Stock Market Prediction Using News Feed and Historical Data

Prof. Kimi Ramteke[1], Pranay Motghare[2], Armin Bodhanwala[3], Shreya Vinchankar[4], Manisha Sagar[5]

[1] *Professor, Computer Engineering, International Institute of Information Technology Pune, Maharashtra, India*
[2] *Student, Computer Engineering, International Institute of Information Technology Pune, Maharashtra, India*
[3] *Student, Computer Engineering, International Institute of Information Technology Pune, Maharashtra, India*
[4] *Student, Computer Engineering, International Institute of Information Technology Pune, Maharashtra, India*
[5] *Student, Computer Engineering, International Institute of Information Technology Pune, Maharashtra, India*

ABSTRACT

The aspiration of any investor is to forecast the market behavior with the aim of making the best decision when it comes to buying or selling shares of stocks seeking to maximize his profits. Some bond traders, who have a deep understanding of the stock market, utilize the actual stock data, stock market information, and survey data from all around to forecast the future development of the stock market direction, which can be combined with the historical data of the stock market and the recent development. However, the trend of the stock market has the characteristics of nonlinear fluctuations, and now the scientific method has not been able to fully achieve the accurate prediction of the stock movements. Stock index value prediction system may predict the stock value based on historical data and news feed.

Keyword: - *Data Mining, Machine Learning, Prediction, NLP, Stock Market, Time Series Analysis*

1. INTRODUCTION

Stock index prediction has been an important issue in the field of finance due to its potential financial gain. As a high amount is traded through the stock market, the stock-market is seen as a field of high investment outlet. Researchers are working hard to prove the predictability of the financial market. Hence, Stock Index prediction has always been researchers' best interest. Many scientific attempts have been made but no method has been discovered to accurately predict stock price movement.

The main goal of our project is to predict accurate stock values for the next few days or months.
These predictions will be made based on the available historical data as well as stock related news articles and blogs. When these two are combined together to make predictions, accuracy would be comparatively higher. The algorithms which we are considering to use in this project are time series analysis using Holt Winters for historical data, Latent Dirichlet Allocation for news articles, blogs and Akaike Information Criterion for computing accuracy of the model. Along with this, ensemble modeling would be used to combine both the prediction algorithms.

2. LITERATURE REVIEW

Many research groups are exploring stock market trend prediction using social media analytics. SENSEX's Index dataset has been used for time series forecasting using Rapid miner in [01]. The model is designed in rapid miner to predict the time series data. The research work is done to predict the SENSEX index in upcoming days.

But, time series analysis alone with rapid miner gives accuracy of around 47-49% only. Also, consideration of external factors such as investor sentiments, economic outlook, terrorist activity, economic policies and governance may affect the predictions. Reference [02] includes both technical and fundamental analysis. Technical analysis is done using historical data of stock prices by applying machine learning and fundamental analysis is done using social media data by applying sentiment analysis. Using social media data in addition to numeric data increases the quality of the input and gives improved predictions. The aid of big data technology allows predictions at real-time.

Sentiment analysis on news articles has also been used to see its effect on stock prices. Datasets are collected using Bing API for specific company. Specialized sentiment dictionary only meant to analyze stock articles is created. Gradient Descent provides an average accuracy of only 59.5%. Results takes into consideration the change in stock value over the entire time period. Mapping the analysis of the sentiment of the news article to immediate changes in the market will be more useful [03]. Correlation between the sentiment score of event related tweets and stock price changes was observed in [04]. Keyword analysis was performed to show that the keywords mentioned in the tweet are more related to the event. Deep learning model in [05] combines a convolutional layer with a recurrent layer for stock price change prediction and uses technical indicators and news titles as inputs. The results of lexicon emotion analysis conducted on crowd annotated news to extract various types of public emotions from news articles are utilized. An emotion scoring model to predict public emotions based on news articles is proposed [06].

3. ARCHITECTURE DIAGRAM

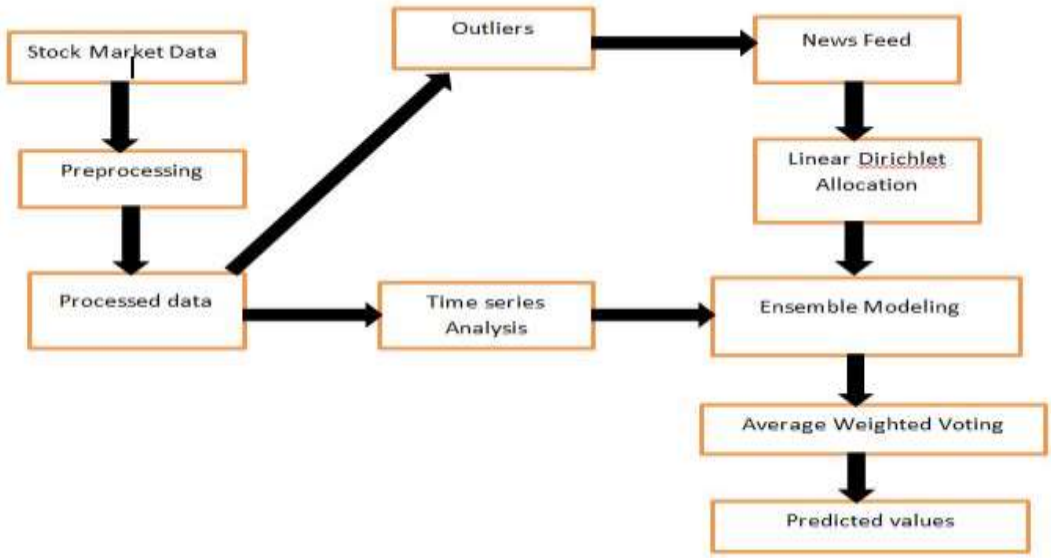

Fig 1: Architecture Diagram

4. PROPOSED SYSTEM

Some of the existing systems for stock market prediction do not consider external factors like terrorist activity, natural calamities, economic policies and governance . Sentiment analysis is done using summative assessment of the sentiments in a particular news article or tweet which could be improved for better predictions. The aim of our project is to predict stock value index accurately. For this, we are going to consider two different types of data: a) Historical Data. b) News Feed Articles and Blogs. Historical data would be fetched by taking previous stock values of different agencies into consideration. The collected data would be tokenized and normalized using different techniques of preprocessing. The processed data will then be analyzed to find similar patterns and will get classified. News feed can be obtained using twitter sentiment analysis. It can also be obtained from stock based news articles, blogs, column from magazines, etc. The major source of collecting this data for our project is Yahoo Finance and similar websites. Here, all the stock related information which keeps on changing

dynamically will be collected and will be processed further. An *outlier,* abnormal stock index value from other values, can rise because of certain events or environmental factors. The algorithms that can be used for Time Series Analysis(TSA) are Holt, Simple Exponential Smoothing and Holt Winters. Holt Winters considers both seasonal variations and trends. Latent Dirichlet Allocation(LDA) algorithm will be used for topic modeling of data which will be collected from news articles. Ensemble modeling will contribute in combining the data obtained from both the sources i.e. from TSA and LDA. Average Weighted voting - Predictions would be done based on the analyzed data. Different data mining algorithms can be used for prediction. The stock index value is predicted by combining both TSA and LDA. Risks include unpredictable fluctuations occurring in the stock index values. Thus, presence of outliers gives rise to risk. Additionally, Sentiment analysis is not always accurate. Natural calamities also affect the stock market indices.

5. CONCLUSIONS

This paper proposes a new approach in designing a system which can predict the values of stock index using combined analysis of historical stock record and news feed. The ensemble modeling would help in acquiring a high accuracy in the prediction model. Considering peoples' opinions in the form of news feed may improve overall efficiency of the system. A large percentage of population invests in stock market so, this system would make the predictions accurate to a maximum level, thus increasing profit of investors.

6. REFERENCES

[01] Udit Agarwal, A Sai Sabitha, Amity university, U.P,India, "Time Series Forecasting of Stock Market Index ", 2016 IEEE.

[02] Girija V Attigeri , Manohara Pai M M, Radhika M Pai, Aparna Nayak Manipal Institute of Technology, Manipal- 576104, India , "Stock Market Prediction: A Big Data Approach", 2015 IEEE.

[03] Vaanchitha Kalyanaraman and Sarah Kazi VESIT, Mumbai, India ,"Sentiment Analysis on News Articles for Stocks ", 2014 8th Asia Modelling Symposium.

[04] Satyabrata Aich, Hee-Cheol Kim, Mangal Sain, Bijay Bhaskar Deo Department of computer engineering, University, S. Korea,"Analyzing stock price changes using event related Twitter feeds", ICACT2017 February 19 .

[05] Manuel R. Vargas, Beatriz S. L. P. de Lima and Alexandre G. Evsukoff COPPE/Universidade Federal do Rio de Janeiro" Deep learning for stock market prediction from financial news articles", 2017 IEEE.

[06] Chayanin Wong, and In-Young Ko, "Predictive Power of Public Emotions as Extracted from Daily News Articles on the Movements of Stock Market Indices", IEEE/WIC/ACM International Conference on Web Intelligence 2016.

[07] Sornpon Wichaidit and Surin Kittitornkun, "Predicting SET50 stock prices using CARIMA(Cross Correlation ARIMA)", IEEE 2015.

[08] Phayung Meesad and Jiajia Li Faculty of Information Technology King Mongkut's University of Technology North Bangkok ,"Stock Trend Prediction Relying on Text Mining and Sentiment Analysis with Tweets ", IEEE International Conference on Information and Financial Engineering (ICIFE), 2016.

[09] K. Kim, "Financial time series forecasting using support vector machines," Neurocomputing, vol. 55, no. 1–2, Sep. 2003.

[10] Andrew Beattie, 10 Risks That Every Stock Faces, https://www.investopedia.com/articles/stocks/11/risks-every-stock-faces.asp

[11] Thiagogm, Computing and visualizing LDA in R, https://www.r-bloggers.com/computing-and-visualizing-lda-in-r/