How to Utilize Big Data for Business Intelligence in the Stock Market

Walaa Bajunaid
King Abdulaziz University
Jeddah
Saudi Arabia

Maram Meccawy, PhD
King Abdulaziz University
Jeddah
Saudi Arabia

ABSTRACT
Big Data plays a vital role in the stock market, especially for traders who need real-time information. Due to the bulk and nature of such data, several data mining technologies have been developed and employed in their collection, classification, storage and analysis, putting them in a form that is useful to traders. In the stock market, big data is useful in fundamental and technical analysis as it captures both historical trends as well as market sentiment. This paper discusses the possible uses of big data for business intelligence by investors in the stock market.

General Terms
Your Big Data, Business Intelligence.

Keywords
Finance, Big Data, Business Intelligence, Stock Market, Prediction, Decision making.

1. INTRODUCTION
Big data has an important impact on the financial services. The big data technique is used in transforming services and profit margins of business organizations via efficient decision-making. The big data are promising for the financial analysts and the investors. Nevertheless, the effective predictions, as well as preventions of future adverse events, are some of the objectives of any trading company. Big data, nonetheless, bring the great power of not only the ability to process a large volume of data and variety of it at high speed but also the capacity to bring value to the company. Big data could be defined as a term used to portray a large data volume that is both unstructured as well as structured, which inundates a firm on the daily running of its core activities [1]. However, it is not the data amount that matters to the enterprise but what a firm does with the data. The big data is significant in analyzing insights that may lead to some favorable decisions as well as strategic moves of the business entity. This paper examines an approach that can be employed while making efficient decisions in the stock market using big data.

2. THE STOCK MARKET
The process of the stock market is unpredictable and is influenced by several factors. The prediction of the stock exchange is one of the significant exertions in business and finances. The two kinds of analysis that are possible for prediction, include the fundamental and the technical approaches. The technical analysis of the stock market based on big data approach is undertaken with the help of historical stock price data from the machine learning application. The fundamental analysis of the stock exchange with the aid of big data is conducted with the help of data from the social media. This approach is made through applications of sentimental analysis. The data from the social media has great influence nowadays and may assist in predicting the stock market trends. This approach involves collecting news as well as data from the social media through extractions of sentiments expressed by various individuals. The correlation between the stock values and the views are analyzed. The learned model is momentous in predicting the future values of the stock market since the method can predict individuals' sentiments and the performances of the stock [2]. One of the primary objectives of big data is to provide firms with proficiency, in addition to accessible analytics, to assist in predicting the future marketability as well as profitable return. A proficient stock market analyzed using big data does not assist in exchange only, but also help the inexperienced and the experienced traders in the stock market. For a company or an individual to trade effectively in the stock market, one must depend on his/her ability to analyze the market trends. Big data play a vital role in analyzing these stock markets in an attempt of any firm or individual making profit in the stock market. Big data have enabled all traders, whether individuals or companies have access to an efficient analysis of these markets. This accessibility, on the other hand, is boosted by the availability of some applications on the internet as well as websites. Via common software on the Internet, for instance, Google Trends, the analysis has turned out to be easier as well as profitable for every company or individual trading in the stock market [3]. Businesses use the bulk information assembled over the message boards in stock markets in determining the trends of the stock market. Once information is gathered from the message boards, it is classified regarding the sentiment of the user so as to create a correct and profitable prediction. The classification is done by integrating some of the historical data, and the financial volatility operations analyze the historical data. These processes or models include the Supervised Machine Learning, Financial Volatility and the Naive Bayes Classifier [4].

3. SENTIMENT ANALYSIS
Sentiment analysis is a process to identify and classify user's view from their reviews or feedback from social media. 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 [5]. There are two main tasks: (1) product features are identified from the reviews of users, and (2) the comments are classified as positive or negative. These are the features to be extracted from each statement of the review and classified as polarity represented in figure 1. These are very challenging tasks. The classification can be done with text level, sentence level, phrase level, etc., using algorithms as machine learning technologies divided to supervised and unsupervised algorithms [6].

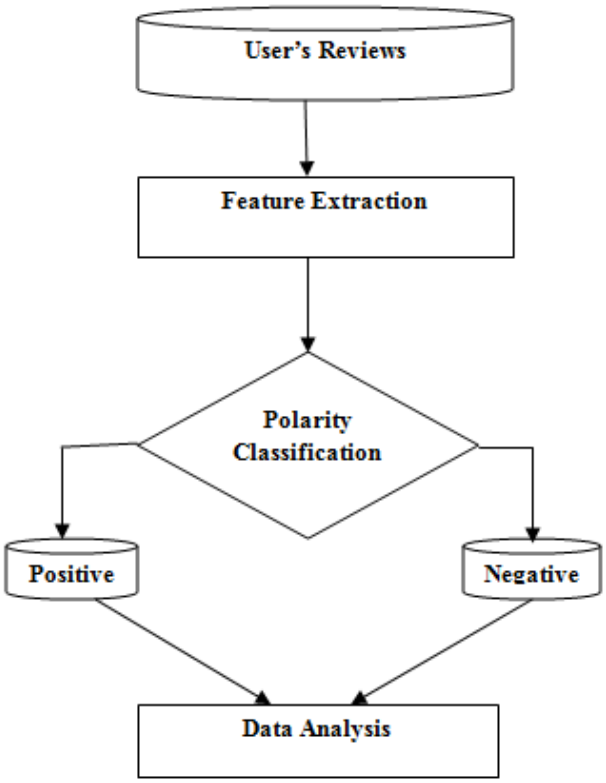

Fig 1: Sentiment Analysis

3.1 Unsupervised Machine Learning

Unsupervised learning is that of trying to search out the hidden structure in untagged knowledge. Because the examples provided to the learner are untagged, there is no error or reward signal to gauge a possible resolution. 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 [7]. These methods uses sentiment lexicon to identify entire document's sentiment polarity [7] - [8]. Lexicon Based Approach is the sentiment based on each word and phrase. It is identified by Turney as semantic orientation of the reviews [9]. 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 [10].

3.2 Supervised Machine Learning

Supervised machine learning (SVM) uses a trained label set to classify the sentiment on data. The training corpus is used for learning new classification of data [11]. 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 SVM often produces higher accurate results than other techniques [6]. Supervised machine learning uses Naive Bayes classification and Support Vector Machine. Naive Bayes classification method is a relatively simple method produces good results with reasonable accuracy. It uses a bag of words for the classification of self-content. It is based on naive Bayes rule assuming conditional independence, which is the main drawback of this category [12]. For the classification of a document d and class c using Bayes' theorem is given by the equation.1, and the equation. 2, [16].

$$p\left(\frac{c}{d}\right) = \frac{p\left(\frac{d}{c}\right)p(c)}{p(d)} \qquad (1)$$

$$c^* = argmaxP\left(\frac{c}{d}\right) \qquad (2)$$

Support Vector Machine (SVM) is a classification and regression model used for data analysis. It constructs a set of hyperactive planes used for linear classification and regression shown in figure 2. It uses kernel mapping for nonlinear classification [13], which is used for erroneous classification. It produces more accurate results on both classification and regression than other machine learning techniques [6].

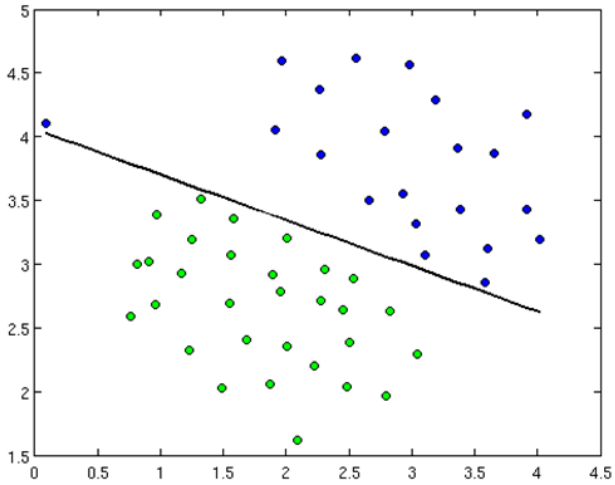

Fig 2: SVM Linear Classifier

4. FINANCIAL VOLATILITY

Financial volatility is predictable and deals with the historical data, which is a time series model. These large data is handled with some classic methods of financial volatility to determine the future. It is very important for financial investments of investors [14]. Volatility is the change in prices of assets during particular period. It is measured with the standard deviation continuously with the specific period. Gaussian process is the widely used process for this measure but the efficient method ARCH was first introduce by Engle [15]. Later it is extended to GARCH, which is more persistent even other models are derived from it for the stock market volatility.

5. BIG DATA IN DECISION MAKING

Developing into an analytic-driven kind of organization aids a company in minimizing costs and increasing its revenues. These are some of the main objectives of a company, in addition to improving competitiveness. Business intelligence, as well as analytics, continues to be a priority for Chief Intelligent Officers (CIO's) of all organizations. There are five major pathways employed in business organizations to optimize data with the analysis of big data to come up with efficient decision-making processes. These include better analytical capability, updated systems, obtaining faster hardware, sharpening the operational, organizational intelligence and getting into cloud computing. These approaches help the CIO's in profitable decision making for their respective companies [1]. Stock market firms are using big data primarily in five different ways as outlined below:

5.1 Trading Analytics

Big data is useful in predictive analytics, high frequency trading and pre-trade decision support analytics. The latter includes temporal analytics and sentiment measurement among others. Most trade decisions are guided by prevailing market sentiment [4].

5.2 Risk Analytics

The stock market is prone to fraud and other illegal activities. Big data is useful in fraud mitigation, on-demand risk management, tracking rogue trading, capturing Know Your Customer (KYC) information and anti-money laundering efforts [4]. It is also useful in monitoring trading activities by individuals and companies in real time.

5.3 Financial and Reference Data Management

Big data from various asset classes from various vendors is used in supporting historical trading, and for overall referencing.

5.4 Regulation

The participants of the stock market use big data in preparation for regulation by the governments. Real time data is crucial in regulations such as Dodd Frank, audits, Solvency II and others that are important to the capital market [4].

5.5 Data Tagging

Data tagging helps in monitoring trades and economic events in real time. Corporate actions can be analyzed and signs of financial stress detected early enough. Monitoring and reporting helps in matching and reconciling trades across platforms that use different symbols. In fact, while technical challenges of using big data are very real, the most critical aspect is its impact on how decisions are made and who gets to make them. If the use of big data for companies is statistically significant and economically important as it reflects in measurable increases in stock market valuations, for public entities this sounds like a revolution in decision-making. Occasionally, when the data are rare, because either it is not available in digital form or expensive to obtain, it makes sense to let well-placed people to make decisions, and this is what to do on the basis of the experience that has accumulated and patterns and relationships that have noticed and internalized. This is what is commonly called the HIPPO way to make decisions, which the highest-paid person's opinion. Therefore, to move from this approach to an ideal approach which based on the big data is an administrative challenge. Access to a torrent of data certainly improve their performance only if used in the right way. As Andy Rooney said: "Computers make it easier to do a lot of things, but most of the things they make it easier to do don't need to be done". Companies and the public sector gather more and more data, and maybe more than it knows what to do with. To turn this flood of new information "into gold", that requires skill and new leaders led more "*what-if approach.*"

6. CONCLUSION

Previously the big data had been regarded as a technical problem, but nowadays has turned out to be a business opportunity and a useful tool in helping companies in predicting trends in the exchange markets. The big data approach also assists various organizations in developing efficient decisions to run their activities in a smooth manner. This big data help business to have the adequate business knowledge and the more knowledge a company has, the greater the confidence it will have in making significant decisions. In general, big data are critical in the stock market decision-making process, and it is always perfect for a company to develop a trading plan as soon as it has an abundance of information to do so.

With the continuous maturation of big data and cloud computing technologies, the focus has turned from how to store and process large amounts of data to make accurate decisions, effective, and inexpensive. This may require a quantum leap in how to manage and analyze the data.

7. REFERENCES

[1] R. Walker, From big data to big profits: Success with data and analytics. New York: Oxford University Press, 2015

[2] H. Regina, "Big data Analytics the Next Big Learning opportunity", Academy of Information & Management Science, 2015.

[3] S. Mehta and S. Srinivasa, "Big Data Analytics: Third International Conference", BDA 2014, New Delhi, India, December 20-23, 2014 Proceedings. Cham Springer, 2014.

[4] T. Madjid and K. Puranam, Handbook of Research on Organizational Transformations through Big Data Analytics. Hershey: Business References, 2015.

[5] R. Prabowo and M. Thelwall, "Sentiment analysis: A combined approach", Journal of Informetrics, vol. 3, no. 2, pp. 143-157, 2009.

[6] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques." Proceedings of the ACL-02 conference on Empirical methods in natural language processing-, Volume 10. Association for Computational Linguistics, 2002, pp. 79-86.

[7] 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, vol. 11, no. 21, pp. 122-129, 2010.

[8] T. Wilson, J. Wiebe, and P. Hoffmann, "Recognizing contextual polarity in phrase-level sentiment analysis", Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing (HLT '05), pp. 347–354, October 2005.

[9] P. Turney. 2002. "Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews", Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL'02), Philadelphia, Pennsylvania, pp.417-424, 2002.

[10] M. Hu and B. Liu, "Mining and summarizing customer reviews", Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 168-177, 2004.

[11] E. Boiy, K. Deschacht, P. Hens and M. Moens, "Automatic Sentiment Analysis in On-line Text", Proceedings of the international Conference on Electronic Publishing, Vienna, Austria, pp 349-360, June 2007.

[12] P. Gamallo and M. Garcia, "Citius: A Naïve Bayes Strategy for Sentiment Analysis on English Tweets", Proceedings 8th International Workshop on Semantic Evaluation, pp. 171-175, Dublin, Ireland, 2014.

[13] Christopher J.C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition", Kluwer Academic Publishers, Boston, pp 1–43.

[14] S.-H. Poon and C. W. J. Granger, "Forecasting volatility in financial markets: A review". J. Econ. Lit., vol. 41, no. 2, pp. 478–539, 2003.

[15] P. Ou and H. Wang, "Financial volatility forecastingby least square support vector machine based on GARCH, EGARCH and GJR models: Evidence from ASEAN stock markets". Int. J. Econ. Fin., vol. 2, no. 1, pp. 51–64, 2010.

[16] Bankar, "A Survey on Web based Traffic Sentiment Analysis", International Journal of Advanced Research in Computer Science and Software Engineering, vol. 5, no. 11, pp. 685-688, 2015.