Predicting the Trends of the Stock through Sentiment Analysis

[1]Sunil Kumar Sharma, [2]Sanjai K Dewakar, [3]Rajat Sharma* and [4]Anurag Kumar

[1]Master Scholar,Department of Computer Science
Abdul Kalam Azad Technical University
[2,3]Research Scholar, Department of Computer Application, Shri Ramswaroop Memorial University
[4]Assistant Professor, CSE Department, Goel Institute of Technology and Management.

Abstract : **Stock market is often viewed as a way to grow the wealth for the individuals or a company. Investors need a greater insight to know the company in which they can invest in, by analyzing the company performance. With the advancement of information technology and analytical software, there is a scope for prediction of future stock values. Stock market prediction model is the most challenging field of the computer science. The thesis aims at investigating various methods which allow the investors to take informed decisions regarding the company they want to invest in. The main ingredient in decision making is the sentiment analysis, which automatically extracts the opinions or sentiments expressed in the news articles concerned with the company that is being investigated. In order to find the impact of the sentiments, the thesis first aims at achieving the correlation between the actual stock values and predicted stock values with sentiments. A neural network has been used for prediction of future stock value.**

IndexTerms - Naïve Bayes Classifier, Artificial Neural Network, Back propagation, Stock

I. INTRODUCTION

It is a proven fact that investing in the stock market fetches more profits in the longer period of time but that is not always the case. Choosing which stock to invest in plays a crucial role in the profit to be obtained. Investing in a stock is nothing but committing our money to a particular company listed in the stock market, with an expectation of getting an optimized profit. A solid investment is not possible without a prior homework on the investor's part. A proper analysis is an important step, before agreeing to invest on a particular stock. Randomly chosen stocks for investment can lead to adverse losses.

Many newbie investors today do not have any idea of how the future of the stock will look like. For them, it's more like a show of hands, invest in any randomly chosen stock, if its price goes up, lucky enough else better luck next time! In reality, it doesn't work that way. The apt prediction of the future stock price is one of the most trending and debated topics of all times concerning to various fields which include statistics, finance, trading and computer science, its motivation unquestionably, to forecast the direction of the future stock price, so that they can be bought and then sold at greater profits

1.1 Challenges in Stock Market Prediction Model:

- Randomness
- Violent Fluctuations
- Noise
- Contribution of many factors
- Unlimited number of potential influencers
- Accuracy

1.2 Area/Field covered: The main objective of this thesis is to study the impact made by the sentiments, in the online news articles pertaining to the financial institutions listed in the stock market, on the direction of the stock prices. To investigate the impact of the news articles on the stock price, a different set of text mining techniques can be implemented. Based on this relationship between news articles and stock prices a prediction system model would be designed. The overall scope of this work is to provide the predictive power to the investor in the web environment so that he could take informed decision of whether he can invest in the company in question, and yield high profits.

II. A REVIEW OF STOCK MARKET PREDICTION METHOD

Knowledge discovery in databases (KDD) is considered to be the intersection of databases, artificial intelligence, pattern recognition, information retrieval and expert systems. Data Mining had been evolved as a step in KDD. Data mining refers to the process of producing the useful patterns, by applying the data analysis methods and algorithms. These methods and algorithms take computational efficiencies into consideration

The various steps in the KDD process are shown in the Figure 1. In line with KDD, KDT (Knowledge Discovery in Text) showed its existence as well. According to the definition given by (Fayyad, 1996), KDT is the process of identifying the useful, novel and understandable patterns in the unstructured text data.

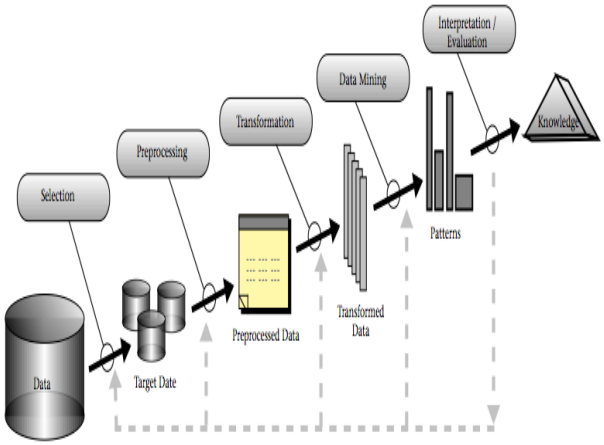

Fig1: An overview of steps in KDD process.

2.1 Methods of stock market Analysis:

Fundamental Analysis: Fundamental analysis (Abarbanell et.al.,1997) is a way to evaluate the stock for predicting the stock price movement. It uses a method called "financial analysis" to achieve the same. The information that has to be taken into consideration, for analysis, includes the annual financial statements and reports of the company, its balance sheet, its health, its future prospects, industry comparisons, market environment and changes in the government policies etc. The stock market investor can use all these facts and figures to decide upon the feasibility to invest on that particular company. They can additionally use the ratios for further analysis like Price/Earning (P/E ratio), Price /Book value, Debt/Equity, Return on Equity, Current Ratio and Net Profit Margin.

Technical Analysis: The technical analysis, on the other hand, is a research on the stock prices in the stock market with the intent of making profits and/or investment decisions (YingziZhu et al., 2009). The technical analysis, when applied to the stock market, predicts the direction of the future stock prices based on their historic data. With a close examination of the previous price movements of a stock, the investor can predict the future price movements of that particular stock. But again, this forecasting too may not be 100% accurate but just like the weather forecasting.

Sentimental Analysis: Sentiment analysis is a research area where the sentiment of a news article, or a review is defined as the opinion or sentiment expressed in that news article or the review (Turney, 2002). Sentiment analysis is also referred to as opinion analysis (Lee et al, 2008). The terminology, history and information regarding these terms were in their work. They concluded that both terms are similar. We use the term sentiment analysis through this thesis. For analyzing sentiment of a news article there exists three methods (Lee et al., 2008), Machine learning, linguistic and lexicon methods (Liu, B., 2012) showed that the sentiment can be detected at three levels namely document, sentence and entity.

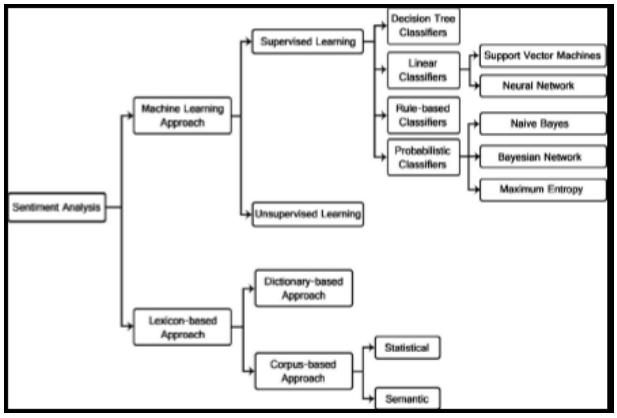

Fig2 : Techniques for semtimental classification

III. SYSTEM MODEL:

The overall system architecture is shown in the Figure 3. The main components of the system architecture are data extraction, data pre-processing, Sentiment classification and the Extraction of historical values. The main module of system are

- Neural Network Predictor
- Combined Technical and Sentiment Analysis
- Price prediction Model

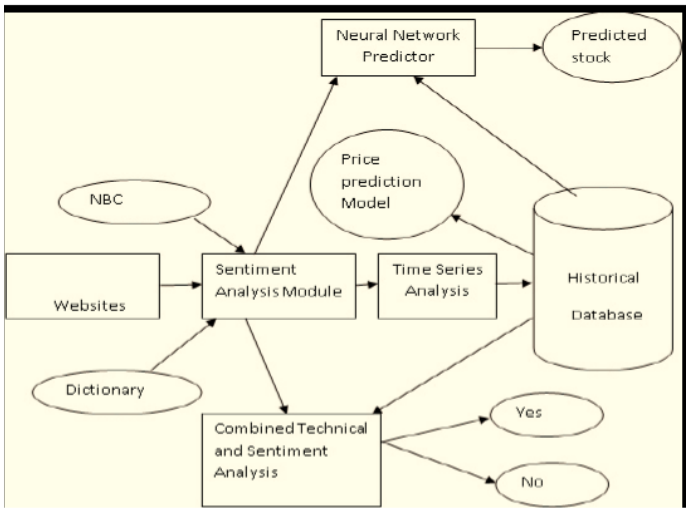

Fig3 : System Model

IV. METHODOLOGY

The overall process of predicting the direction of the stock market consists of various steps which include data collection, text pre-processing and feature selection. In order to carry out the overall research the programming is needed. R language, python, Java was used. The packages of those tools were used to implement the proposed algorithms.

The data needed for our problem is of two types. The historical stock prices and the news articles from which the sentiments are to be extracted. In contrast to the other systems, which used the static data, our system is based on the streaming data as well as the static data. The crawler crawls on the specified website and extract the news articles of the specified company for which the future direction of the stock is to be predicted. Since the stock prices have to be correlated with the news articles, the news articles have to be extracted along with the timestamps. These news articles are then served as the input to the sentiment analysis module.

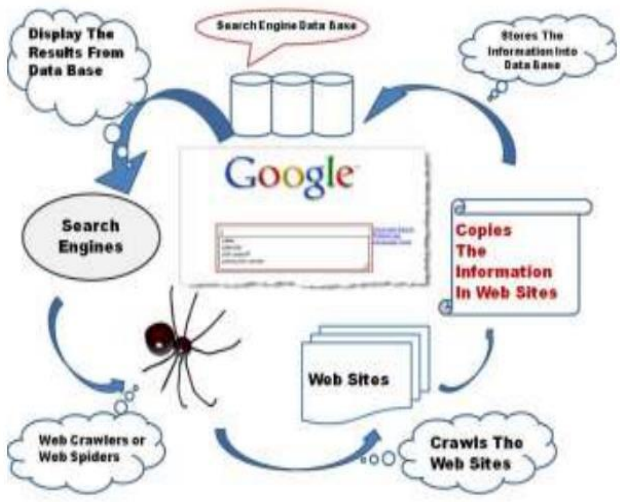

Fig4 : Components of web crawler

V. SENTIMENT ANALYSIS ALGORITHM:

The steps involved in sentimental analysis are:
Step 1: Gathering data from internet is solely based on the (SOR) Subject of Reference (e.g. ICICI bank). We use web mining techniques (ex. crawler) to gather all web pages where the SOR is mentioned.
Step 2: Text Extraction can be done using several data mining or text mining techniques starting from simple 'keyword matching' to 'DOM structure mining' to 'neural networks' methods. The major challenge here is that web documents are highly unstructured and no single method can give 100% clean text extraction for all documents.

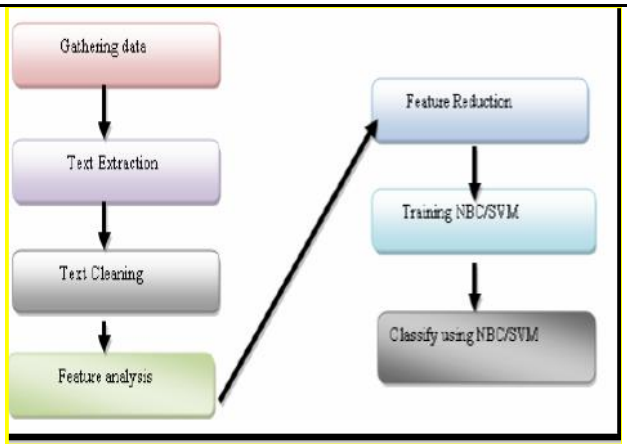

Fig5 : Steps in sentimental analysis

Sentiment Analysis

After the news articles pertaining events are extracted the sentiment analysis is done to classify to check whether it is a negative event, positive event or neutral one. As mentioned in the previous chapters the main approaches for automatic extraction of sentiment are lexicon and machine learning approaches. The first approach is based on considering semantic orientation of the words in the news articles (Turneyet al.2002). In our system, we follow the first approach of sentiment analysis that uses a dictionary of words with polarity. In lexicon based approaches the dictionaries can be created manually (Tong., 2001) or automatically (Turney et.al.,2003). Most of the researchers in the researcher community made use the generalized dictionaries for lexicon based sentiment analysis. The Loughran McDonald Dictionary is available, which contains the list of the positive and negative words pertaining financial domain. Our work has made use of this dictionary. Most of the researchers working on the prediction of stock market used this dictionary, but almost all of them extracted the news articles of the companies listed in the foreign stock markets like NYSE etc. Since we focus on the companies listed in BSE, it was thought of building a dictionary manually in combination with LoughranMcDonald Dictionary so that it is customized accordingly, (YoosinKim et.al, 2014) manually built a dictionary for sentiment classification. The steps involved are shown in the figure.

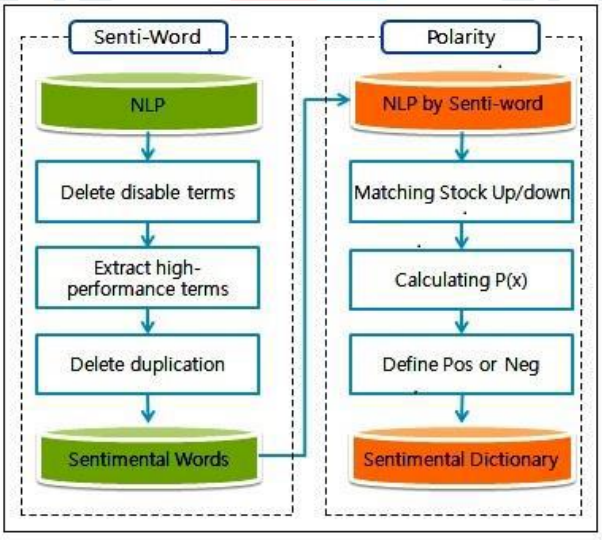

Fig 6: Steps involved in creating dictionary

VI. CONCLUSION AND FUTURE WORK

This thesis attempted to achieve objectives such as providing evidence that the stocks have a correlation with the sentiments in the news items by building a real time system to predict the future stock price of a company.

We also aimed at establishing a system which gives overall performance of a financial institute based on technical and sentiment analysis. We focused on building a system that would analyze and predict the variations in stock prices over a timeline, based on the sequence of events. The thesis focused on the news articles published on a particular website whose authenticity is not questionable. But there are some other key role players apart from news articles. Many investors seek guidance from some financial experts. These websites provide the expert advice also. These advices can be included in the sentiment analysis.

A future work in the experiment with neural network can be to extend the results obtained here using Deep Multilayer Neural networks containing more than two hidden layers to determine the sentiment of the text

VII. REFERENCES

1. GiliYenandChengFew Lee (2008), Efficient Market Hypothesis (EMH): Past, Present and Future, *Review of Pacific Basin Financial Markets and Policies*,Vol. 11, issue 2. Griffin, J. M., N. H. Hirscheyand P. J. Kelly(2011), How Important Is The Financial Media InGlobal Markets?, *Review of Financial Studies* 24, pp 3941–3992.

3.Grob-Klubmann. A and N. Hautsch(2011), When Machines Read the News: Using AutomatedText Analytics to Quantify High Frequency News-Implied Market Reactions, *Journal of Empirical Finance* 18, pp 321–340.

4. GS Atsalakis and KP Valavanis (2009), Surveying stock market forecasting techniques–Part II: Soft computing methods, *Expert Systems with Applications*,*Elsevier*, pp5932–5941.

5. Haddi. E, Liu. X and Shi.Y(2013),The role of text pre-processing in sentiment analysis, *Procedia Computer Science*, pp 26–32.

6. Hao Han, Tomoya Noro and Takehiro Tokudab(2009),An Automatic Web News Article Contents Extraction System Based on RSS Feeds. *Journal of Web Engineering*, Vol. 8, No. 3, pp268–284.

7. Jerome Kuperman, Manoj Athavale, and Alan Eisner (2003),Financial analysts in the media: Evolving roles and recent trends, American *Business Review*, 21(2), pp 74-80.

8. Jin W., Ho H. H. and Srihari R. K. (2009), Opinionminer: a novel machine learning system for web opinion mining and extraction. *KDD*, pp 1195–1204.

9.K. Michalak and P. Lipinski (2005), Prediction of HighIncreases in Stock Prices using NeuralNetworks, *Neural Network World*, Vol.15,pp. 359–366.

10. Khurshid Ahmad and Yousif Almas (2005), Visualizing Sentiments in Financial Texts?,*IEEE, pp 362-368.*

11. Koppel, M and Shtrimberg. I (2009), Good News or BadNews? Let the Market Decide, *AAAI Spring Symposium onExploring Attitude and Affect in Text*, pp86-88.

12.Li. X , Wang. C , Dong. J and Wang. F (2011), Improving stock market prediction by integrating both market news and stock prices, *Lecture Notes in ComputerScience: Database and Expert Systems Applications*, Vol. 6861, pp 279–293.

13.Long Sheng Chena, Cheng-Hsiang Liub and HuiJuChiua(2011), A neural network based approach for sentiment classification in the blogosphere, *Journal of Informetrics, Elsevier*, pp313–322.

14. Loughran, T. and B. McDonald (2011), When Is A Liability Not A Liability? Textual Analysis,Dictionaries, And 10-Ks, *Journal of Finance 66*, pp 35–65.

15. M. Cecchini, H. Aytug,G.J. Koehler and P. Pathak (2010), MakingWords Work: Using Financial Text as a Predictor of FinancialEvents, *Decision Support Systems*, vol. 50, no. 1, pp 64-175.

16. M. Cohen, P. Damiani, S.Durandeu, R. Navas, H. Merlino and E. Fernandez,Sentiment analysis in micro blogging: apractical implementation, *RedUNCI*, pp 191-200. http://sedici.unlp.edu.ar/bitstream/handle/10915/18642/Documento_completo.pdf?sequence=1

17. M. L. Mitchell and J. H. Mulherin (1994), The impact of publicinformation on the stock market, *Journal of Finance*, vol. 49(3), pp 923-950.

18. Maite Taboada, Julian Brooke, Milan Tofiloski and Kimberly (2011),Lexicon-Based Methods for Sentiment Analysis, *ACL*. Vol. 31, No 2, pp 267-306.

19. Marina Resta (2000), Towards an Artificial Technical Analysis of Financial Markets, *IEEE*, pp117-122.

20.MasoudMakrehchi, Sameena Shah and WenhuiLiao (2013), Stock Prediction Using Event-based Sentiment Analysis, *IEEE*, pp 337-342.