

---

## A Textual Analysis of Financial Articles for Prediction of Polarity

---

<sup>1</sup>Nischal S. Puri, <sup>2</sup>Dr. Avinash J. Agarwal, <sup>3</sup>P.S. Prasad

<sup>1,3</sup>Department of Computer Science & Engineering, PIET, Nagpur

<sup>2</sup>Department of Computer Science & Engineering, Shri Ramdeobaba College of Engineering & Management, Nagpur

Email: \*nischalspuri@gmail.com, avinashjagrawal@gmail.com

Received: 20<sup>th</sup> September 2018, Accepted: 11<sup>th</sup> October 2018, Published: 31<sup>st</sup> October 2018

### Abstract

Finding Polarity has been used in various domains. In this approach our domain is Stock market which is a promising domain. [11] Using information such as discussion a thread, blogs of expert in their respective fields, technical data, quarterly results of company etc. has been generated such enormous information is available on internet. Generally information is referred as 'Unstructured Data'. In this paper we discussed how we can use this unstructured data for determining the polarity of text. In Data mining techniques many applications have used to interpret such data. However, Stock data has dependencies among different parameters, and its characteristics change with time and natural scenario. [11] Here we discussed techniques which can be helpful for predicting the prices of stocks. However, there is only outlined structure for the problem in our interested domain. The discussed process will identify words in an article and compares with the dictionary. This dictionary contains words which will be helpful to find polarity. A weight is assigned to each word in order to calculate the polarity. The calculation of the final sentiment score produced by the model. The application of our technique is openly available in the open-source code

### Keywords

*Textual Analysis; Data Mining; Machine Learning; Bags of Words; Stock Market.*

### Introduction

In [11] discussed a huge amount of data, company annual or periodic meetings, company's quarterly results of etc has been available on various Internet websites. Therefore, the important problem is to extract knowledge from these information and make use of information. For this, knowledge about techniques involved in data mining is required. In now days, it has been able to study the huge data because of developments in technology.

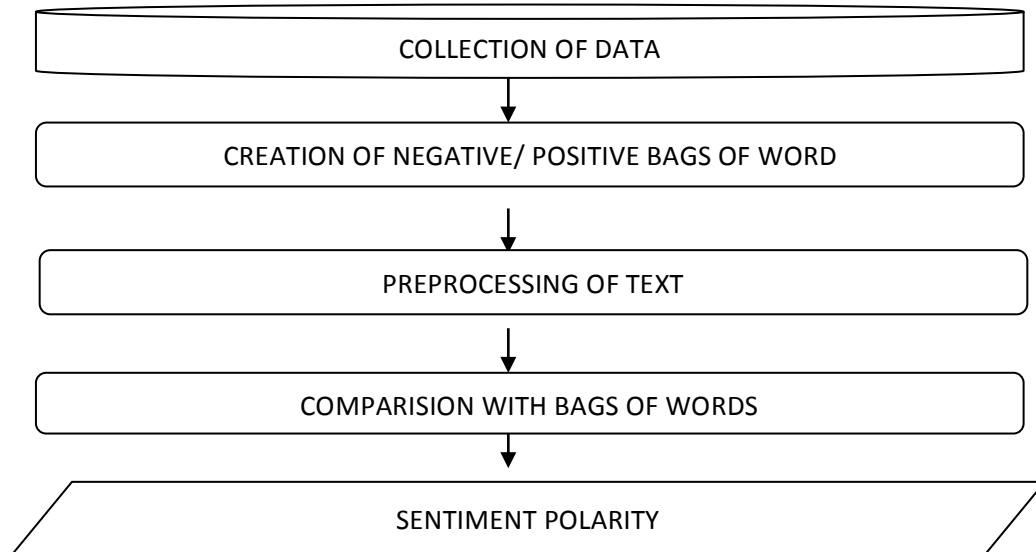
Here, we discuss the Analysis of extraction of text techniques for trading in stocks to find the rise and fall of stocks before the actual movement occur in order book profit either by increase / decrease in the stock price. In General practice, two popular Prediction Methodologies in domain of Stock are as follows [11].

Fundamental Analysts carried out, by this method is more focused with the basics of the company than the stock. The analysts takes their conclusions based on the past working of the company, the earnings prediction etc. [11]. When put on Machine Learning to Data of Stock, we are more concerned in Technical Analysis to understand if our algorithm can precisely learn the primary patterns in the time series of stock. This thought, Learning can also show a major role in assessing and predicting the performance of the stock and other parallel parameters supportive in Analysis of fundamental. Indeed, the most effective automated prediction of stock and recommendation structures use some hybrid analysis model relating both Technical and Fundamental Analysis

### Analysis of text for Prediction of Market Based on Financial Articles

A large amount of information generated related to the market of finance is available on the network. A greater part of this information originates from News of Financial Articles, Reports of Company and Recommendations of Expert (Blogs from legal informants can also turn as a resource of information) Utmost of this data is in format of text as supposed to numerical format which makes it difficult to use. Thus the domain of problem can now be considered as one that leads to Mining of Textual Documents and data which contains Analysis of Time Series data simultaneously. Defining the news influence on a individual stock: Positive (Optimistic), Negative (Adverse), and Neutral (Disinterested) is one of the process which has been considered. If the price of stock rises (or dips) news is considered well-thought-out to have a positive influence (or negative influence) significantly for a time period, after the news has been announced. If the price of stock does not modify vividly after the news is announced, then the news is considered as neutral. The over-all structural design is as follows. The learning model beholds for particular information in the document containing text (i.e. displays like "profits rise" inside a just announced news article, or "share prices will go down" on the blog of a expert Trader/Speculator etc. A (B.O.W) Bag-Of-Words encompassing

of Terms containing Positive Polarity and Terms containing Negative Polarity. Every time a term from the Positive Polarity transpires in an article, a weight to Positive term is assigned to that article [11]. The final score of the sentiment is calculated on the basis of the occurrences of positive and negative terms in the document.



**Figure 1: Textual Analysis Flow Diagram**

### **Model Based on Textual Analysis**

Here In this model we applied the “BOW’s” Bags of Words method to analysis of sentiment. The process identifies positive terms and negative terms within the text article. This requires a large dictionary that includes sentiment. Each term in the referred dictionary can be assigned with a weight. The final score of sentiment was produced by the model.

We will assess our model on the talk of the higher management official; text considered from the quarterly results of a leading automotive company of India .The intent of model is to quantify the opinion expressed in their quarter earnings.

#### **a) Loading of Package & Dictionaries**

[11] We are exploiting the mining of text through package named “tm” and “Rweka” in order to construct the model. We get the positive terms and the negative terms polarity by scanning the two records through the libraries and next load. To train these we have spent through former conference call records prior to the quarterly statements. Here we had also added the terms which correlate to the Motorcycle Industry which will be more helpful for more accuracy Apart from regular positive or negative terms.

#### **b) Inserting & Cleaning of Data**

The speech of CEO’s which contains the text commentary in using “corpus” function then load the document containing text (quarterly statements) . For this we have gathered the commentary text document in the folder named Text Mining [11].Subsequent step is to immaculate the text. Here all terms converted into lowercase, eliminate punctuations, eliminate numbers, and strip off the whitespace. The function facilitates us to understand the text post the sanitization.[11] We tokenize the textual data which was sanitized above. The activity of break into a stream of text into a string of words is called as Tokenization. This generates N-grams of text.[11] N-grams are essentially a set of co-occurring terms within a given terms containing text. For example, take into account this sentence “The Movie is amazing”. If  $n=2$ , then the ngrams would be:

- the movie
- movie is
- is amazing

Thereafter we construct a matrix term document (called “terms” in the program) also referred as TDM, corpus contains all the occurrences of terms which a matrix that list down.

### c) Calculation of Positive/Negative Score

[11] Further we assess if the positive/negative term polarity in the dictionary are represent in the textual document. [1] We unearth all the positive/negative terms from the text document which attained with the terms in our dictionary. The program lines compute the positive/negative sentiment score, and finally the overall polarity score.

### d) Investigational Results

The final score of sentiment polarity was 10 the model founds 14 positive terms and 4 negative terms, which is the difference between both types of terms. The management's viewpoint on quarterly results is considerable and respectable on the basis of above calculated results. The term shows certain positive/negative words that be located from the textual document on which we proceeded the model.[11] Authenticate our polarity analysis model – let's assess the quarterly result's performance numbers to check the positive polarity score spawned by our model. As it can be seen, announced a strong quarter result. EBIT growth was worthy on a robust sales volume of motorcycles. The robust results were even though the production closure for a small amount of times which was instigated by the floods undergone during the considered quarter at its production center. The technical charts on the right demonstrate the market's reaction to good results on the date of earnings result declaration. The stock initiated at around levels, made immense move moving an intraday high.

### Conclusion

The model can be made further strong and adapted in future. In time ahead we attempt to shape a model near and try to uncover other procedures to study the polarity of text.

Thus we recommended a basic polarity analysis model in R Language. **Polarity analysis** is the study of the sentiments and feelings (i.e. thoughts, reactions and opinions) which are articulated in the news information /blog columns /twitter messages etc., using processing of natural language tools. [11] The purpose is to mine data from unstructured and semi-structured data unearthed in these tweets/blogs/articles. Due to usage of computational linguistics, artificial intelligence, and computer science enables formulization of NLP. To find the polarity of text various NLP models can be used for more accuracy. In areas like Finance, Politics, Policy making, Business intelligence polarity analysis is a popular topic now a days, and has accepted widely etc.

### References

- [1] Wilson T, Wiebe J, Hoffman P. Recognizing contextual polarity in phrase-level sentiment analysis. In: Proceedings of HLT/EMNLP; 2005.
- [2] Binoy B. Nair, M. Minuvarthini, Sujithra B. , V.P. Mohandas” Stock market prediction using a hybrid neuro-fuzzy system ” International Conference on Advances in Recent Technologies in Communication and Computing, 2010
- [3] Wilson T, Wiebe J, Hoffman P. Recognizing contextual polarity in phrase-level sentiment analysis. In Proceedings of HLT/ EMNLP ; 2005
- [4] Liu B. Sentiment analysis and opinion mining. Synth Lect Human Lang Technol 2012.
- [5] Yu Liang-Chih, Wu Jheng-Long, Chang Pei-Chann, Chu Hsuan- Shou. Using a contextual entropy model to expand emotion words and their intensity for the sentiment classification of stock market news. Knowl-Based Syst 2013;41:89–97.
- [6] Michael Hagenau, Michael Liebmann, Dirk Neumann. Automated news reading: stock price prediction based on financial news using context-capturing features. Decis Supp Syst; 2013.
- [7] Tao Xu, Peng Qinke, Cheng Yinzhaoh. Identifying the semantic orientation of terms using S-HAL for sentiment analysis. Knowl-Based Syst;35: 279–89, 2012.
- [8] Maks Isa, Vossen Piek. A lexicon model for deep sentiment analysis and opinion mining applications. Decis Support System , 53:680–8, 2012.
- Pang B, Lee L. Opinion mining and sentiment analysis. Found Trends Inform Retrieval , 2:1–135, 2008.
- [9] Cambria E, Schuller B, Xia Y, Havasi C. New avenues in opinion mining and sentiment analysis. IEEE Intell Syst , 28:15–21, 2013.
- [10] QuantInsti Quantitative Learning Pvt Ltd Websites:
- [11] Nischal Puri, Aviansh Agarwal A Technique for Stock Prediction based on Textual Analysis of Financial Articles. International Journal of Emerging Trend in Engineering and Basic Sciences (IJEBS), Vol:Special Issue, 2017