

# Stock Market Prediction using Financial News Articles

Siddhanth M<sup>1</sup>, Shравan Bhat<sup>2</sup>, Sampath Kumar<sup>3</sup>

<sup>1,2,3</sup>Under the guidance of: Dr. Rekha B Venkatapur, Head of Department of Computer Science, K.S Institute of Technology, Karnataka, India

\*\*\*

**Abstract** - Stock market prediction is one of the most unpredictable markets and hence very difficult to predict. Our system first extracts the data from trusted news sources. Then, the extracted text is cleaned using natural language processing libraries and then sentiment analysis is applied on this text to get the polarity of the news (positive or negative). After determining the polarity, it is then given as input to the machine learning model which applies an efficient algorithm to predict the price of the stock in the future (for the next 10 minutes or so).

**Key W Keywords:** Stock Market prediction, Financial News, Machine Learning, Natural language processing, Sentiment Analysis, Data Mining, Stock price movement

## I. INTRODUCTION

There are a lot of online sources that publish financial news on the Internet to help investors for shaping their investments. Both, current and historical news about companies, economic and political events are available on these sources. Availability of this huge amount of financial data in digital media creates appropriate conditions for a data mining research.

The decision of when to buy or sell shares is an interesting research challenge in the stock market. Such decisions are being negotiated daily in the stock market across the globe. Indices were created as a way to measure the relative value of stocks in a determined group in the market.

The idea about whether stock markets can be predicted has kept economists and investors very busy for decades. The two distinct trading philosophies for stock market prediction are fundamental and technical analysis. While technical analysis focuses on the study of market action through the use of charts, fundamental analysis concentrates on the economic forces of supply and demand that cause the stock price to move higher, lower, or remain the same.

Stock price prediction has attracted many researchers in multiple disciplines such as computer science, statistics, economics, finance, and operations research. These research efforts have so far produced several methods for forecasting the future direction of the stock market.

In this study, we propose a system predicting stock price movements by first extracting the data from trusted news sources. Then, the extracted text is cleaned using natural language processing libraries and then sentiment analysis is applied on this text to get the polarity of the news (positive or negative). After determining the polarity, it is then given as input to the machine learning model which applies an efficient algorithm to predict the price of the stock in the future (for the next 10 minutes or so).

There are a lot of substantial work done on prediction of stock prices. These works are basically text categorization systems targeting to predict stock price movement by classifying financial news articles as positive or negative. Since the problem is converted to a text categorization problem, several feature selection and classification methods are used in these works. In term frequency – inverse document frequency technique is used as a feature selection method. This relatively low success rates are caused by nature of stock price movements, which are the result of decisions of investors, since it is hard to predict human behavior.

## II. Background

Traditional technical analysts have developed many indices and sequential analytical methods that may reflect the trends in the movements of the stock price. However, in addition to historical prices, the societal mood is seen to be playing a significant role in stock price movement. The overall social mood with respect to a given company is now being considered as an important variable which affects the stock price of that company. Online information such as public opinions, social media discussions, and Wikipedia activities are being used to study their impact on pricing and investors' behaviors taking into account the risk factors due to biased and malicious posts. Researchers have shown that integrating models based on historical stock prices with the data from the mainstream and social media platforms can improve the predictive ability of the analytical models of stock price prediction.

Sentiment analysis is an NLP technique for mining and evaluating public opinions expressed in text/speech. The increasing interest in sentiment analysis is partly due to its potential applications. Major elements of sentiment analysis methods include the sentiment (opinion expressed), target object or topic, time of expression, person expressing the opinion, the reason for the expression, the opinion qualifier, and the opinion types.

Sentiment analysis is divided into several levels according to different discourse granularity, such as document, sentence, entity, and aspect. It can also be implemented in a supervised, unsupervised, or semi-supervised approach. This paper adopts a dictionary-based learning approach at the entity level to analyze the effects of sentiments expressed in news articles on stocks. The bag of words for this system is around 5000 which may increase based on the amount of raw data.

### III. SYSTEM TRAINING AND TESTING

The dataset is divided into 70-30 format where 70% is used for the training and the remaining 30% is used for testing.

The system proposed by us can be divided into 4 phases:

- 1) The user is given the choice of adding the news sources he/she trusts.
- 2) The news sources, after being selected, are then used to extract the headlines using text mining libraries like BeautifulSoup4. This library is written in Python and hence provides efficient text mining and extraction of the news articles.
- 3) After data is mined from the news sources, which are trusted by the user, it is then passed to the natural language processing phase.
- 4) The natural language processing phases mainly involves cleaning the news articles to extract the keywords which will then be verified with our bag of words. Those words which match the bag of words are then passed on to the sentiment analysis phase.
- 5) The sentiment analysis phase is basically employed by popular libraries which determine the polarity of the cleaned words from the previous phase. The libraries that we have used are spaCy and Vader\_Text and TextBlob. These libraries are then optimized to our needs which maximises the consistency of the polarity of the news and gives us the best results.
- 6) After the polarity is determined, it is then passed to the machine learning model which tries to predict whether the price of the specific company's shares of stock has increased or not.
- 7) The machine learning model or the ML model uses an algorithm which can efficiently determine whether the price of the stock increases or decreases in real-time. This is done by querying the news API to get the news and then it is preprocessed before it is passed to the ML model.

- 8) The ML model uses, in this case, Linear Regression model to plot all the price points of the stocks and then checks whether there is a similar pattern in the price movements before and if there exists a similar pattern then that pattern is correlated to predict the price points of the stock.
- 9) After the model predicts the price points of the stocks with the training data, we then calculate the accuracy of the model for that training data.
- 10) Finally, after training the model with satisfactory results, we moved on to test for real-time data and predicted the accuracy of the entire system. To calculate the accuracy, we have used the statistics formulae of accuracy specificity and precision and the output of the accuracy of the entire system is displayed.

### FLOW DIAGRAM



Chart -1: Overall system flow diagram

### IV. CONCLUSION

This study is an application of sentiment analysis in gauging the effects of news sentiments on stocks for the pharmaceutical market sector. Our system automatically analyzes and classifies news articles and generates recommendations for investors. The sentiment scores obtained from the analysis of the news articles is a powerful indicator of stock movements and can be used to effectively leverage the prediction of short-term trends. We believe that the model created for this particular sector by researching and leveraging domain expertise will be helpful in further research and analysis

### ACKNOWLEDGEMENT

We would like to thank our mentor, family and our friends for helping us through the project.

### REFERENCES

[1] Dev Shah, Haruna Shah, Farhana Zulkernine, "Predicting the effects of news sentiments on the

stock market,” 2018 IEEE conference on Big Data(Big Data), ISBN:978-1-5386-5035-6/18 .

- [2] Yasef Kaya, M. Elif Karsligil, “Stock price prediction using financial news articles”, 2010 IEEE , ISBN: 978-1-4244-6928-4/10.
- [3] HD Huynh, LM Dang, D Duong, “A New model for stock price movements prediction using Deep Neural Network”, SoICT, 2017, pp.57-62: ACM
- [4] Cicil Fonseka, Liwan Liyanage, “A data mining algorithm to analyse stock market data using lagged correlation”, 2008 IEEE, ISBN:978-1-4244-4/08.
- [5] Pegah Falinouss, “Stock trend prediction using news articles”,2007,ISSN:1653-0187 ISRN:LTU-PB-EX--07/071--SE
- [6] Pengyu Wei, Ning Wang, “Wikipedia usage pattern help to predict the individual stock movement”,ISBN:978-1-4503-2138-9
- [7] Kroha, P.Kroeber, K Janetzko, “Case study: Classification of stock exchange news by Support Vector Machines”, 2007