

Survey on Factors Determining the Share Price Movement in Indian Stock Market - A Big Data Approach

Nitha K. P.*¹, Dr. S. Sivakumari²

¹Research Scholar, Department of Computer Science and Engineering, Faculty of Engineering, Avinashilingam Institute for Home Science and Higher Education for Women Coimbatore, Tamil Nadu, India

²Professor and Head, Department of Computer Science and Engineering, Faculty of Engineering, Avinashilingam Institute for Home Science and Higher Education for Women Coimbatore, Tamil Nadu, India

ABSTRACT

The aim of investors of any stock is to earn reasonable returns. This paper presents a framework on share price movement and investigates the factors determining and predicting the share prices in Indian stock market through big data approach. This study is also trying to focus on the fundamental factors, technical factors and sentimental factors that stimulate the share price movement. The methodology used mainly focuses on the implications of the theory of financial markets and features affecting the share price movement. Stock market is one of the burning areas where data is growing day by day. Big data Analytics is a very promising area and buzzword for the next generation information technologies. Because of the heterogeneity and other complexities of data, big data architecture and design, is needed which specifically deals with the stock market data and analyze the heterogeneous data for the future prediction of the market. Variety and large Volume of factors exists. Moreover, the velocity with which data generates and the effect of them in the share price movement emphasizes, role of big data in predicting the share prices in Indian Stock Market.

Keywords : Stock Market, Big Data, Stock valuation model, Technical, Fundamental and Sentimental Factors

I. INTRODUCTION

For an investor, predicting the share price is significant and critical. Data required for the stock market prediction is placed and generated at distinct places of the world. Industry specific factors such as EPS, book value, P/E ratio and dividend yield has influenced the share price movement of Indian companies a lot. The present study has been undertaken with an objective to review the existing literature by examining the relationship between stock prices and company specific intrinsic factors or fundamental factors such as book value per share, dividend per share, earning per share, price earnings

ratio, dividend yield, and dividend payout. This study also examines the importance of technical factors and sentimental factors, which highly influences the share price movement of Indian companies.

A. FUNDAMENTAL FACTORS

- 1) **Book Value (BV)** : It is also known as net asset value per share because it measures the amount of assets, which the corporation has on behalf of each equity share. BV shows the net investment per share made in the business by the shareholder. It is the value at which an asset is carried on a balance sheet.
- 2) **Earning Per Share (EPS)**: The ratio of the profit

after tax of the company for any financial year after **payment** of preference dividend. The equity shareholders are the sole appellants to the net earnings of the corporation after making payment of dividend to the preference shareholders. The significance of this ratio flows from the fact that higher the earnings per share the more is the scope for a higher rate of dividend and of retained earnings, to build up the internal power of the company.

- 3) **Dividend per share (DPS):** Dividend is the portion of the profit after tax, which is distributed to the shareholders for their investment bearing risk in the company. It has a significant influence on the market **price** of the share. DPS shows how much the company has paid out as dividend. It refers to the actual amount of dividend (gross) declared per share. The net profit after taxes belong to shareholders but the income that they really receive is the amount of earnings distributed and paid as cash dividend.
- 4) **Price Earnings Ratio (P/E ratio):** This ratio enables an investor to make appropriate calculation of the time required to cover the investment in a company's stock. P/E ratio expresses the relationship between the market price of a company's share and its earnings per share. It indicates the extent to which the earnings of each share are covered by its price

B. TECHNICAL FACTORS

Technical indicators are mathematical formulas that, when applied to security prices clearly flash either buy or sell indications. Price data includes any combination of the open, high, low, or close over a period and most of the indicators use only the closing prices, technical indicators are usually shown in a graphical form above or below a security's price chart for the analysis. Once shown in graphical form, an indicator can then be

associated with the matching price chart of the security. Technical indicators offer many uses such as confirm the trends, generate buy/ sell signals, forecast the trend of future prices. Technical analysis is a technique, which gives an idea about future share prices of selected companies in which we invest. Based on the knowledge of technical analysis one can predict the perfect investment decision of the stock market. By using the nominal pointers, the future market of securities will know in which stock to invest. The more accurate prediction of stock prices of selected companies the investor to carry out fundamental analysis of stock prices, they can predict of future trend of stock prices.

- 1) **Moving Averages:** Most visual aids shows a lot of discrepancy in price movement. This can make it difficult for traders to get an idea of a security's overall trend. One simple method traders use to combat this is to apply moving averages. A moving average is the average price of a security over a set amount of time. They smooth a data series and made it easier to spot tendencies, something that is especially helpful in volatile markets.
- 2) **Simple Moving Average (SMA):** A simple moving average is formed by computing the average (mean) price of a security over a specified number of periods. For example: a 10 day moving average(SMA) is taken as the closing prices for the last 10 days added together and dividing the total by 10.
- 3) **Exponential Moving Average (EMA):** Exponential moving average is the moving average that is formed by applying weight to the recent price changes.

$$EM A(c) = ((price(c) EM A (p) * K) + EM A (p))(1)$$

Where c=current, p=previous

The formula for the smoothing constant K is:

$$K = 2 / (1 + N) \quad (2)$$

Where, N = number of periods for EMA

4) **Relative Strength Index (RSI)** : The relative strength index (RSI) is an extremely useful oscillating momentum indicator that was developed by J. Welles Wilder and is one of the most widely used indicators in technical analysis. RSI oscillates between zero and 100. The most popular is the 14 days RSI where the RSI is calculated based on 14 days values. Traditionally the stock is considered to be over bought when RSI is above 70 and over sold when RSI is below 30. Signals can also be generated by looking for divergences and central line crossovers. The basic formula is:

$$RSI = 100[100/(1 + RS)] \quad (3)$$

Where, RS = average of upward price change over a select number of days / average of downward price change over the same number of days

C. SENTIMENTAL FACTORS

Sentiment analysis extracts subjective information in source materials with the help of natural language processing, text analysis and computational linguistics. Sentiment analysis aims to determine the attitude of a speaker or a writer with respect to some topic or the overall contextual polarity of a document. The elementary job in sentiment analysis is to categorize the feelings of a given text in the document, sentence, and the feature/aspect level of the expressed opinion in a document, or a sentence or an entity feature/aspect is positive, negative, or neutral. Advanced, “beyond

polarity” sentiment classification looks, for instance, at emotional states such as “angry,” ”sad,” and “happy.”

Existing methods applied to extract the sentiment analysis, which grouped mainly into four categories: keyword spotting, lexical affinity, statistical methods, and concept-level techniques. Keyword spotting categorizes text based on the presence of unambiguous affect words such as happy, sad, afraid, and bored. Lexical affinity detects an affect words and assigns a probable affinity to particular emotions. Statistical methods has control on elements from machine learning such as latent semantic analysis, support vector machines, “bag of words” and Semantic Orientation Point-wise Mutual Information.

Concept-level approaches leverage on elements from knowl- edge representation such as ontologies and semantic networks and, hence, are also able to detect semantics that are expressed in a subtle manner, e.g., through the analysis of concepts that do not explicitly convey relevant information, but which are implicitly linked to other concepts that do so. The main data sources for the sentimental analysis are taken from Twitter as they provide an API to extract the tweet data from the site. The following features capture useful characteristic of Twitter and writers for opinion retrieval.

II. METHODS FOR STOCK MOVEMENT PREDICTION

The Support Vector Machine (SVM) been recognized as being able to efficiently handle high dimensional data and has shown to perform well on many classification and clustering techniques

1) **Price only:** In this method, only past data or prices predicts the stock movement. The

purpose of this method is to investigate whether there are patterns in the history of the stock or not. In addition, this model was used as a standard to estimate whether combination of the sentiments is effective for the approach by comparing with other sentiment models. Features used for the training of SVM are price t_1 and price t_2 which are the price movements (up, down) at the transaction dates t_1 and t_2 , respectively.

- 2) **Human sentiment:** In addition to historical prices, this model integrates the sentiments annotated by human into the prediction model. According to the Message Board dataset, the users explicitly attach a sentiment label with their posts. The sentiment tags like strong buy, buy, hold, sell and strong sell are usually seen labels attached with the posts. Instead of using all the messages, only the messages with annotated sentiments by the users need to be considered, and all the other irrelevant messages can be discarded. From these messages, only the explicit sentiment will be taken and remove other information such as message content. The perseverance of this method is that how attitude interpreted by human can be used to predict the stock. Because human marks the sentiments, this feature is one of the strongest features for stock prediction.
- 3) **Sentiment classification:** To utilize the remaining 84.4% of the messages without the explicit sentiments, the build model extracted the sentiments for those messages. A classification model will be trained from the messages with annotated sentiments on the training dataset. Then it can be used to classify the remaining messages into five classes (Strong Buy, Buy, Hold, Sell, and Strong Sell).
- 4) **LDA-based method:** In this model, we consider each message as a mixture of hidden

topics. Latent Dirichlet Allocation (LDA) (Blei, Ng, & Jordan, 2003) is a generative probabilistic model of a corpus. The basic idea is that documents are represented as random mixtures over latent topics, where each topic is characterized by a distribution over words. Therefore, the LDA is chosen as a simple topic model to discover these hidden topics.

- 5) **JST-based method:** The method refers to opinion often expressed on a topic or aspect. When people post the message on the social media to express their opinion for a given stock, they tend to talk their opinions for a certain topic such as profit and dividend. Based on pairs of topic-sentiment, they would think that the future price of that stock would go up or down. From that intuition, a new feature topic-sentiment for the stock prediction model results.
- 6) **Aspect-based sentiment:** As an alternative of considering the blends of hidden topics and sentiments as in the previous model, in this model the combinations are not hidden. Each message is represented as a list of topics and their corresponding sentiment values are considered. The topic selected as the consecutive nouns in the sentence will provide tremendous results in prediction.

III. IMPORTANCE OF BIG DATA IN SHARE PRICE PREDICTION

Data required for the stock market prediction is placed and generated at different places of the world. Distinguished agencies are maintaining this data in its own format. So, to cater to the need of the current scenario, big data architecture for the stock market has been framed. Big data architecture for the stock market is divided into three parts. i) **First part** is to identify different data and its sources required for the future prediction of the market. ii) **Second Part** is to store, acquire and process the heterogeneous, unstructured and temporal

data.iii) **Last part** of the architecture is to attain its goal, the prediction of the stock price.

News and Information flows from any corner have impact on investment. Every second data 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 world countries have acute scarcity of resources. Big data architecture with latest technologies will provide easy access of information flow; it is relatively affordable to all investors. Little, and slow access of information became big reason for the big loss in the market for small investors. All information need would be integrated very fast and provide information to the small and medium investors at very reasonable & affordable cost.

Traditional database will not be able to handle and integrate this type of heterogeneous, temporal, unstructured and dynamic data stored at different places, so a new architecture is required for the integration of the heterogeneous database. Data will not be available at one place or one database and it is in the different format too. It requires an integrated mechanism which provides all temporal data at one place and easy to process for the decision-making process in desired format.

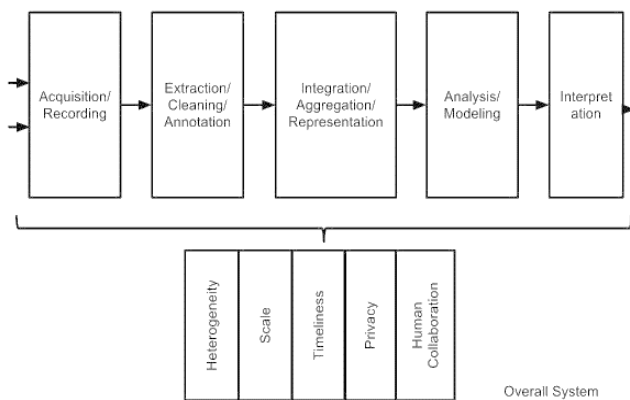


Fig. 1. Big Data Analysis Pipeline Architecture

Stock market is affected by all information. As per the efficient market hypothesis, stock market must react on all of the information. Small and Middle class investors have not such type of resources to know all happenings instantaneously and take their positions respectively. Loss fear factor had been playing a crucial role for the small and middle class population and that is why they are keeping distance from the market. Easy availability of this integrated data from all corners of the globe is the key issue.

Due to temporal nature of the data required in the stock market, big data architecture ,being implemented in such a way that all data must be updated and integrated on timely basis and reports should be generated based on the current information. To achieve efficient market hypothesis, these information must reach all participates of the market. Due to shared resources and other above dis- cussed benefits of big data technologies, small investors too have equal opportunity to access excess relevant data. This will increase participation, participation will increase more investments, investments will increase more stable market with market hypothesis, more realistic market hypothesis will lead to more accurate prediction of the market and finally it leads to stable and mature market.

Cloud based big data engine is used to integrate different heterogeneous and unstructured data. Cloud based technology will provide all analytical services to the clients at reasonable costs. Distributed file systems like HDFS (Hadoop Distributed File System) and parallel processing in the cloud will increase efficiency in the processing of data with quick response. Big data integration will affect greatly on the growing markets like India. There is no need to store same data repeatedly. Intermediaries and other interested parties gain benefits by two ways. First, they can get data without wasting time and cost as per their own desired format. Secondly, they exploit

the benefit of other existing assistances of Big or cloud computing like SaaS, PaaS, and IaaS for forecasting the share prices. A model developed based on the framework of big data analytics along with the blend of the fundamental, Technical and Sentimental factors will surely be a promising model for the accurate prediction of stock prices in Indian stock market.

IV. REFERENCES

- [1]. V. Sandhiya, T.Revathi, A.Jayashree, A.Ramya, S.Sivasankari "Stock Market Prediction on Bigdata Using Machine Learning Algorithm", International Journal of Engineering Science and Computing, Volume 7 Issue No.4, April 2017
- [2]. Rupawari Jadhav, M. S. Wakode "Survey : Sentiment Analysis of Twitter Data for Stock Market Prediction", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 6, Issue 3, March 2017.
- [3]. Venkata Sasank Pagolu, Kamal Nayan Reddy Challa, Ganapati Panda, Babita Majhi "Sentiment Analysis of Twitter Data for Predicting Stock Market Movements", International conference on Signal Processing, Communication, Power and Embedded System (SCOPE)- 2016
- [4]. Varunesh Nichante, Prof. Sulabha Patil "A Review: Analysis of stock market by using big data analytic technology", International Journal on Recent and Innovation Trends in Computing and Communication, Volume: 4 Issue: 1, ISSN: 2321-8169, January 2016
- [5]. Aparna Nayak, M. M. Manohara Pai and Radhika M. Pai Prediction Models for Indian Stock Market, , Twelfth International Multi-Conference on Information Processing-2016 (IMCIP-2016), pp : 441-449
- [6]. P.G Preethia, V. Umab, Ajit kumar, Temporal Sentiment Analysis and Causal Rules Extraction from Tweets for Event Prediction, , International Conference on Intelligent Computing, Communication & Convergence. Procedia Computer Science pp: 84-89, July 2015
- [7]. Raghavendra Kune, Pramod Kumar Konugurthi, Arun Agarwal, Raghavendra Rao Chillarige and Rajkumar Buyya "The anatomy of Big Data Computing", SOFTWARE: PRACTICE AND EXPERIENCE, Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/spe.2374, 9 October 2015
- [8]. Dr B Charumathi and Suraj E S "An Artificial Neural Network Approach on Improving Relative Valuation Accuracy for Bank Stocks in India Asian Research Journal of Business Management Issue 2 (Vol. 3) Issn: 2321-9246, 2015
- [9]. Kavitha S, Raja Vadhana P, Nivi A N "BIG DATA ANALYTICS IN FINANCIAL MARKET", IJRET: International Journal of Research in Engineering and Technology, Volume: 04 Issue: 02 - Feb-2015
- [10]. Aditya Bhardwaja, Yogendra Narayanb, Vanrajc, Pawana, Maitreyee Duttaa Sentiment Analysis for Indian Stock Market Prediction Using Sensex and Nifty, , 4th International Conference on Eco-friendly Computing and Communication Systems. Procedia Computer Science 70 (2015) pp: 85-91
- [11]. Krishna Kumar Singh, Priti Dimri, Krishna Nand Rastogi "The Implications of Big Data in Indian Stock Market", International Journal of Computer Applications, Volume 99, Issue: 5, August 2014
- [12]. Dr B Charumathi and Suraj E S "Comparing Stock Valuation Models for Indian Bank Stocks International Journal of Accounting and Taxation, Vol. 2, No. 2, pp. 111-127, June 2014
- [13]. Carson Kai-Sang Leung, Richard Kyle MacKinnon, Yang Wang "A Machine Learning Approach for Stock Price

- Prediction", IDEAS14, Porto, Portugal, ACM 978-1-4503-2627-8/14/07, July 2014
- [14]. Yoosin Kim¹, Seung Ryul Jeong¹, Imran Ghani "Text Opinion Mining to Analyze News for Stock Market Prediction", *International Journal in Advanced Soft Computing and Applications*, Volume 6, Issue:1, March 2014
- [15]. Qasem A. Al-Radaideh, Adel Abu Assaf, Eman Alnagi "Predicting Stock Prices Using Data Mining Techniques", *The International Arab Conference on Information Technology (ACIT2013)*
- [16]. Dr B Charumathi and Suraj E S "Value relevance of earnings and book value for bank stocks An empirical analysis using ohlson valuation model *Global Journal of Management and Research*, July 2013.
- [17]. Phaisarn sutheebunjard "Stock exchange of Thailand Index prediction using back propagation networks *Second international conference on computer and network technology*, IEEE 376-381, Thailand, 2010 (2010).

Cite this article as :

Nitha K. P., Dr. S. Sivakumari, "Survey on Factors Determining the Share Price Movement in Indian Stock Market - A Big Data Approach ", *International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT)*, ISSN : 2456-3307, Volume 5 Issue 1, pp. 63-69, January-February 2019. Available at doi : <https://doi.org/10.32628/CSEIT195113>
Journal URL : <http://ijsrcseit.com/CSEIT195113>