ISSN XXXX XXXX © 2018 IJESC

Research Article Volume 8 Issue No.3

Stock Market Prediction System using Hadoop

Akshay M. More[1], Pappu U. Rathod[2], Rohit H. Patil[3], Darshan R. Sarode[4]
BE Student[1, 2, 3, 4]
Department of Information Technology
NDMVP'S KBT COE, Nashik, Maharashtra, India

Abstract:
Stock Market has high profit and high risk features that's why its prediction must be in parallel of accuracy. Stock Market contains very huge amount of data that is working in terabytes and petabytes, these are very complex and can be learned by a data mining methods. Stock market has very large amount of investors who wants to invest their money into shares a through selling shares or buying the shares. Through the our propose system the investors are able to obtain the time stock information, finding stock information, finding stock chart, news and research on internet that can help users to find the right investments strategies with good profit. The input stock market data is in very huge amount and for accessing such huge amount of data we uses the data mining algorithms. We uses data mining techniques such as map reduce; classification, prediction etc. The input to system i.e. data is in very large amount so their requirements are also high. We uses Hadoop framework where we can access large amount of stock market data in parallel manner. Hadoop allows us to work on clusters with thousands of nodes. We also uses Naïve Bayes algorithm. This algorithm is for making decisions and inferential statistics that deals with probability interference. It is uses to knowledge the prior data and predict the future trends. By using different technique we can able to get accurate reliable prediction result which give consumer better solution for where to invest their valuable money.

Keywords: Data mining, Map reduce, Naïve Bayes

I. INTRODUCTION

The use of mobile devices and internet has become a very important part of our daily routine. Today we can hardly pick up newspaper, turn on television, overhear a conversion or talk to a friend without mentioning Internet. Using nothing more than an Internet connection and an account with an online broker, one can sell or buy shares of stocks.

The number of people using the Internet to invest is growing fast. As the stock runs up and down, of record territory, investors are increasingly turning to the Web to research, discuss the trade stocks and securities. The stock market is characterize by high-risk, high-yield, so investors are concerned about the analysis of the stock market and trying to get prediction of the stock market. However, stock market is impacted by the politics, economy and many other factors, coupled with the complexity of its internal law, such as price changes in the non-linear, and shares data with high noise characteristics, therefore the traditional mathematical statistical techniques to predict the stock market has not produced suitable results.

To analyze the large volume of data and to process it, is difficult and challenging. So to analyze this data we use the Hadoop framework. Hadoop is a very fast way for massively parallel processing. Hadoop analyze the scattered data of stock market and predict the future trends and solutions which would benefit to the investor. It has a file system that provides an interface between the user's applications and the local file system, which is the Hadoop Distributed File System (HDFS).

We introduce a system where user can able to finding stock information, finding stock chart and previous results that can help users to find the right investments strategies with good profit. This system also provides the analysis of previous information. This system helps to user to get the accurate prediction results.

II. LITERATURE SURVEY

Many researchers have contributed to the development of Stock market prediction system using Hadoop. Techniques used by these researchers are summarized below:

This article is based on Cloud Based Stock forecasting [1] using neural network and Cloud as Hadoop. Stock Market has high profit and high risk features, on the stock market analysis and prediction research has been paid attention by people. The stock price trend is complex nonlinear function so the price has certain predictability. Map Reduce[2] is a programming model and an associated implementation for processing and generating large data sets. Users specify a map function that processes a key/value pair to generate a set off intermediate key/value pairs, and a reduce function that merges all intermediate values associated with the same intermediate key. Many real world tasks are expressible in this model, as shown in the paper. Twitter disposition [3] predicts the share trading system.

Here Author research says whether estimations of collective mind-set states got from large scale twitter encourages. They failure the content material of every day Twitter bolsters by two state of mind following devices, to be specific solution Finder that measures positive versus bad inclination and Google-Profile of Mood States that trials temperament regarding 6 measurements. Our result suggest that the precision of DJIA forecasts can be enhanced by the consideration of particular open state of mind measurements however not others. Inter day news-based expectation of stock costs and exchanging volume.

This inspects the perceptive force of online news on single day stock cost up or down changes and high or low exchange volume of 19 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[4] 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 robotized archive grouping model is made and used for prediction.

The model does not succeed in prediction the single day stock value changes, however the rate of effectively named reports in the one-day exchange volume analysis, proposing that online news contains some important prescient data. Content Opinion Mining[5] to Analyze News for Stock Market Prediction Writer strategy comprises of completing the Natural Language Processing of news, depicting its components, ordering and eliminating the conclusions and opinions communicated by the essayists.

The strategy then recognizes the relationship in the middle of news and securities exchange variances. The calculation removed trials can be used to make expectations about securities exchange growths.

III. PROBLEM STATEMENT

The above literature survey gives an idea about the working on stock market prediction system using Hadoop in this stock market system the Stock market has very large amount of investors who wants to invest their money into shares through selling shares or buying the shares. Through this system the investors are able to obtain the time stock information, finding stock information, finding stock chart, news and research on internet that can help users to find the right investments strategies with good profit. Hadoop is the framework where we can access large amount of data in parallel manner. Hadoop allows us to work on clusters with thousands of nodes. A distributed file system in the Hadoop framework is the main component of Hadoop ecosystem and isused to manage the file system. Hadoop also implements a parallel computation algorithm i.e. Map Reduce, is the key algorithm. The Hadoop Map Reduce engine uses to distribute work around cluster. The number of people using the Internet to invest is growing fast. As the stock runs up and down, of record territory, investors are increasingly turning to the Web to research, discuss the trade stocks and securities. By using different technique we can get accurate reliable prediction result which give consumer better solution for where to invest their valuable money.

IV. SYSTEM OVERVIEW

User or new stock holder is the end user of the application. He searches for the company shares to invest money. And can also sell the shares. Money control is the leading financial information source. Manage your finance with our online investment portfolios Live Stock Price, Stock trading News Live etc. Server act as an intermediate for communicating with the database and the user. It is responsible for generating appropriate query to retrieve information from the database. It consist the data of the reviews that are fetched from the Money Control Website.

A. MAP REDUCE

In the proposed system, map reduce component is used to allow to work on small modules and work parallel. This system also uses HDFS distributed file system for managing the files. Map reduce is the high level programming system.

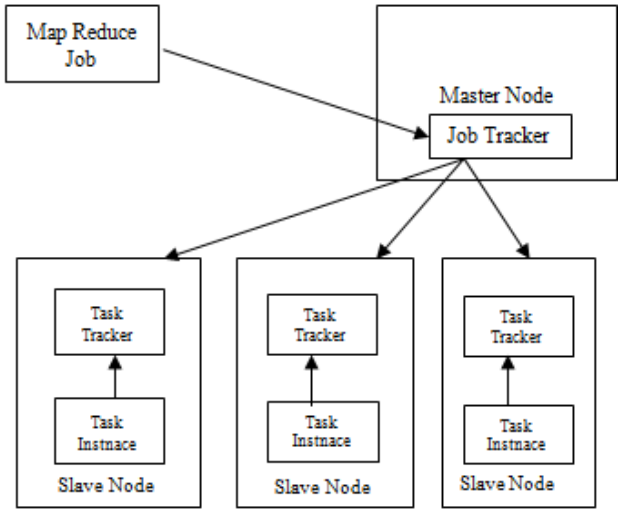

Figure.1. Map Reduce Architecture

This helps in doing the computation of the problem in parallel using all the connected machines so that the output, results are obtained in efficient manner. DFS also provides data replication up to three times to avoid data loss in case of media failures. The Master node stores the huge data HDFS and runs parallel computations on all the data i.e. Map Reduce.

B. HDFS

HDFS is a distributed file system that provides a limited interface for managing the file system to allow it to scale and provide high throughput. HDFS creates multiple replicas of each data block and distributed them on computers throughout a cluster to enable reliable and rapid access. When a file is loaded into HDFS, it is replicated and fragmented into blocks of data, which are stored across the cluster nodes; the cluster nodes are also called the Data Node. The Name Node is responsible for storage and management of metadata, so that when Map Reduce or another execution framework calls for the data, the Name Node informs it where the data is needed resides data, the Name Node informs it where the data is needed resides.

C. Naïve Bayes Theorem

This algorithm is for making decisions and inferential statistics that deals with probability interference. It is uses to knowledge the prior data and predicts the future trends.

V. ALGORITHM

Naïve Bayes Algorithm:

Derivation:

D: Set of tuples

- Each tuple is an ‘n’ dimensional attribute vector
- X: (x1,x2,x3,…,xn)

Let there be ‘m’ classes: C1, C2, C3,…. Cm

- Naïve Bayes classifier predicts belongs to Class Ci if
- P(Ci/X)>P(Cj/X) for l<=j<=m, j<>i

Maximum Posterior Hypothesis

- P(Ci/X)= P(X/Ci) p(Ci)/ P(X)
- Maximize P(X/Ci) P(Ci) as P(X) is constant.

With many attributes, it is computationally expensive to evaluate P(X/Ci).
Naïve Assumption of "class conditional independence"
$P(X/ Ci)= \Pi P(xk/ Ci)$
$P(X/ Ci)=P(x1/ Ci)*P(x2/ Ci)* \ldots.*P(xn/Ci)$

VI. CONCLUSION

Stock market has very large amount of investors who wants to invest their money into shares through selling shares or buying the shares. Through this system the investors are able to obtain the time stock information, finding stock information, finding stock chart, news and research on internet that can help users to find the right investments strategies with good profit. 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.

VII. REFERENCES

[1]. KushagraSahu, RevatiPawar, SonaliTilekar, Reshma Satpute, Stock-exchange forecasting using Hadoop MapReduce technique, International Journal of Advancements in Research & Technology, Volume 2, Issue04,April 2013.

[2]. Jeffrey and Sanjay Map-Reduce: Simplified processing on large cluster, Google Research Publication 2004.

[3]. J. Bollen, H. Mao and X. Zeng, "Twitter mood predicts the stock market," journal of Computational Science,Vol.2, 2011, pp.1–8.

[4]. 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.

[5]. H. Chen and D. Zimbra, "AI and Opinion mining," IEEE Intelligent Systems, May/June 2010, pp.74-80.

[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]. 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.

[8]. R. P. Schumaker and H. Chen, "Textual Analysis of Stock Market Prediction Using Breaking Financial News: The AZFin Text System," ACM Transactions on Information Systems, Vol.27, No.2, Article 12, February 2009.