# A STUDY ON ANALYSIS OF STOCK MARKET BY USING BIG DATA AND FINANCIAL ANALYTICS

**Manjunatha S[1]**
Research Scholar
Department of commerce and management
Maharani Women's arts,
Commerce & Management College
Sheshadri road, Bangalore
Research Centre in Management

**Dr K Ramchandra[2]**
Professor and Research Guide
Department of commerce and management
Maharani Women's arts,
Commerce & Management College
Sheshadri road, Bangalore
Research Centre in Management.

**Abstract:** - Big data analytics plays a vital 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.

Huge information examination can be utilized in numerous spaces for precise forecast and investigation of the enormous measure of information. They encourage the disclosure of critical data from huge information, which is concealed generally. In this paper, we depict a methodology for investigation of the securities exchange to comprehend its unstable nature and anticipate its conduct to make benefits by putting resources into it. We initially give writing overview of past deals with this domain. Then we provide a methodology of our approach which contains data collection and machine learning algorithms.

Keywords - Big Data, Stock Market forecasting, stock market, temporal data, big data architecture.

## 1. INTRODUCTION

Big data means a lot in financial services in the transformation of the organization services, profits, etc. It is more promising for financial analysts and also the investors for their services and investments. The large information gathered over stock message boards is being large assets for nearly 91 percent of organizations and it also uses the historical time series data for accurate predictions of the stock market. It also gains new insights of financial organizations as well as investors. Big data technologies create a value for these types of data in financial market. Once the data are collected from message board

(unstructured) they have to be classified based on the users sentiment on data in order to predict the correct results by integrating the historical data. The historical data is analyzed by financial volatility models. This paper deals with the models that can be used with predictive analytics of big data in financial market for better predictions.

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 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 holing 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.

**1.1 Sentiment Analysis** Sentiment analysis is the process of identifying user's view from their reviews or feedback from social media. These reviews are only unstructured and it is handled by the tools of big data which implements the machine learning algorithm for analyzing sentiments. The Sentiment Analysis is focused when it comes to big data analytics because of the reviews of customers who are majorly involved in the benefit of the organization. The sentiment of the reviews is based on the subjectivity f the content. For example the phrase like "not bad" means "good", the words directly give positive feedback are "good", "amazing", etc. the negative feedbacks include "poor", "not good", etc., These are the features to be extracted from each and every statement of the review and classified as polarity represented in figure 1. It is mainly used in fields like marketing and advertisement to improve the quality and profit of the product by various organizations. In financial analysis it is the reviews on stock exchange dash board from the investors or analysts or by any organizations.

**1.2 Financial Volatility**: Even though online reviews are analyzed using sentiment analysis methods the analysis of financial volatility using traditional volatility methods is important to make accurate predictions on stock or other financial data. The veracity will be there in the mere analysis of online reviews and hence it requires historical data analysis for producing better result than complete uncertainty in data. In finance, volatility could be a measure for variation of worth of a financial mechanism over a period of time. Historic volatility springs from statistic of past market costs. An implied volatility springs from the value of a market listed spinoff (in explicit, an option). The symbol σ is employed for volatility, and corresponds to std. deviation, that shouldn't be confused with the equally named variance, that is instead the square, σ2. In stock market, the shares may go up and down day by day which is not consistent is said to be the volatility. The historic data can be structured but it holds a large amount of data that could be integrated with reviews.

**2. Sentiment Analysis:** Sentiment analysis is the classification of sentiment features from the real life data such as comments posted on review boards. It plays a major role in big data analytics to provide predictive results with the machine learning algorithms. The sentiment is categorized into positive, negative and neutral. There are two main tasks product features are identified from the comments of reviewers and the comments are classified as positive or negative. These are very challenging tasks. The classification can be done with document level, sentence level, phrase level, etc., using algorithms such as machine learning technologies such as supervised and unsupervised algorithms

**2.1 Unsupervised Machine Learning**: Unsupervised learning is that of tries to search out the hidden structure in untagged knowledge. Because the examples provided to the learner are untagged, there's no error or reward signal to gauge a possible resolution. The unsupervised learning is distinguished from supervised learning and reinforcement learning when there is no error. Unsupervised learning traditionally uses the lexicon based approach for sentiment classification. These methods uses sentiment lexicon to identify entire document's sentiment polarity.

**2.2 Lexicon Based Approach:** It is the sentiment based on each word and phrase. It is identified by Turney as semantic orientation of the reviews. Later the lexicon based approach is used for sentiment extraction. A lexicon based approach is a practical and easy approach for Sentiment Analysis of data without a requirement for training. A Lexicon based approach is good on how the lexicon is used. A lexicon-based approach is mainly projected to perform task using opinion bearing words (or simply opinion words).Opinion words are the commonly used words to express positive or negative opinions (or sentiments), For example, "good", "bad", "poor" and "excellent". The count of positive and negative opinion words is used to determine the feature of the product in each sentence of the review. When the positive opinion words are more than the negative opinion words, then final conclusion on the feature will be positive or otherwise negative

**2.3 Supervised Machine Learning:** Supervised machine learning uses a trained label set to classify the sentiment on data. The training corpus is used for learning new classification of data. A set of training examples will be there for training data. Each example consists of an input object (vector) and a desired output value (supervisory signal).A literature study shows often SVM yields higher accuracy results than other techniques.

3. **Components of Big Data**

**Volume –**Volume refers to the vast amounts of data that is generated every second. The ninety percent (90%) of the world's data created in the last two years, the volume of data that is being collected daily is what presents immediate challenges for businesses.

**Velocity –**Velocity refers to the speed at which new data is generated and the speed at which it moves around.

**Variety –** Variety Refers to the different forms of data that we collect and use. Data comes in different formats, such as structured and unstructured. To make matters even more challenging, because of the explosion of data generated by social media sources, eighty to eighty five percent of all the world's data is now unstructured (text, audio, video, click streams, log files and so on)

**Veracity –** Veracity refers to the uncertainty surrounding data, which is due to data inconsistency and incompleteness, which leads to another challenge, keeping big data organized.

**Value –** Through effective data mining and analytics, the massive amount of data that we collect throughout the normal course of doing business can be put to good use and yield value and business opportunities.

### 3.1 Importance of Big Data

The main importance of big data consists in the potential to improve efficiency in the context of use a large volume of data, of different type. If big data is defined properly and used accordingly, organizations can get a better view on their business therefore leading to efficiency in different areas like sales, improving the manufactured product and so forth.

Big Data can be used effectively in the following areas:

- In risk assessment by analyzing information from the transactions on the financial market. In the future we propose to analyze the potential of big data and the power that can be enabled through big data analytics.
- In improving services and products through the use of social media content. By knowing the potential customers preferences the company can modify its product in order to address a larger area of people.
- In the detection of fraud in the online transactions for any industry.

## 4. Big Data Analytics

The world today is built on the foundations of data. Lives today are impacted by the ability of the companies to dispose, interrogate and manage data. The development of technology infrastructure is adapted to help generate data, so that all the offered services can be improved as they are used.

Big data analytics and the Apache Hadoop open source project are rapidly emerging as the preferred solution to business and technology trends that are disrupting the traditional data management and processing landscape. Enterprises can gain a competitive advantage by being early adopters of big data analytics. Even though big data analytics can be technically challenging, enterprises should not delay implementation. As the Hadoop projects mature and business intelligence (BI) tool support improves,

big data analytics implementation complexity will reduce, but the early adopter competitive advantage will also wane.

Technology implementation risk can be reduced by adapting existing architectural principles and patterns to the new technology and changing requirements rather than rejecting them. Big data analytics can be differentiated from traditional data processing architectures along a number of dimensions:

- Speed of decision making being very important for decision makers
- Processing complexity because it eases the decision making process
- Transactional data volumes which are very large
- Data structure data can be structured and unstructured
- Flexibility of analysis consisting in the amount of analysis that can be performed on it.

The big data analytics initiative should be a joint project involving both Information Technology and business. Information Technology should be responsible for deploying the right big data analysis tools and implementing sound data management practices. Both groups should understand that success will be measured by the value added by business improvements that are brought about by the initiative.

### 4.1 Big Data Architecture Design for Stock Market Data

Data required for the stock market prediction is placed and generated at different and distinct places of the globe. Distinguished agencies are maintaining this data in its own format. So, to cater need of current scenario, big data architecture for stock market is divided into three parts. News and Information flows from any corner have impact on investment. First job is to identify different data and its sources required for the future prediction of the market. Store, acquire and process of heterogeneity, unstructured and temporal data is the second major challenge. Last part of the architecture is to attain its goal. In this section, analyze, visualize, reporting and decision making has done. Every second data is generated for the market and its format is different from previous dataset. To store and integrate these temporal data for the stock market is the biggest challenge. In current scenario, cost required for these data is affordable by few big investors. Investors from third and fourth word countries have acute scarcity of resources. Big data architecture with latest technologies will provide easy access of information flow and it is relatively affordable to all investors.

Data required for the stock market prediction is placed and generated at different and distinct places of the globe. Distinguished agencies are maintaining this data in its own format. So, to cater need of current scenario, big data architecture for the stock market is framed as in Fig. 03. Big data architecture

for stock market is divided into three parts. News and Information flows from any corner have impact on investment. First job is to identify different data and its sources required for the future prediction of the market. Store, acquire and process of heterogeneity, unstructured and temporal data is the second major challenge. Last part of the architecture is to attain its goal. In this section, analyze, visualize, reporting and decision making has done. Every second data is generated for the market and its format is different from previous dataset. To store and integrate these temporal data for the stock market is the biggest challenge. In current scenario, cost required for these data is affordable by few big investors. Investors from third and fourth word countries have acute scarcity of resources. Big data architecture with latest technologies will provide easy access of information flow and it is relatively affordable to all investors

### 4.2 Financial Analytics

Financial analytics are widely used in research in the equity and corporate bond markets. Financial analytics form a key role in fundamental analysis. Fundamental analysis methods focus on what should happen to stock or corporate bond prices by examining why they move. Those who perform fundamental analysis look at the economic, political and environmental factors that could affect a company's or an industry's performance; they also study financial information and use a range of financial analytics to review past performance and future prospectus.

Company fundamental data is a principal research resource and serves as basis for financial analytics. Investment managers and analysts look for services that provide company annual and interim results over several years in order to track performance. They want quick access to key sections of companies' accounts and to the financial measures commonly used in analytics. These range from multiple profit measures and revenue and cash flow figures to pre calculated ratios, such as earnings or dividend per share.

While portfolio managers and analysts use analytics based on historical information for a view of past performance, they also need access to analytics that project future performance. As a result, they want services that additionally provide forecast information ranging from detailed brokers estimates to company guidance on future financial results.

For financial analytics to be meaningful, they need to draw on consistent information across all companies and over time. Users look for services that offer accurate and timely data, presented in a standardized format, and that adopt uniform analytics methodologies.
They also want purpose-built analytics tools and functionality that enables them to find and filter information quickly, search by multiple criteria, and chart and download data to perform their own analytics.

## 4. LITERATURE REVIEW

1. Supposition mining and conclusion examination This review covers methods and methodologies that guarantee to specifically empower supposition arranged data looking for frameworks. Our attention is on techniques that try to address the new difficulties raised by conclusion mindful applications, when contrasted with those that are now present in more customary certainty based examination. We incorporate material on rundown of evaluative content and on more extensive issues with respect to security, control, and monetary effect that the advancement of sentiment situated data access administrations offers ascend to. To encourage future work, a dialog of accessible assets, benchmark datasets, and assessment battles is likewise given.
2. Content Opinion Mining to Analyze News for Stock Market Prediction 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.
3. Programmed forecast of organization stock citations The paper contains a review of stock citations expectation strategies identified with Artificial Intelligence space. A forecast framework which utilizes stock citations and daily paper articles identified with an organization is displayed.
4. Robert Engle: The usefulness of ARCH and GARCH models are described in financial forecasting. The standard deviation is calculated to find the usefulness of both the models. The weights for forecasting the variance is determined by the ARCH model. GARCH model also estimate the weights but it never decline it to complete zero. It is easy and simple form to estimate by conditional variances. Both the models are used in various applications but successful especially in financial forecasting due to the risk measurement. These models and their derivatives provide the storage to analyze and test the pricing and its change over time.

## FINDINGS

Big data integration will impact greatly on the growing markets like India. There is no need to store same data again and again. Intermediaries and other interested parties will be benefited by two ways. First, they can get data without wasting time and cost as per their own desire format. Secondly, they exploit the benefit of other existing benefits of Big or cloud computing like SaaS, Paas, IaaS etc. Implementation of big data technologies and above design architecture will reduce use of resources with same efficiency and effectiveness. Implementation of green computing will also helpful in the forecasting technologies or methods. Indian stock markets are still working with the traditional database and only upper class of investors has capacity to afford information from all database maintained by different agencies. Because of it, market theories and definitions like efficient market hypothesis, fractal

market hypothesis etc are not applied properly in the third world markets. Extensive, heterogeneous, unstructured and temporal data required for the market prediction will handled only by big data technologies, i.e. cloud computing, parallel computing, high performance computing, distributed computing etc. Unique feature of cloud computing like Sharing of resources must be incorporated in the market for more sustainability of the investment process and economies.

## 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. Rudy Prabowo and Mike Thelwall, "Sentiment Analysis: A combined approach," Journal of Informatics 3 (2009) 143-157.
2. David J. Hand, "Data, Not Dogma: Big Data, Open Data,and the Opportunities Ahead", LNCS 8207, pp. 1-12,2013.
3. Emerson, J.W., Kane, M. J.: Don't drown in data.Significance Vol. 9(4), pp. 38–39, 2012.
4. FergusonGT. Strategy in the digital age: Role ofinformation technology in corporate strategic planning.Journal of Business Strategy, Vol. 17(6), pp. 28–31, 1996.
5. Gartner (2012), http://www.gartner.com/ DisplayDocument? id=2057415&ref =client FriendlyUrl.
6. B. Pang, L. Lee, and S. Vaithyanathan. "Thumbs up? Sentiment classification using machine learning techniques". In Proceedings of ACL, pages 79–86, 2002.
7. xia.hu, jiliang.tang, huiji.gao, huan.liu, "Unsupervised Sentiment Analysis with Emotional Signals", WWW'13, May 13–17, 2013, Rio de Janeiro, Brazil. ACM 978-1-4503-2035-1/13/05.
8. B. O Connor, R. Balasubramanyan, B. Routledge, and N. Smith. "From tweets to polls: Linking text sentiment to public opinion time series". In Proceedings of ICWSM, 2010. J. Wiebe, T. Wilson, and C. Cardie. "Annotating expressions of opinions and emotions in language". Language Resources and Evaluation, 39:165–210, 2005.
9. T. Wilson, J. Wiebe, and P. Hoffmann. "Recognizing contextual polarity in phrase-level sentiment analysis". In Proceedings of HLT and EMNLP, 2005.

10. P. Turney. 2002. “Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews”. In ACL-2002.
11. Robert Engle, “GARCH 101: The Use of ARCH/GARCH Models in Applied Econometrics”, Journal of Economic Perspectives - Vol 15, Number 4- Fall 2001-Pages 157-168.
12. Chien-Jen Huang, Peng-Wen Chen, Wen-Tsao Pan, “Using multi-stage data mining technique to build forecast model for Taiwan stocks”, Springer-Verlag London Limited 2011.
13. David J. Hand, “Data, Not Dogma: Big Data, Open Data, and the Opportunities Ahead”, LNCS 8207, pp. 1-12, 2013.
14. Emerson, J.W., Kane, M. J.: Don’t drown in data. Significance Vol. 9(4), pp. 38–39, 2012.
15. Ferguson G. Strategy in the digital age: Role of information technology in corporate strategic planning. Journal of Business Strategy, Vol. 17(6), pp. 28–31, 1996.