

Social Media Mining for Price Prediction of Stock Market Using Map Reduce Framework

^[1] Megha Rani Raigonda, ^[2] Gangamma
^[1] Assistant Professor ^[2] Post-Graduate Student
^{[1][2]} Department of MCA,VTU P.G Centre, Kalaburgi

Abstract— Big data may be an evolving expression that depicts large data sets that may be analyzed computationally to reveal patterns, especially relating to the behaviour and interaction of the human..At whatever voluminous add up of structured, semi structured and unstructured information that need those possibility with a chance to be mined to majority of the data in the stock market. The stock market price prediction is the one of the most difficult task because the price of the stock is changing instantaneously. In the traditional, the stock are only prices are predicting the based on the market sentiment, usually for the short-term periods for the small-caps. For the long term stocks, the statistical analysis methods are not beneficial. In our project work for the stock price prediction we are using the social media mining technology to evaluating the stocks of the market segment. We are using the big data analytic and the map reducing technique to predict the price of the stocks for long term for the user selling or buying the stocks in the best time.

Keywords— Big data ,map-reduce technique, big data analysis, statistical analysis.

1. INTRODUCTION

Stock prediction has recently grown to be a huge research area in the field of predictive analytics, big data analytics and statistical analysis. The field of stock prediction has used machine learning techniques as well as recently predictive analytics to predict stock prices. This paper presents description of big data analytics and stock prediction and its methodologies. It go on to describing the current methods of stock prediction methodologies that exist in the current system where the prediction methodologies follow statistical analysis methods and predictive analysis over social media and hybrid of the two as well. On portraying the flow approaches, we demonstrate that there is a great deal of change and research going ahead in the field of stock expectation and it will keep on improving the accuracy of stock forecast later on with more research going on..

There are many rules and factors that prompts the ascent and fall of the stock exchange. The expectation of the cost in stock markets and settling on the choice on there are very difficult task. For the short term investors the statistical analysis methods stand a good factors.

To produce high predictive accuracy of the stock value development. The data mining systems are profitable. A large number of the financial analysts and stock market investors are convinced. They can make benefit by utilizing one of the specialized examination ways to deal with estimate money markets.

2. RELATED WORK

In [1] Stock amount is a affectionate of time alternation in banking domain. The fluctuations of banking time alternation are dynamically, selectively, and nonlinear, non-stationary and with a lot of noises, which accomplish it difficult to anticipation.

In [2] Effectively predict banal amount by application abstracts mining or apparatus learning techniques in the abreast approaching has become one of the most significant issues. However, it is harder to accomplish predictions from the assumption of the able bazaar hypothesis, since the banal amount will chase a accidental plan pattern.

In [3] A anchored anticipation action is as well not possible because investors will anon ascertain such strategies and those acknowledged forecasting rules will advance to self destruct.

In [4] Social media mining originates from the relevant field of data mining, which mines patterns from structured data instead of unstructured. It is also related to other fields like information retrieval, web mining, statistics, computational linguistics and natural language processing

In [5] A valuable application of social media mining is text sentiment analysis, also referred to as opinion mining.

In [6] The purchase of the shares for the long term ,it follows the „long term trail“ technique.

In [7] Baylor University's Zhao, Lei. The proposed an outlier mining algorithm to detect anomalies on the basis of volume sequence of high frequency tick-by tick data of stock market. Using anomalies on distributions of trading volume to predict upward trends of stock prices.

In [8] ANN is usually chosen as a stock prediction tool compared to other methods. However, these approaches cannot work alone because the market value which is always subject to external impact. The stock market is affected by system uncertainties and other unknown factors.

In [9] The decision makers in the computer and web literate, Most of them predicting for this, they using computerized tools to support their work.

In [10] The volumes of the Sale Ranker prediction is done carefully by the „hand-crafted“ queries produce matching the post.

3. ARCHITECTURE



Figure 1: System Architecture

The biggest challenge for the AI community is the prediction of the stock market price. The various results are held by using the technical analysis, fundamental analysis and the statistical indicators in the proposed system. In recently the stock prediction movements are proved by using the artificial intelligence

Step 1: In the step1 understanding the objective. The objective is to generate the prediction based on the previous stock details which will assist the stock investors in their decision for the investment. The stock investors will view the prediction for the assistance before they invest in the market.

Step 2: The data has been collected by the collected by the user.

Step 3: In this step the data pre-processing and data wrangling tasks are takes place. In the data pre-processing technique the raw data is transformed into an understandable data. In the data wrangling processing the raw data is transforming and mapping into another format.

Step 4: In this step the data is processing and training.

Step 5: In this step the results are forecasting using the mapreduce technique.

Step 6: In this step the results are visualized.

Step 7: In this step the user can view and analysed the result

4. MODULES

Data set module

In the beginning the was gathered, the numeric ceaseless esteems are chosen from the estimations of the quality. The data as exchanged by applying the big data ideas. At first the information are in the mapping position. The quality esteems are exchanged relying upon the past stock closing. Assume the present stock cost is more than the past closing stock, it will be positive esteem. In the event that the present stock cost is not exactly of the present stock esteem at that point, it will give the negative esteem.

Analysis module

In this module the arranged and the exchanged data has been analyzed. In this step we are using the map reducing technique. Initially the information has been investigated at that point, Here we are arranging all the stock costs utilizing the map() function and diminishing the information utilizing the reduce() function.

Predictive module

The map reduce system is helping us to foresee the best time for the offering or purchasing the stocks .The user can see the all the stock costs of the current and the past stocks information, that can help the user to offer or purchase the stocks.

5. CONCLUSION

The stock forecast field has utilized machine learning strategies and the predictive analysis to expectation of the costs. The utilization of the statistical analysis of the stock costs are not beneficial for long term investor. To accomplishing more accuracy for long term investor the financial specialist we are utilizing the map reduce technique. it is exceptionally effective to contrast with the past ideas. Here we are utilizing the conveyed framework with the guide decreasing system ,It is quick preparing for getting the stock data, That is help to the client expectation. The proposed framework is quick handling and extremely productive.

REFERENCES

- [1] Wang, Y.F., "Mining stock price using fuzzy rough set system". "Expert Systems with Applications", 24.pp. 13-23,2003.
- [2] Wu, M.C., Lin, S.Y, and Lin, C.H., "An effective application of decision tree to stock trading", "Expert Systems with Applications".31, pp. 270-274,2006.
- [3] Al-Debie, M. Walker. M. , "Fundamental information analysis: An extension and UK evidence", "Journal of Accounting Research". 31(3), pp. 261– 280,1999.
- [4] ZHANG G, PATUWO B E, HU M Y, "Journal of economic perspectives. Forecasting with artificial neural networks". "The state of the art [J]. International journal of forecasting",14(1)pp. 35-62.,2003
- [5] TANG L, LIU H. "Community detection and mining in social media[J]". "Synthesis Lectures on Data Mining and Knowledge Discovery",vol 2(1): 1- 137,2010.
- [6] Jure Leskovec Lada A. Adamic Bernardo A. Huberman, "The dynamics of viral marketing". "Proceedings of the 7th ACM Conference on Electronic Commerce",2006.
- [7] ZHAO L, WANG L. "Price Trend Prediction of Stock Market Using Outlier Data Mining Algorithm" proceedings of the Big Data and Cloud Computing (BDCloud)", 2015 IEEE Fifth International Conference on, F, 2015 [C]. IEEE.
- [8] Kay-Yut Chen Leslie R. Fine Bernardo A. Huberman, "Predicting the Future". "Information Systems Frontiers",vol. 5 no. 1 pp. 47-61 ,2003.
- [9] Ramesh Sharda Dursun Delen "Predicting box-office success of motion pictures with neural networks". "Expert Systems with Applications",vol.30pp.243-254,2006.
- [10] Gruhl Daniel R. Guha Ravi Kumar Jasmine Novak Andrew Tomkins, "The predictive power of online chatter" ."SIGKDD Conference on Knowledge Discovery and Data Mining", 2005.