

Analysis of Stock Market by Using Big Data Analytic Technology

Varunesh Nichante, Prof. Sulabha Patil

Abstract: Big data plays a serious role within the business for creating higher predictions over business information that is collected from the real world. Finance is that the new sector wherever the big data technologies like Hadoop, No SQL are creating its mark in predictions from financial data by the analysts. It's a lot of fascinating within the stock exchange choices which might predict on a lot of profits of stock exchange. For this stock exchange analysis each regular information and historical information of specific stock exchange are needed for creating predictions. There are varied techniques used for analyzing the unstructured information like stock exchange reviews (day-to-day information) and historical statistic of economic information severally. This paper involves discussion regarding the strategies that square measure used for analyzing each varieties of information. We also propose a system for stock prediction.

I. INTRODUCTION

The quick advance in computerized information securing has prompted the quickly developing measure of information put away in databases, information distribution centers, or different sorts of information stores. Albeit important data might be holding up behind the information, the mind-boggling information volume makes it troublesome for individuals to concentrate them without capable instruments. Simple and brisk accessibility to news data was unrealistic until the start of the most recent decade. In this period of data, news is currently effectively open, as substance suppliers and substance locators, for example, online news administrations have grown on the World Wide Web. Persistent accessibility of more news articles in computerized structure, the most recent advancements in Natural Language Processing (NLP) and the accessibility of speedier PCs lead to the inquiry how to concentrate more data out of news articles.

Budgetary investigators who put resources into securities exchanges for the most part don't know about the share trading system conduct. They are confronting the issue of stock exchanging as they don't know which stocks to purchase and which to offer keeping in mind the end goal to acquire benefits. Every one of these clients realize that the advancement of the share trading system depends a ton on important news and they need to arrangement day by day with unlimited measure of data. They need to break down every one of the news that shows up on daily papers, magazines and other literary assets. Yet, examination of such measure of monetary news and articles so as to concentrate valuable information surpasses human capacities. Content mining methods can help them consequently extricating the valuable information out of literary assets.

We would build up a framework which can utilize content mining strategies to show the response of the share trading system to news articles and foresee their responses. Thusly,

the financial specialists can anticipate the future conduct of their stocks when significant news are discharged and act promptly upon them. As information we utilize constant news articles and intra-day stock costs of a few organizations in Bombay Stock Exchange. The general motivation behind study can be compressed in the accompanying exploration questions:

- How to anticipate the response of stock value pattern utilizing literary monetary news?
- How information and content mining strategies produce this prescient model?

II. RELATED WORK

Writer strategy comprises of completing the NLP (Natural Language Processing) of news, portraying its components, arranging and removing the conclusions and sentiments communicated by the essayists. The strategy then recognizes the relationship in the middle of's news and securities exchange variances. In our test, we demonstrate that our strategy can be utilized to comprehend unstructured huge information, and we likewise uncover that news' opinion can be utilized as a part of anticipating stock value vacillations, whether up or down. The calculation extricated trials can be utilized to make expectations about securities exchange developments [1].

Here Author research whether estimations of aggregate mind-set states got from huge scale Twitter encourages are related to the estimation of the Dow Jones Industrial Average (DJIA) after some time. We break down the content substance of every day Twitter bolsters by two state of mind following apparatuses, to be specific Opinion Finder that measures positive versus negative inclination and Google-Profile of Mood States (GPOMS) that measures temperament regarding 6 measurements (Calm, Alert, Sure, Vital, Kind, and Happy). We cross-approve the subsequent state of mind time arrangement by contrasting their capacity with distinguish general society's reaction to the presidential

race and Thanks giving day in 2008. A Granger causality examination and a Self-Organizing Fuzzy Neural Network are then used to research the speculation that open mind-set states, as measured by the Opinion Finder and GPOMS mind-set time arrangement, are prescient of changes in DJIA shutting values. Our outcomes demonstrate that the precision of DJIA forecasts can be essentially enhanced by the consideration of particular open state of mind measurements however not others. We discover a precision of 87.6% in foreseeing the day by day here and there changes in the end estimations of the DJIA and a lessening of the Mean Average Percentage Error by more than 6%. [2] This theory examines the prescient force of online news on one-day stock cost up or down changes and high or low exchange volume of 19 noteworthy banks and money related establishments inside of the MSCI World Index, amid the period from January 1 2009 to April 16 2015. The news information compare to news articles, public statements, and stock trade data, and were acquired by a web-crawler, which filtered around 6000 online hotspots for news and spared them in a database. The news are parceled and named into two classes as indicated by which value change class, or exchange volume class, it relates. A directed robotized archive grouping model is made and utilized for forecast. The model does not succeed in foreseeing the one-day stock value changes, however the rate of effectively named reports in the one-day exchange volume analysis was 78.3%, i.e. an order precision of 78.3% was accomplished, proposing that online news contains some important prescient data. [3]

III. PROPOSED SYSTEM

Methodology for NLP module

To exactly predict the stock price is very complex task till the date. Here we are proposing to make a prediction based on news articles using one of the Text Mining concepts like sentiment analysis. We would like to make the prediction system for Indian Stock market. Implementation steps to be followed to make a prediction system are: 1. Gathering of news articles. 2. Perform sentiment analysis on news articles 3. Get Polarity of the text 4. Make a prediction based on current stock price and calculated polarity of the text.

To collect the news articles

R.S.S feed is the main source. As R.S.S feed is used for news article collection process. Here the Times of India's R.S.S feed is used for business and market related news. It will give results by retrieving top news of Indian stock market. We have to just specify the R.S.S feed address in our code. To Perform Sentiment Analysis and Get Polarity of the text Sentiment analysis (also known as opinion

mining) refers to the use of natural language processing, text analysis and computational linguistics to identify and extract subjective information in source materials.

A basic task in sentiment analysis is classifying the polarity of a given text at the document, sentence, or feature/aspect level - whether the expressed opinion in a document, a sentence or an entity feature/aspect is positive, negative, or neutral. For sentiment analysis and calculating polarity of text two things are used:

1. POS tagger
2. SentiWordNet_3.0.0 POS Tagger

A Part-Of-Speech Tagger (POS Tagger) is a piece of software that reads text in some language and assigns parts of speech to each word (and other token), such as noun, verb, adjective, etc., although generally computational applications use more finegrained POS tags like 'noun-plural'. This software is a Java implementation of the log-linear part-of-speech taggers.

IV. CONCLUSION

In Data Mining to predict stock market here we have created NLP based module & statistical parameter based module which results the sentence polarity & behavior compared to past year data. By using different technique we can get accurate & reliable prediction result which give consumer better solution for where to invest their valuable money. These modules evaluate the news sentences based on grammatical analysis and with the help of historical data also.

REFERENCES

- [1] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," Foundations and Trends in Information Retrieval, Vol.2, Nos.1-2, 2008, pp.1-135.
- [2] Fu, K. Lee, D. Sze, F. Chung and C. Ng, "Discovering the Correlation between Stock Time Series and Financial News," Proceedings of the 2008 IEEE/ WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, 2008, pp.880-883.
- [3] H. Chen and D. Zimbra, "AI and Opinion mining," IEEE Intelligent Systems, May/June 2010, pp.74-80.
- [4] J. Bollen, H. Mao and X. Zeng, "Twitter mood predicts the stock market," journal of Computational Science, Vol.2, 2011, pp.1-8
- [5] Kim, Y., N. Kim, S.R. Jeong, "Stock-index Invest Model Using News Big Data Opinion Mining," Journal of Intelligence and Information Systems, Vol.18, No.2, 2012.6, pp.143-156.
- [6] L. Zhuang, F. Jing and XY. Zhu, "Movie Review Mining and Summarization," Proceedings of the 15th ACM International Conference on Information and Knowledge Management, November 2006, pp.43-50.

-
- [7] M. A. Mittermayer and G. F. Knolmayer, "NewsCATS: A News Categorization And Trading System," Proceedings of the International Conference in Data Mining, 2006.
 - [8] R. P. Schumaker and H. Chen, "Textual Analysis of Stock Market Prediction Using Breaking Financial News: The AZFinText System," ACM Transactions on Information Systems, Vol.27, No.2, Article 12, February 2009.
 - [9] S. Ahn and S. B. Cho, "Stock Prediction Using News Text Mining and Time Series Analysis," Proceedings of the KIISE 2010 conference, Vol.37, No.1, 2010.6, pp.364-369.
 - [10] V. Sehgal and C. Song, "SOPS: Stock Prediction using Web Sentiment," Seventh IEEE International Conference on Data Mining – Workshops, 2009, pp.21-26.13 Text Opinion Mining to Analyze News for Stock Market Prediction