial & stock market are backbone of this economy. Stock market prices are frequently changes several time during trading day. This change in stock market can be due to internal as well as external factors. Many studies have predicted the closing values of market using internal factors but external factors are also responsible for market fluctuation. In this project we will analyse the market closing value with the help of external factors. so that investors can predict & plan their future investment opportunities.

Two methods are used in this which are supervised and unsupervised study. The results obtained have shown that the model can predict the closing price using the classification algorithm and regression algorithms with accuracy greater than 91% and that the regression algorithm used in predicting the stock prices with a correlation coefficient equal to 0.865

Keywords: *Big data, Business Intelligence, Natural Language Processing, Regression Analysis, Voting Feature Intervals, Genetic Algorithm Comparison Survey Security and Reliability, dividend yield.*

INTRODUCTION

Economy is the key factor in the development of country. Financial policy and the stock market plays main role in economy base of country. Financial and stock markets are the driving-force for economic growth and stability as they significantly contribute to the flow of resources which leads to productive investment opportunities. Stock and securities prices frequently change in stock markets. During particular times of the year, it is quite common to notice that stock prices are frequently appreciated every morning, and this may take place several times during a given day for some stocks. This indicates that stock prices are affected by many factors.

There is no perfect system existed that indicates the exact fluctuation of the stock prices. However, the factors behind increases or decreases in the demand and/or supply of a particular stock could be due to company's fundamentals, internal or external factors, and other market behaviors.

There are ample research work investigated the factors affecting stock prices. Several studies have examined the association of the stock prices and selected internal or external factors. However, it is quite difficult to pinpoint to a particular factor due to the various conditions and scenarios that surround each stock market environment. Different economy or market has its own policy, procedures, rules and regulations, country location, type of the investors, and other unique factors.

The issues of data quality in predicting stock prices have an important role in improving the accuracy of the predicted class. The main factors that comprise data quality are accuracy, completeness, consistency, timeliness, believability and interpretability. Inaccurate, incomplete and inconsistent data are common-place properties of financial data.

There is a general agreement that the basic internal factors influencing stock prices could be due to company performance, a change in management or ownership, the creation of new assets, earnings, dividends, etc. On the other hand, some of the external factors include government policy, rules and regulations, economy stability, inflation, and other economic conditions, such as Gross Domestic Product (GDP), money supply, oil/gas/gold prices or environmental circumstances which could have direct links with the production of the company. Some other internal factors also include, earning per share, dividend per share, book values. However, the most important internal factors remain to be the earning per share (EPS), dividend yield (DY), and dividend per share (DPS).

Improve the predictive accuracy of the classifiers. Regression analysis is widely used in statistics to analyze data. It provides a conceptual method to investigate functional relationships among variables. The standard regression analysis approach takes the data, fit a model and then evaluate the fit using statistics .in this we applied multiple regression analysis and business intelligence techniques using natural language processing to define the economic and financial variables, which have a strong relationship with the output result.

The motivation behind this study is to assist investors in the Financial Market, "FM," regarding the closing prices of stocks and when to buy or sell stocks based on the stock predictions.

Time series algorithms require at least fifty, or preferably one hundred and more, sets of historical data to yield the desired results. ANNs have some limitations in learning the patterns because stock market data has tremendous noise and complex dimensionality, and ANNs often exhibit inconsistent and unpredictable performance on noisy data. The fuzzy-logic algorithm also has some limitations, as most of the parameters are defaulted and cannot be changed, which is a serious limitation, and the algorithm fails to handle the issue of missing data.

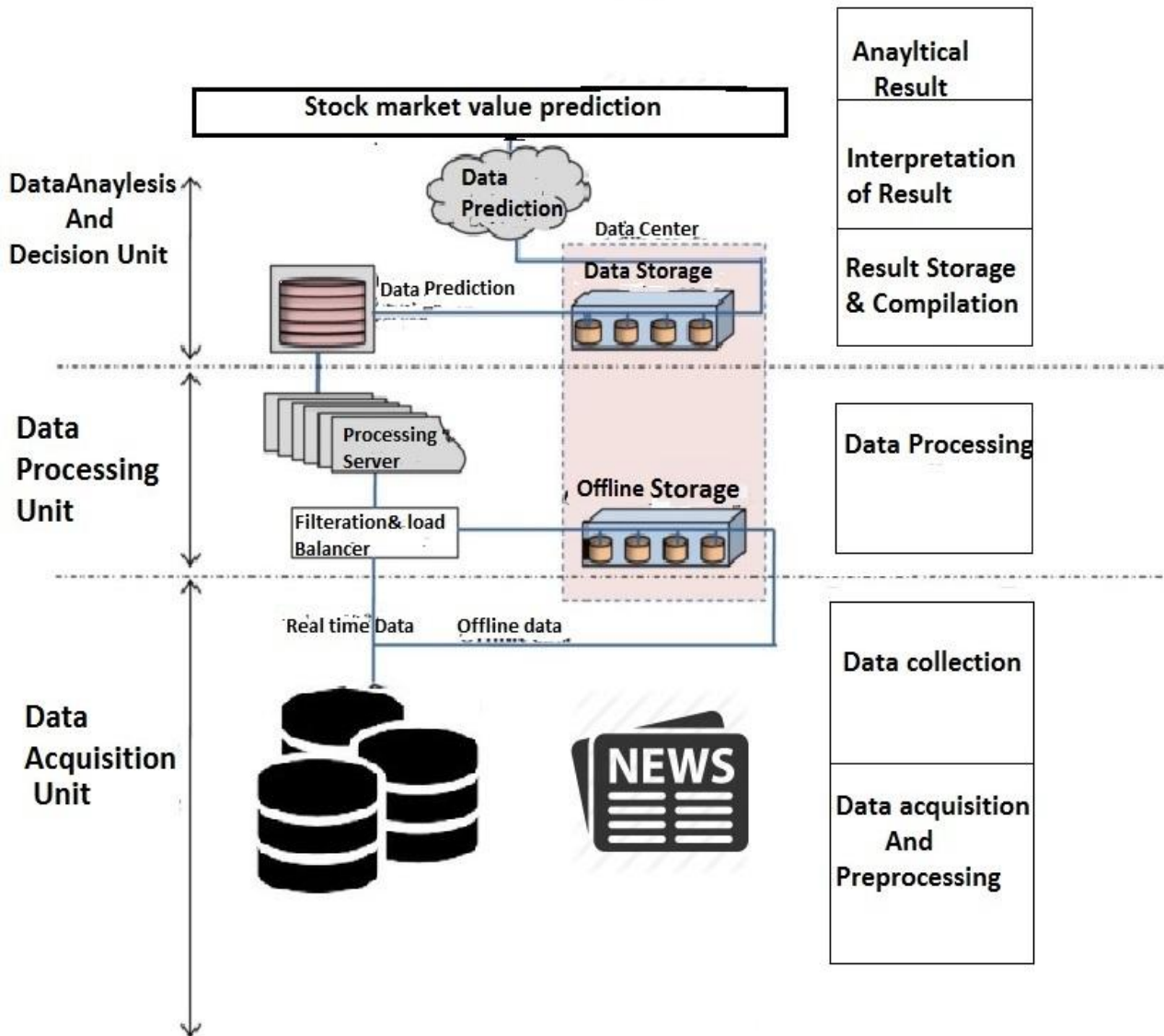
Due to the mentioned limitations of time series, ANN and fuzzy-logic algorithms, our objective in this study is to experiment with classifications and regression analysis along with the help of natural language processing in predicting stock market prices. We also aim to deal with small amount of data for building the learning models and identify factors and attributes potentially affecting the prediction of stock market prices.

RELATED WORK

Lee et have proposed a system that forecasts companies' stock price changes in response to financial events. They reported that predicting next day's price movement was improved by 10% if text is considered. Ichinose and Shimada have proposed a combination method with a filtering and a machine learning technique for a stock market prediction task. They reported that the filtering method improved the classification accuracy. These studies handled financial documents and web news about financial topics. On the other hand, some researchers focused on Social Medias, such as Message boards and Twitter. Nguyen and Shirai have proposed a stock market prediction model with a topic-sentiment feature and a new topic model. By using the feature and topic model, the accuracy of the proposed method increased by 6% as compared with some baselines. Bollen et al. have proposed a method with 6 mood dimensions on Twitter for the stock market prediction. They obtained the high accuracy rate; 86.7%. These studied only focused on the accuracy rates of the classification models. However, there was no mention whether the improvement of the accuracies contributed to making a profit. One of the most important points for personal investors is whether they eventually gain a profit by using the prediction model. Izumi et al. have proposed a text-mining method for long-term market analysis using market reports published by financial professionals and institutions on the web. For Japanese government bond 2-year, 5-year, and 10-year markets, the proposal method forecast in higher accuracy about both the level and direction of long-term market trends. The task was forecasting long-term market trends of Japanese government bond. On the other hand, the target and resources of our task are the Nikkei Stock Average and news articles on the Web. Schumaker and Chen have

proposed a SVR-based method that made a discrete prediction of what +20 minute stock should be. The method, AZ Fin Text, used financial news articles and stock price quotes. They compared AZ Fin Text against the top 10 quants for one year. As a result, AZ Fin Text outperformed six of the top 10 quant funds. They considered the effectiveness of the method in terms of real profit-and loss. However, they evaluated the investment performance at the end of period. We also discuss a criterion based on property changes during the testing.

SYSTEM ARCHITECTURE



Stock market value prediction system architecture

1. Data acquisition and pre-processing

In this system there are two places from which data can be gathered from nse india and rss feed nse india is to extract historical data and from rss feed real time data is extracted data obtained from rss feed is categorical and to perform operations on it we used natural language processing for the processing of data from rss feed to calculate the impact of the each news on the stock market value.

In this stage unwanted data is removed using setting filter limits.

2. Data collection

Data from nse india can be downloaded easily through there site and the data of rss feed is categorical and can be collected by calling object of the data which is present at rss feed class. Our method consists of two part; (1) one-day classifiers and (2) strategies for trading. The purpose of one-day classifiers is to classify the next day into “raising” or “dropping” by using the articles of a day. We explain two strategies for trading after generating one-day classifiers. *A. Classifier* We apply machine learning techniques to the one-day classifiers; Support Vector Machines (SVM) and linear perceptron. For SVMs, we use *SVMLight2* by as the implementation. We introduce three types of features to the methods:

BOW: simple Bag-of-words with binary features

VOL: Bag-of-words that are weighted by the absolute value of a volatility score ($V S$)

AVS: Bag-of-words that are weighted by the actual value of the $V S$

We use nouns, verbs and adjectives as the BOW features. The volatility score ($V S$) denotes stock market volatility in each day. It is computed as follows: $V S = hp - lp(1)$ where hp and lp are the highest and the lowest prices in the day. We combine the two methods and the three features. As a result, we compare six classifiers (SVM BOW, SVM VOL, SVM AV S, Perceptron BOW, Perceptron VOL and Perceptron AV S) in the experiment. The target value of each classifier is +1 (Raising in the next day) or -1 (Dropping in the next day). We generate a one-day classifier by using news on other days. For example, the one-day classifier to evaluate the stock market on January 5th is generated from data from January 6th to December 13th. The one-day classifier for January 6th is also generated from data of January 5th and from January 7th to December 13th. *B. Trading settings and strategy* We evaluate our method with simulated trading using real stock prices. For the evaluation, we need to set a virtual market and a virtual investor with a strategy. • Virtual market: We use the past data on the Nikkei Stock Average.

Each datum of a day consists of the opening price, closing price, highest price, and lowest price. • **Virtual investor:** This is a classifier described in Section IV-A. We set the initial money of each investor to 2 million yen (1 million yen for Spot buying and 1 million yen for Short selling). The investor buys stocks if the classifier answers that the price of the next day will be raised and the investor sells stocks if the classifier answers that the price of the next day will be dropped. There are many strategies for the buying and selling. We apply two strategies in this paper. The strategy *S1* is as follows: • **Buying:** The investor buys stocks by using all his/her money in the opening. Then, the investor sells all stocks in the closing.

• **Selling:** The investor sells stocks in the opening (short selling³). Then, the investor buys all stocks that sold in the opening in the closing⁴. The strategy *S2* is as follows:

• **Buying:** The investor waits until the stock price becomes lower than the closing price of the previous day⁵, and then buys stocks by using all his/her money. If the day does not meet this requirement, the investor does not trade stocks at the day. Finally, the investor sells all stocks in the closing.

• **Selling:** The investor waits until the stock price becomes higher than the closing price of the previous day, and then the investor performs short selling. If the day does not meet this requirement, the investor does not trade stocks at the day. Finally, the investor pays back the borrowed stocks to the securities company in the closing.

ALGORITHMS

Algorithm I. Data Acquisition and Filtration Algorithm

Input: Live Data Feed process data set

Output: filtered data in key value pair and send this to processing Mechanism

Steps:

1. Filter related data i.e. Processed data. All other unnecessary data will be discarded.
2. Divide the Data into Appropriate Key Value Pair.
3. Transmit Unprocessed data directly to aggregation step without processing.
4. Assign and transmit each distinct data block of Processed data to various processing steps in Data Processing Unit.

Description: This algorithm takes live RSS Feed Data and then filters and divides them into segments and performs approximation algorithm.

In step 1, related details filtered out.

In step 2, filtered data are the association of different key value pairs and each pair is different numbers of sample, which results in forming a data block. In Next steps , these blocks are forwarded to processed by Data Processing Unit.

Algorithm II. Processing and Calculation Algorithm

Input: Filtered Data

Output: Normalized News data into Numerical comparable form Along with Historical Values.

Steps:

1. For each event data, relevant Historical Data is extracted.
2. Normalize this for all the live feed.
3. Persist the data into data store and forward it.

APPLICATION

- E-commerce.
- Banking.
- Defining company policy.
- Product design policy.
- Enrich the value prediction method.
- Access to all users.
- Financial budget .

EXPECTED RESULT

At end, result of the news is analyzed and decision will be taken on whether each result is correct or not a correct. Such result is helpful to users for making their respective decisions while investing money in stock market.

FUTURE SCOPE

Several researches and discussions are going on across the world among technologists, researchers, academicians, vendors, operators, and governments about the innovations, implementation, viability, and security concerns of stock market and stock market prediction. Since there is possibility of predicting stock market results on the basis of closing value with the help of the classification and regression algorithms. This system enhances the chances of accuracy while predicting the stock market value on the basis of natural language processing and the business intelligence. This is designed to provide incredible and remarkable data capabilities, unhindered call volumes, and immeasurable data processing within the latest parallel processing system. Hence, it is more intelligent technology, which will interconnect the entire world stock limits and values without limits. Likewise, our world would have universal and uninterrupted access to stock market value information that will open a new dimension to our lives and will change our life style meaningfully.

Moreover, governments, companies, investors and regulators can use this technology as an opportunity for the good governance and can create healthier environments, which will definitely encourage continuing investment in stock market.

CONCLUSION

In this paper, the system acts as stock market value predictor. This study involves analyzing the market data closing process using two types of data mining analysis techniques, which are supervised classification algorithms and unsupervised regression algorithms. This study builds a model for predicting the closing stock prices for the companies. In addition, this study assists investors in predicting the closing pricing in the future.

This proposal is still open for future work to improve the prediction accuracy; that might be achieved by continuous data sampling, monthly or daily, and testing the data against the model accuracy might also be enhanced by using more classifiers in the classification algorithm.

ACKNOWLEDGEMENT

We take this special opportunity to express our sincere gratitude towards our team members and all the people who supported us during our project work. We will like to express our gratitude to our guide and HOD **Prof. Shyam Ingle** and also the project coordinator **Prof. Ashvini Jadhav** for providing us special guidance. We

would also like to thank our who always has enough time to solve everyone's problems at hour of the day. Finally thanks to all our teachers who are always supportive at us.

REFERENCES

- [1] J. Bollen, H. Mao, and X.-J.Zeng, "Twitter mood predicts the stock market," *Journal of Computational Science*, vol. 2, no. 1, pp. 1–8, 2011.
- [2] H. Lee, M. Surdeanu, B. MacCartney, and D. Jurafsky, "On the importance of text analysis for stock price prediction," in *Proceedings of the 9th edition of the Language Resources and Evaluation Conference(LREC)*, 2014, pp. 1170–1175.
- [3] T. H. Nguyen and K. Shirai, "Topic modeling based sentiment analysis on social media for stock market prediction," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing*, 2015, pp. 1354–1364.
- [4] Y. Tsuji, K. Komiya, and Y. Kotani, "Stock price prediction corresponding to more than one companies on the web news," in *IEICE technical report SIG-IBISML SIG-NLC (in Japanese)*, vol. 110, no. 476, 2011, pp. 109–113.
- [5] K. Ichinose and K. Shimada, "Stock market prediction from web news using filtering and machine learning," in *IEICE technical report SIGNLC(in Japanese)*, vol. 115, no. 70, 2015, pp. 91–96.
- [6] K. Izumi, T. Goto, and T. Matsui, "Trading tests of long-term market forecast by text mining," in *Proceedings of 2010 IEEE International Conference on Data Mining Workshops (ICDMW)*, 2010, pp. 935–942.
- [7] R. P. Schumaker and H. Chen, "A discrete stock price prediction engine based on financial news," *Computer*, vol. 43, no. 1, pp. 51–56, 2010.
- [8] V. N. Vapnik, *Statistical Learning Theory*. Wiley, 1999.
- [9] F. Rosenblatt, "The perceptron: a probabilistic model for information storage and organization in the brain," *Psychological Review*, vol. 65, pp. 386–408, 1958.
- [10] T. Joachims, *Learning to Classify Text Using Support Vector Machines – Methods, Theory, and Algorithms*. Kluwer/Springer, 2002.

Publish Research Article

Dear Sir/Mam,

We invite unpublished Research Paper, Summary of Research Project, Theses, Books and Book Review for publication.

**Address:- North Asian International Research Journal Consortium (NAIRJC)
221, Gangoo Pulwama - 192301**

Jammu & Kashmir, India

Cell: 09086405302, 09906662570,

Ph No: 01933212815

Email:- nairjc5@gmail.com, nairjc@nairjc.com , info@nairjc.com

Website: www.nairjc.com

