International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 05 Issue: 04 | Apr-2018 www.irjet.net p-ISSN: 2395-0072

Improving Hadoop Performance by Using Metadata of Related Jobs for Stock Market Analysis

Pratibha Chaudhari[1], Tejaswini Halwar[2], Anuja Kamble[3], Khushboo Kothari[4]

[1,2,3,4] *BE Student, Computer Science, Savitribai Phule Pune University, Maharashtra, India*

Abstract - *Stock market is generating great deal of significant information that's operating in terabytes and petabytes. Currently a days securities market gains additional attention. The prediction of stock markets is thought to be a difficult job. Stock analyst, investors and stock brokers are attempting to analyze stock commerce information to predict future movement of stocks. This paper target acceptive stock market connected data for investors and analyze stock market information to predict stock future movements victimization Naïve Bayes and Hadoop. We planned associate increased Hadoop design that reduces the computation value associated with stock market analysis. It provides associate well-organized data processing approach for stock market analysis victimization Naïve Bayes. Our focus is on extracting options and building a data table on that naïve Bayes data processing rule is employed to predict the coming movements.*

Key Words*: *BigData, Hadoop, MapReduce, Stock market, Naïve Bayes, Text Data.

1. INTRODUCTION

Stock Market has high profit and high risk characteristics that tells why its prediction should be near correct. Exchange prediction is the act of attempting to see the future worth of a corporation stock or made prediction of a stock's future worth may yield extensive profit different money instrument listed on a money exchange. The [1].Stock Market handle the knowledge regarding the share market. The most problems regarding such knowledge sets are that these are terribly complicated nonlinear functions and may solely be learnt by a unique ways to spot the longer term market trend. To analyse the large volume of information and to method it, is tough and difficult and there are totally different ways. Hadoop could be an in no time manner for massively data processing. Hadoop analyse the scattered knowledge and predict the longer term trends and business intelligence solutions which might profit the enterprise and shopper all at once.

Hadoop could be a Java based mostly open supply framework that uses straightforward programming models to permit storing and process of huge knowledge in an exceedingly distributed computing setting across clusters of computers. It's a locality of the Apache code foundation. Hadoop runs applications exploitation map cut back rule that could be a programming framework for distributed computing. Here divide and conquer technique is employed to interrupt giant complicated knowledge. In this system prediction is employed for improve the performance of the system.

2. RELATED WORK

Applying data processing Techniques to stock market Analysis by Gabriel Fiol-Roig, Margaret Miro-Julia, and Andreu Pere Isern-Deya in 2010, Springer [2]. Here the researchers viewed the stock market analysis as an artificial intelligence drawback. First, data processing techniques are wont to evaluate past stock costs and acquire helpful data through the calculation of some money indicators. Next they applied computing methods to construct decision making trees and predictions square measure created. Clustering-Classification primarily based Prediction of securities market Future Prediction by Abhishek Gupta, Dr.Samidha D Sharma in 2014, IJSCIT [3]. The methodology enforced here for the prediction of stock market is bunch classification primarily based prediction like applying bunch rule like K-means and decision tree rule. the actual ways has been applied in term of two stages and prediction is created.

Stock price Prediction using K-Nearest Neighbour (kNN) algorithmic program by Khalid Alkhatib Hassan Najadat Ismail Hmeidi mohammed K. Ali Shatnawi in 2013 , IJBHT [4]. During this methodology, the researchers applied k-nearest neighbour algorithmic program and non-linear regression approach so as to predict stock costs. In classification approaches, a knowledge set is split into two sub sets like training data set and testing set. KNN algorithmic program uses similarity metrics to match a given test entity with the training information set to assist within the prediction method. As mentioned several researchers are successful in analysing the data set associated with exchange using various approaches like Regression based mostly data mining, clustering, classification and totally different algorithms like Kmeans clustering. Most of them used Neural Network approach to create the prediction model. So of these prediction models were developed exploitation data mining tools like Oracle data miner, Weka etc.

3. PROPOSED FRAMEWORK

The projected system focuses on building a prediction model victimization Hadoop map cut back technique. Projected construct deals with providing information by victimization Hadoop tool. The Hadoop software package

library offers a serious advantages of distributed process of huge information sets and it additionally provides high accessibility of information. Hence, Hadoop has been taken as a framework for developing the prediction model. during this system, an organization's daily stock information set is chosen as a coaching information set, and when analyzing the information set fully, a prediction model is developed, it will be wont to analyze however the stock are going to be for the longer term trends. We tend to get results with less time, high outturn and maintenance price is incredibly less [5, 7].

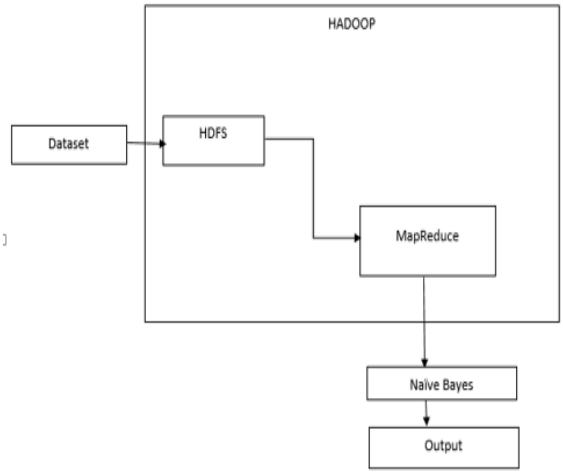

Fig 3.1: Architecure of proposed system

3.1. MapReduce

A MapReduce job is an access and process-streaming job that splits the input dataset into freelance chunks (blocks) and stores them in HDFS. Throughout MapReduce, multiple Maps area unit processed in parallel followed by scale back tasks additionally processed in parallel. Relying upon applications the numbers of maps will be totally different than that of reduces. Storing knowledge in HDFS has totally different forms like idea to determine the given parameter (Key) and to retrieve the desired result (Value) at the top of the job [6].

For instance, a "WordCount" job counts variety of replication of every word within the knowledge files. Figure one explains MapReduce example "WordCount" as a typical example to use MapReduce in such unstructured knowledge like books. As an input data, it consists of a sequence of characters that square measure separated by house, thus we will take into account the space as a delimiter that separates words. First step, Hadoop divides the data to blocks within the splitting section. Then, the Mapping section will for every word (e.g. Then, Shuffling section collects the values of identical key to be in one intermediate result. After that, the Reducing section provides the addition of prices to possess one final value for every key. Finally, NameNode provides a effect that has all keys and their values in concert effect from the MapReduce job.

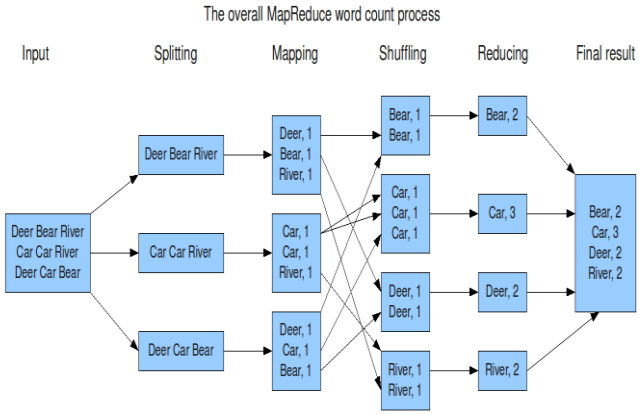

Fig 3.2: MapReduce Framework Example

3.2.Naive Bayes

Naïve mathematician rule could be a classification technique that generates Bayesian Networks for a given dataset supported Bayes theorem. It assumes that the given dataset contains a selected feature in an exceedingly category that is unrelated to the other feature. For instance, associate degree object is taken into account to be A owing to some options. These options presence could rely on different or on other options however all of the options presence severally contribute to the chance that this object could be a. which is that the reason it's referred to as „Naïve?.

Advantages of Naïve Thomas Bayes rule are it's straightforward to make and helpful for terribly giant datasets and even renowned to crush extremely refined classification techniques. Following were the necessary steps to be performed during this rule.

1. The given dataset is to be born-again into a frequency table.

2. Calculate possibilities of the events and victimization the possibilities produce probability table.

3. Victimization the Naive Bayesian equation, calculate the posterior chance for all categories.

4. The category with the best posterior chance is that the outcome of prediction.

4. CONCLUSION:

The studies reveal a high potential of Naïve bayes algorithmic rule in predicting the come back on investment within the share market. From the above analysis we discover the businesses who have created profits from every

industries. Most of the Investors prefer to invest in such company that is performing well within the equity market. The data should be helpful for analyst and users, who works within the exchange and analyse all the past records of an organization to recommendation their purchasers for investments.

ACKNOWLEDGEMENT:

We wish to take this opportunity to express my gratitude to principal Dr.K.S.Holkar and HOD Dr.V.S. Pawar for motivating and providing me the best facilities which were required in our project. We would like to thank our project guide Prof. P.P.Shinde for guiding us throughout the project development phase. We also thank our friends and lab staff for helping us in collecting information and solving tricky problems.

REFERENCES:

[1] Amarinder Cheema, Ateet Vora, Chetan Jain,Puneet Kataria ,Ronak Shah," web based stock forecasters" 7 may, 2008.

[2] Gabriel Fiol-Roig, Margaret Miro-Julia, and Andreu Pere Isern-Deya, "Applying Data Mining Techniques to Stock Market Analysis",2010, Springer

[3] Khalid Alkhatib Hassan Najadat Ismail Hmeidi Mohammed K. Ali Shatnawi, "Stock Price Prediction Using KNearest Neighbor (kNN) Algorithm", 2013 , IJBHT

[4] Abhishek Gupta, Dr.Samidha D Sharma, " ClusteringClassification Based Prediction of Stock Market Future Prediction", 2014, IJSCIT

[5] Alshammari, H., J. Lee, and H. Bajwa, "H2Hadoop: Improving Hadoop Performance using the Metadata of Related Jobs". IEEE Transactions on Cloud Computing, 2016.

[6] Hamoud alshammari, "improving hadoop performance by using metadata of related jobs in text datasets via enhancing mapreduce workflow", 2016.

[7] Mrs. Lathika j shetty, ms. Shetty mamatha gopal, "developing prediction model for stock exchange data set using hadoop map reduce technique", volume: 03 issue: 2016, irjet