American Journal of Engineering Research (AJER)
e-ISSN : 2320-0847 p-ISSN : 2320-0936
Volume-02, Issue-09, pp-265-275
www.ajer.org

Research Paper Open Access

Analyzing Public Emotion and Predicting Stock Market Using Social Media

Jahidul Arafat[1], Mohammad Ahsan Habib[2] and Rajib Hossain[3]
[1]Researcher, HT Research and Consultancy, UK.
[2]Department of Information and Communication Technology, Mawlana Bhashani Science and Technology University, Bangladesh
[3]Freelancer, oDesk, Bangladesh

Abstract: - The focus of this research was to build a cloud based architecture to analyze the correlation between social media data and the financial markets. From analytical point of view this study refurbish the viability of models that treat public mode and emotion as a unitary phenomenon and suggest the needs to analyze those in predicting the stock market status of the respective companies. With the aim to justify the correlation between social media and stock market prediction process our result reveals a proportional correspondence of pubic emotion over time with the company's market viabilities. The major significance of this research is the normalization and the conversion process that has utilized vector array list which thereby strengthen the conversion process and make the cloud storing easy. Furthermore, the experimental results demonstrate its improved performance over the factor of emotion analysis and synthesizing in the process of prediction to extract patterns in the way stock markets behave and respond to external stimuli and vice versa.

Keywords: - *Public Emotion, JSON, CSV, Hashtag.*

I. INTRODUCTION

Micro blogging is an increasingly popular form of communication on the web. It allows users to broadcast brief text updates to the public or to a limited group of contacts. In this micro blogging, stock market prediction has attracted much attention from academia as well as business. But can the stock market really be predicted? Early research on stock market prediction [1], [6], [34]-[36] was based on random walk theory and the Efficient Market Hypothesis (EMH) [4]. According to the EMH stock market prices are largely driven by new information, i.e. news, rather than present and past prices. Since news is unpredictable, stock market prices will follow a random walk pattern and cannot be predicted with more than 50% accuracy [2], [34].

However this research focused on to fetch the public emotions associating the several companies of UK and store them into a cloud. It enables the researcher to do further analysis on public sentiments to depict how the emotions change over time. It thereby assists the companies to decide on their stock market operational pattern.

This paper is organized as follows: section 2 presents literature reviews on the public emotional context while section 2.1 defines the public emotion and characteristics of tweets are described in section 2.2. Furthermore, section 3 and its subsequent sections presets technologies incorporated for sentiment management such as Query Generation and Tweet Management Technique and section 3.2 presents the Opinion Detection system that can be applied on tweets. Later section 3.3 portraits several existing software and technologies for analyzing those tweets and depicting the real cause of sentiments. Next to it is section 4 which presents the research's aim with that of the design, implementation, evaluation strategies and outcomes and both the limitations and advantages of it. Recommendation on further development is depicted in section 5. At the end, section 6 presents the degree upon which the research outcomes meet the objectives and section 7 comes out with some concluding remarks.

II. RELATED LITERATURE ON SENTIMENT ANALYSIS

An increasing number of empirical analyses of sentiment and mood are based on textual collections of public user generated data on the web. Sentiment analysis and opinion mining are now a research domain in their own which sometimes referred to as \subjectivity analysis" | whose methods and applications were extensively surveyed in much details in [10].

Different methodological approaches have been used to extract sentiment from text. Some methods are grounded in natural language processing (NLP) and rely on word constructs (n-grams) found in text to extract sentiment towards a subject (favorable or unfavorable). NLP methods have been used to extract sentiment and opinion from texts such as camera [12] and pharmaceutical reviews [11]. Other techniques of sentiment analysis, rooted in machine learning, use support vector machines (SVM) to classify text in positive or negative mood classes based on pre-classified training sets. SVM has been used to classify noisy customer feedback data [9] and movie reviews using a 5-point scale [8]. A number of hybrid methods that blend NLP and machine learning techniques have also appeared in the literature [15].

Besides the textual content discussed thus far (movie reviews, camera reviews, customer feedback), sentiment analyses have touched on many different kinds of personal online content. Personal websites such as blogs and online journals are often awash with emotive information and have been extensively used to deduct everyday happiness [10], explore trends and seasonality [14], forecast mood [17] and predict sales of books [18] and movies [20]. Some similar analytical tools operate entirely on the web: We Feel Fine5 constantly harvests blog posts for occurrences of the phrases "I feel" and "I am feeling" and provides statistics and visualizations of past and current geo-tagged mood states. A similar online site, Mood views, constantly tracks a stream of Live journal weblogs that are user-annotated with a set of pre-defined moods.

The results generated via the analysis of such collective mood aggregators are compelling and indicate that accurate public mood indicators can be extracted from online materials. Using publicly available online data to perform sentiment analyses reduces enormously the costs, efforts and time needed to administer large-scale public surveys and questionnaires. These data and results present great opportunities for psychologists and social scientists. Yet, while blogs have largely been analyzed for mood patterns, not much research has yet addressed social networking sites and micro blogging platforms. Recently, emotion has been extracted from public communication on Myspace [22], [28] and status updates on Facebook, but this study could not find any large scale sentiment analysis of Twitter, other than a study focused on micro bloggers' response to the death of Michael Jackson [13], [15]. This may be due to the fact that micro blogging and social networking sites are fairly recent forms of online communication (at least when compared to blogs).

Scale may be an issue as well. Sentiment analysis techniques rooted in machine learning yield accurate classification results when sufficiently large data is available for testing and training. Minute texts such as micro blogs may however pose particular challenges for this approach. In fact, the Twitter analysis of Jackson's death mentioned above was performed using a term-based matching technique based on the Affective Norms for English Words (ANEW). ANEW provides pre-existing, norm emotional ratings for nearly 3,000 terms along three dimensions (pleasure, arousal, dominance) [21] and has been recently employed to measure mood of song lyrics, blog posts and U.S. Presidents speeches [26]. Since it doesn't require training and testing, this syntactical approach may enable sentiment analysis for very small text data where machine learning techniques may not be appropriate. Choi and Varian [27] shows that using emoticons as labels for positive and sentiment is effective for reducing dependencies in machine learning techniques.

1.1 Defining Public Emotion

To meet the objective of the study, sentiments or emotions are defined to be of personal positive or negative feeling. Table 1 shows some examples of these.

For example, the following tweet is considered neutral because it could have appeared as a news research headline, even though it projects an overall negative feeling about General Motors: *"RT @Finance Info Bankruptcy filing could put GMon road to profits (AP) http://cli.gs/9ua6Sb #Finance"* [12].

In this research, neutral tweets are not considered for further analysis. Only positive or negative tweets has been utilized for deriving sentiment. Many tweets do not have sentiment, so those referred as tweet of question mark.

Table 1. Tweet Example

Emotion	Query	Tweet
Positive	jquery	dcostalis: Jquery is my new best friend.
Neutral	San Francisco	schuyler: just landed at San Francisco
Negative	exam	jvici0us: History exam studying ugh.

1.2 Characteristics of Tweets

Twitter messages have many unique attributes, which differentiates this research from previous research [14], [29], [30] :

1. Length: The maximum length of a Twitter message is 140 characters.
2. Data availability: Another difference is the magnitude of data available. With the Twitter API, it is very easy to collect millions of tweets.
3. Language model: Twitter users post messages from many different media, including their cell phones. The frequency of misspellings and slang in tweets is much higher than in other domains.
4. Domain: Twitter users post short messages about a variety of topics unlike other sites which are tailored to a specific topic. This differs from a large percentage of past research, which focused on specific domains such as movie reviews.

III. TECHNOLOGIES INCORPORATING THE MANAGEMNT OF SENTIMENT ANALYSIS

1.3 Query Generation and Management Technique in Twitter

1.3.1 Twitter Search API

The Twitter Search API is a dedicated API for running searches against the real-time index of recent Tweets. There are a number of important things to know before using the Search API which is explained below [24].

However, one major limitation of this Twitter API is that it is not complete index of all Tweets, but instead an index of recent Tweets. At the moment that index includes between 6-9 days of Tweets. Moreover, one cannot use the Search API to find Tweets older than about a week. Furthermore, Queries can be limited due to complexity. If this happens the Search API will respond with the error: {"error":"Sorry, your query is too complex. Please reduce complexity and try again."}. In addition to this, search does not support authentication meaning all queries are made anonymously and search is focused in relevance and not completeness. This means that some Tweets and users may be missing from search results [12-15]. That's why, Choi and Varien [27] recommends the Streaming API if research require more completeness in data. But the near operator cannot be used by this Search API. Instead the geocode parameter could be used [5], [7], [23].

The best practices of this API ensure all parameters are properly URL encoded [9]. Include a since_id when asking for Tweets. since_id should be set to the value of the last Tweet has received or the max_id from the Search API response. If the since_id is provided older than the index allows, it will be updated to the oldest since_id available. Furthermore, Java et al. [7] suggests to include a meaningful and unique User Agent string when using this method. It will help to identify the traffic when one use shared hosting and can be used by this study to triage any issues that report. However, it limits the searches to 10 keywords and operators [37].
Constructing a Query: it involves the following three basic steps: (a) Run the search on twitter.com/search. (b) Copy the URL. For example: https://twitter.com/#!/search/%40twitterapi and (c) Replace https://twitter.com/#!/search/ with http://search.twitter.com/search.json?q=. For example: http://search.twitter.com/search.json?q=%40twitterapi [32].

1.3.2 Hashtag

A hashtag is simply a relevant word or series of characters preceded by the # symbol [11]. Hashtags help to categorize messages and can make it easier for other Twitter users to search for tweets [15].
When one search for or click on a hashtag he/she will see all other tweets that use the same hashtag (see Twitter Advanced search option) [21]. Only others who are interested in the same topic thread will likely be using the same hashtag. For example, if one search for "Apple company" then "#Apple" will assit in most for having that company oriented information instead of using "Apple" [12].

1.4 Scripting language for advance searching

JSON (JavaScript Object Notation) is a lightweight data-interchange format [22]. It is easy for humans to read and write. It is easy for machines to parse and generate. It is based on a subset of the JavaScript Programming Language, Standard ECMA-262 3rd Edition - December 1999. JSON is a text format that is completely language independent but uses conventions that are familiar to programmers of the C-family of languages, including C, C++, C#, Java, JavaScript, Perl, Python, and many others. These properties make JSON an ideal data-interchange language [33].

JSON is built on two structures: (a) A collection of name/value pairs. In various languages, this is realized as an object, record, struct, dictionary, hash table, keyed list, or associative array [11, 24]. (b) An ordered list of values. In most languages, this is realized as an array, vector, list, or sequence [25]-[27].
However, the structures that the JSON offer are universal in nature. Virtually all modern programming languages support them in one form or another. It makes sense that a data format that is interchangeable with programming languages also be based on these structures [4], [16], [22].

1.5 Opinion detection in Twitter

1.5.1 Emotion Corpus Based Method

Emotion Corpus Based Method is based on vector space model for calculating document similarity. For the emotion detection in tweets, an emotion corpus that is based on 8 basic classes can be used, E= {Anger, Sadness, Love, Fear, Disgust, Shame, Joy, Surprise}[3]. Each class represents a dimension in the Boolean emotion vector of a tweet. Look for emotion words in a tweet, and if found, set the corresponding class dimension in the emotion vector to 1, otherwise it remains 0 [12], [25].

Tweet: I was on Main Street in Norfolk when I heard about tiger woods updates and it made me feel angry, on 2009-12-11. Emotion vector: (1, 0, 0, 0, 0, 0, 0, 0).

For all the tweets in a chosen time interval, a centroid of all corresponding emotion vector dimensions can be calculated. This centroid is considered as a document for each interval [8]. For a given time interval T that contains N tweets, let V ={v1, v2, . . ., vN} be a set of vectors (with l = 8 dimensions each) generated from these tweets. Define centroid $\bar{v}$ for period T as [16]:

$$\bar{v} = \left(\frac{\sum_{k=1}^{k=N} v_k^1}{N}, \frac{\sum_{k=1}^{k=N} v_k^2}{N} \ldots, \frac{\sum_{k=1}^{k=N} v_k^t}{N}\right) \quad (1)$$

After finding centroid vector for each interval, define the opinion similarity between two intervals T1 and T2 by calculating cosine similarity between their centroid vectors as suggested by [24]:

$$Sim(T_1, T_2) = \frac{\overline{v_1}.\overline{v_2}}{|\overline{v_1}||\overline{v_2}|} \quad (2)$$

1.5.2 Set Space Model

Set Space Model prescribes representing each interval by a single document which is the union of the tweets posted in that particular time interval [11]. After removing the stop words and stemming the terms using Porter stemmer 3, collect all terms in a hash set for each interval as suggested by [24]. Define the similarity between two intervals T1 and T2 by calculating Jaccard Similarity [2]:

$$Sim(T_1, T_2) = \frac{|(Set)T_1 \cap (Set)T_2|}{|(Set)T_1 \cup (Set)T_2|} \quad (3)$$

To find the changes, neither corpus based method nor the set space model alone is suitable [19]. For the corpus based method, a change in the centroid can be misleading when the interval has very few emotion words compared to its neighbors [17]. For the set space model, a change in similarity does not by itself imply an opinion change, because not all of the words are emotion words. In this method, first analyze vector space similarity as suggested by [33]. If detect a possible change, validate it by analyzing the Jaccard Similarity [31].

Tn is a time break, if the followings are satisfied in both corpus based method and set space model:

$$Sim(T_{n-1}, T_n) < Sim(T_{n-2}, T_{n-1}) \quad (4)$$

$$Sim(T_{n-1}, T_n) < Sim(T_n, T_{n+1}) \quad (5)$$

1.6 Twitter Public emotion and opinion analysis: Software and Techniques

1.6.1 Opinion Finder (OF)

Opinion Finder (OF) is a publicly available software package for sentiment analysis that can be applied to determine sentence-level subjectivity [38], i.e. to identify the emotional polarity (positive or negative) of sentences. It has been successfully used to analyze the emotional content of large collections of tweets [19] using the OF lexicon to determine the ratio of positive versus negative tweets on a given day. The resulting time series were shown to correlate with the Consumer Confidence Index from Gallup4 and the Reuters/University of Michigan Surveys of Consumers5 over a given period of time. Weadopt OF's subjective lexicon that has been established upon previous work [3], [6], [20].

Like many sentiment analysis tools OF adheres to a uni-dimensional model of mood, making binary distinctions between positive and negative sentiment [25]. This may however ignore the rich, multi-dimensional structure of human mood. To capture additional dimensions of public mood a second mood analysis tools, labeled GPOMS, that can measure human mood states in terms of 6 different mood dimensions, namely Calm, Alert, Sure, Vital, Kind and Happy can further be used [12], [17], [29].

1.6.2 Google-Profile of Mood States (GPOMS)

GPOMS' mood dimensions and lexicon are derived from an existing and well-vetted psychometric instrument, namely the Profile of Mood States (POMS-bi) [4], [22]. To make it applicable to Twitter mood analysis it can expanded the original 72 terms of the POMS questionnaire to a lexicon of 964 associated terms by analyzing word co-occurrences in a collection of 2.5 billion 4- and 5-grams6 computed by Google in 2006 from approximately 1 trillion word tokens observed in publicly accessible Web pages [6]. The enlarged lexicon of 964 terms thus allows GPOMS to capture a much wider variety of naturally occurring mood terms in Tweets and map them to their respective POMS mood dimensions. Then match the terms used in each tweet against this lexicon. Each tweet term that matches an n-gram term is mapped back to its original POMS terms (in accordance with its co-occurrence weight) and via the POMS scoring table to its respective POMS dimension. The score of each POMS mood dimension is thus determined as the weighted sum of the co-occurrence weights of each tweet term that matched the GPOMS lexicon [32]-[34].

To enable the comparison of OF and GPOMS time series it can be normalized to z-scores on the basis of a local mean and standard deviation within a sliding window of k days before and after the particular date. For example, the z-score of time series Xt, denoted ZXt ,is defined as [38]:

$$\mathbb{Z}_{X_t} = \frac{X_t - \bar{x}(X_{t\pm k})}{\sigma(X_{t+k})} \qquad (6)$$

Where, (Xt±k) and (Dt±k) represent the mean and standard deviation of the time series within the period [t−k, t + k]. This normalization causes all time series to fluctuate around a zero mean and be expressed on a scale of 1 standard deviation

The mentioned z-score normalization is intended to provide a common scale for comparisons of the OF and GPOMS time series [39]. However, to avoid so-called “in-sample” bias, Bayir [4] suggests not to apply z-score normalization to the mood and DJIA time series that are used to test the prediction accuracy of the Self-Organizing Fuzzy Neural Network.

1.6.3 Self-Organizing Fuzzy Neural Network

SOFNN has been developed specifically for regressions, function approximation and time series analysis problems [29]. Compared with some notable fuzzy neural network models, such as the adaptive-network-based fuzzy inference systems (ANFIS), self-organizing dynamic fuzzy neural network (DFNN) and GDFNN, SOFNN provides a more efficient algorithm for online learning due to its simple and effective parameter and structure learning algorithm . In some researches work, SOFNN has proven its value in electrical load forecasting, exchange rate forecasting and other applications [31]-[32].

IV. RESEARCH DESIGN

1.7 Research Aim

The aim of this research is to build a cloud based architecture to further analyze the correlation between social media data and the financial markets.

1.8 Complete Design Artifact

This design artifact presents the structure the way the research and the program will work. It presents a skeleton portraying the blocks and sequence of operations.

1.8.1 The Block Diagram

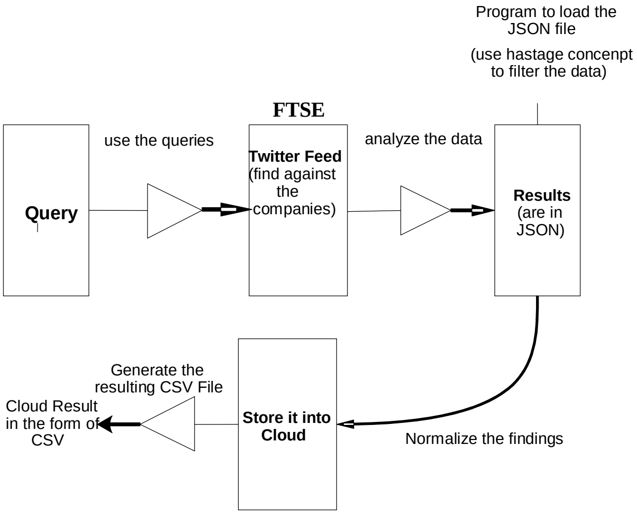

Program to load the JSON file
(use hastage concept to filter the data)
FTSE
use the queries
analyze the data
Query
Twitter Feed (find against the companies)
Results (are in JSON)
Generate the resulting CSV File
Cloud Result in the form of CSV
Store it into Cloud
Normalize the findings

Fig 1: The block diagram

This block diagram is composed of four phases. In the first phase the queries that need to be executed to find out the public mood on the corresponding companies have to be designed and need to be associated in the program. However the structure of these queries will be the same as depicted in the above sections. Once after the queries and their associated codes has been incorporated into the program, the resulting tweeter feed consisting positive, negative or question mark or default feeds will be fetched from the tweeter database. After gathering the required tweet feed, the data has to be analyzed and the results need to be generated in JSON. For that a program which will load that JSON file is preloaded a prior. After the loading and execution of that program the next step is to normalize all the findings and store into a cloud. For that the program will use a conversion technology to store the outcomes into a CSV file. However, before that when fetching down the tweets or feeds the available hashtag technology of twitter will be used which is further details in section 4.3.2.

1.9 Technologies and methods used for implementation

1.9.1 JSON

This research acknowledges the use of JSON data structures that generates the outputs based on the inputs received from queries and display and store the resultant outcomes into cloud.

1.9.2 The tweet fetching and cloud storing process

To normalize all the finding and store into a cloud this program has used a conversion technology to store the outcomes into a CSV file. However, before that when fetching down the tweets or feeds, the available hangtag technology of tweeter is also used. Moreover to identify the emotional polarity (positive or negative) of sentences lexical analysis was done using the keywords provided by tweeter where “love” for positive feeds, “hate” for negative feeds and “traffic” for question mark feeds, “Hate” or “Traffic” for negative question mark feeds, “love” or “traffic” for positive question mark feeds and “mixed” for all kinds of feeds and “Popular” for the recent and by default feeds were used. Finally the public modes of the corresponding companies were converted and stored in cloud in the form of the separate CSV file for further analysis or comparisons. This CSV file conversion process utilized vector array list for getting data and then put that data into JTable. The associated code of this tweet fetching and CSV file storing process details in the next subsequent sections.

As stated in earlier section, the design of the searching process should have the number of tweets to be searched, the number of pages on which the tweets can be displayed at maximum and the company name for which to search.

V. PROGRAM EVALUATION

The evaluation of this research has outcome with a cluster of seven steps as show below where step 1: the main GUI will be “enabled” to do the rest of the operation. Upon choosing the search option a new dialog box has been opened in step 2, where all the necessary search options i.e. number of tweets to be searched, how many pages in where results to be shown, for which company public emotions to be searched along with the category of tweets to be fetched i.e. positive, negative or question marked is enlisted. Upon selection of all these and clicking on the OK button as tweets associating that particular company has been fetched into a textbox in step 3. Next all these resultant tweets have to be saved in a CSV file. Upon choosing the save and location the file automatically converted into the CSV format in step 4. These process has been repeated for all of the enquired companies that the study focused into. The overall evaluation process of this implementation is depicted in the following series of illustration.

1.10 Survey Results

The entire survey results have been clustered around three sessions as tabulated below to show how the public emotion on a particular company of share market changes over time.

Table 2. Session Duration

Session Duration	1/04/2012 to 10/04/2012	11/04/2012 to 20/04/2012	21/04/2012 to 30/04/2012

1.10.1 Result on HSBC

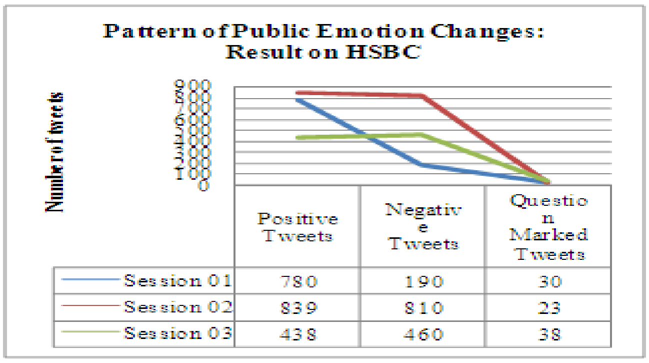
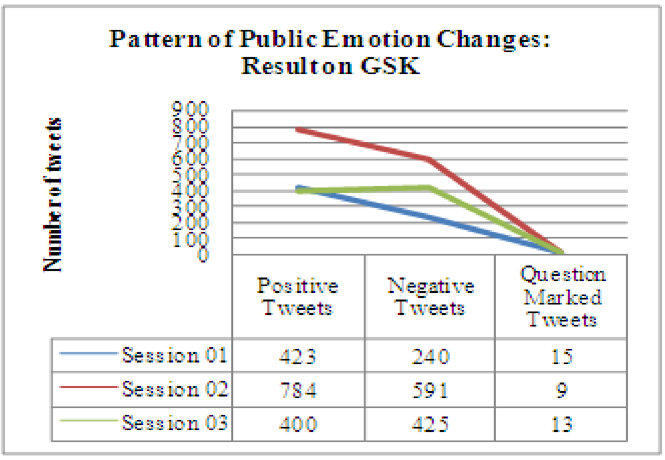


Fig 4: Pattern of public emotion change: result on HSBC

Table 3. Pattern of public emotion change: result on HSBC-Session wise breakdown

	Session 01	Session 02	Session 03
Number of Tweets	1000	1672	936
The oldest one	HSBC offers new deals to first time buyers - HSBC has launched new mortgage products for those with a 10 per...		

1.10.2 Result on GSK



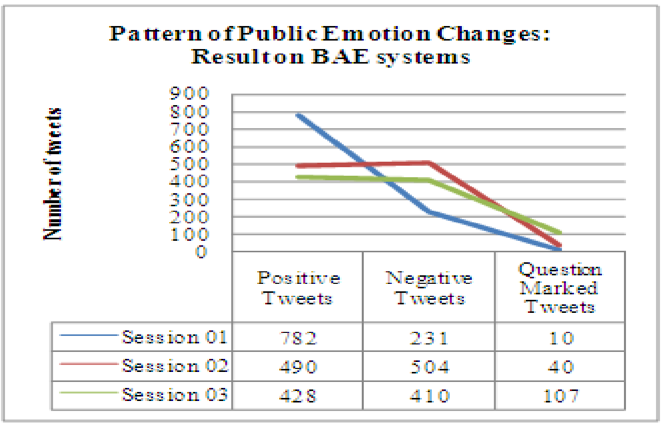
	Positive Tweets	Negative Tweets	Question Marked Tweets
Session 01	423	240	15
Session 02	784	591	9
Session 03	400	425	13

Fig 5: Pattern of public emotion change: result on GSK

Table 4. Pattern of public emotion change: result on GSK-Session wise breakdown

	Session 01	Session 02	Session 03
Number of Tweets	678	1348	837
The oldest one	**GSK** fined measly $90,000 by Argentine court for killing 14 babies in illegal vaccine trials: pewsitter.com/page_1.html#nw… **#FB**		

1.10.3 Result on BAE System



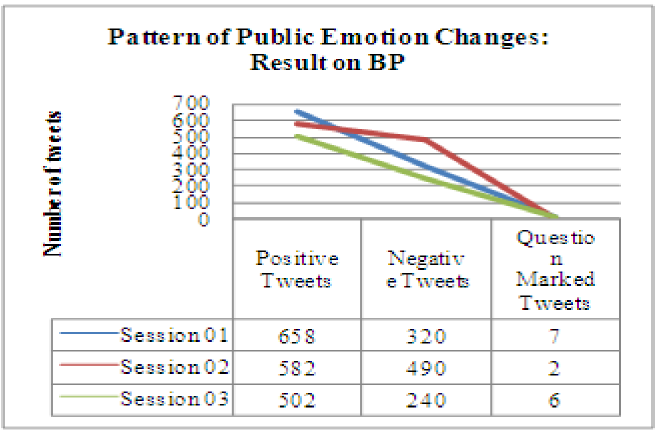
	Positive Tweets	Negative Tweets	Question Marked Tweets
Session 01	782	231	10
Session 02	490	504	40
Session 03	428	410	107

Fig 6: Pattern of public emotion change: result on BAE System

Table 5. Pattern of public emotion change: result on BAE System-Session wise breakdown

	Session 01	Session 02	Session 03
Number of Tweets	1023	1034	945
The oldest one	The Week Ahead: Results due from **BAESystems**, Cable & Wireless, Thorntons ...: Domino's Pizza will roll out more...		

1.10.4 Result on BP



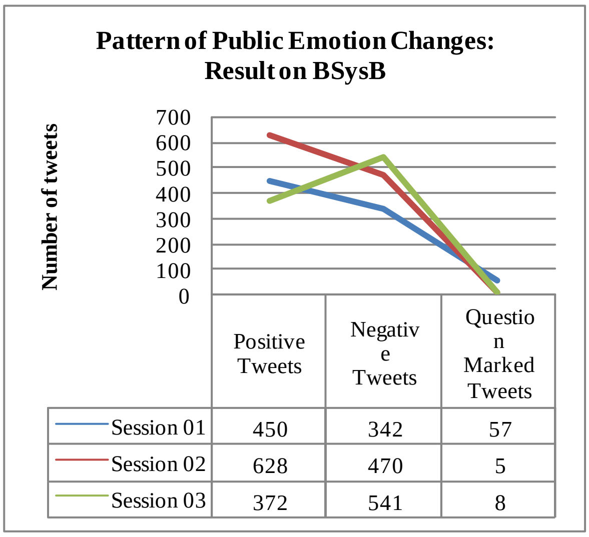
	Positive Tweets	Negative Tweets	Question Marked Tweets
Session 01	658	320	7
Session 02	582	490	2
Session 03	502	240	6

Fig 7: Pattern of public emotion change: result on BP

Table 6. Pattern of public emotion change: result on BP-Session wise breakdown

	Session 01	Session 02	Session 03
Number of Tweets	985	1074	740
The oldest one	**BP** wins exclusion of emails from oil spill trial reut.rs/wuyNMP via @Reuters		

1.10.5 Result on BSysB



	Positive Tweets	Negative Tweets	Question Marked Tweets
Session 01	450	342	57
Session 02	628	470	5
Session 03	372	541	8

Fig 8: Pattern of public emotion change: result on BSysB

Table 7. Pattern of public emotion change: result on BSysB-Session wise breakdown

	Session 01	Session 02	Session 03
Number of Tweets	849	1103	921
The oldest one	**BSysB**is one track for next business run… great news indeed….!!		

1.10.6 Summary of the Results: Impact of the results on predicting share market causes

All these results tabulated above depict how the public emotions had changed over time on a particular company. The emotion carves and charts presents the time variant fluctuations. This found to be high on positive response in some sessions but high on negative responses in other sessions. Thereby these illustrates the people's changing nature and thought regarding a particular company. The oldest tweets on each company as depicted above with that of other fetched tweets. All these portraits somehow the reasons behind all the changing emotion dimensions which is further details in section 5.3.

1.11 Main Findings

Fetching the public emotion, analyzing them and enabling those to predict the share market pattern was the prime concern of this research to undertake. The findings of this study illustrates that the mathematical prediction morphology is of significant interest for the present stock market of UK. The experiments indicate that the proposed emotion fetching and analyzing solution enables companies of UK to figure out the social impact of share market vulnerabilities. These further assist the share market leaders to do justification on their market setting while associating the public concern. These would also help the regulatory body of share market to have a holdings on the market setting by configuring market leaders social responses and activities.

1.12 Evaluation

This study has investigated whether public mood measured from large-scale collection of tweets posted on twitter.com is correlated with market prediction. The results show that changes in the public mood state can indeed be tracked from the content of large-scale twitter feeds by means of rather simple text processing techniques and that such changes respond to a variety of socio-cultural drivers in a highly differentiated manner. Changes of the public mood as that the change in the market ground of those companies found to be proportional while incorporating the social media and stock market prediction processes.

Public mood analysis from Twitter feeds on the other hand offers an automatic, fast, free and large-scale addition to this toolkit that may in addition be optimized to measure a variety of dimensions of the public mood state.

5.4 Advantages of the Research

This research has a number of potential implications i.e. here public mood is measured from large-scale collection of tweets posted on twitter.com and correlated through this designed toolkit. The normalization and the conversion process has utilized vector array list which thereby strengthen the conversion process and make the cloud storing an easy.

5.5 Limitations

The concerned time limit restricted the researcher to focus only on the following two dimensions instead of going for analyzing the whole set of stoke market pattern: (a) calculated the share fluctuations in the form of numeric from the posed tweets and (b) graphically depicting how the emotions changes over time. However, inclusion of other research dimensions could strengthen the research outcomes. Furthermore, more lexical analysis could be done on the posted tweets for robust vector analysis and clustering instead of using just the twitter factory defined keywords. Besides, when fetching the tweets it found to have some junk comments written other than using pure English grammatical syntax. In addition, limiting the scope only in twitter rather than utilizing other forms of social media invokes the further constraint to this study.

5.6 Reflection of the objective

The major objective of this research was to fetch the different dimensions of public emotion associating the stock market of UK and store them into a cloud. Along with this, the research aimed to justify whether the sentiment or emotion attached to a specific company changes over time. Execution of this program however reveals a proportional correlation with the company’s market viabilities over time. This thereby will assist the companies to evaluate their stakeholder's concerns and set their new stock market strategy while assessing the public sentiments and emotions.

VI. FUTURE WORK

This work was a short term social-economic research and many extents of it could not be explored because of time limitation. Thereby this study suggests the following strategies for the further improvement: (a) Store the tweets or feed into Array. (b) Do lexical analysis on the stored feeds to retrieve the meaning of it (c) To aid to this analysis process all the English words associating the meaning to positive, negative or neutral are recommended to be gathered and stored into three different text file such as hate.txt, question.txt and positive.txt. (e) Read those files using java FileReader function or equivalent function and bring the associating words into Array. (f) Compare those words with the each words in the comment field. (g) Finally, categorize the feeds into associating field of emotion.

Furthermore a Self-Organizing Fuzzy Neural Network can be used to train the program on the basis of past DJIA values. The public mood time series will demonstrate the ability to improve the accuracy of the most basic models to predict DJIA closing values. Given the performance increase for a relatively basic model such as the SOFNN, this study is hopeful to find equal or better improvements for more sophisticated market models which will associate new sources and a variety of relevant economic indicators. These results will have implications for existing sentiment tracking tools in which individual's evaluate the extent to which they experience positive and negative effect, happiness, or satisfaction with life. However, due to shortage of time it could not be done but suggest that this program will further be evaluated on it. However, such surveys are relatively expensive and time-consuming, and may not allow the measurement of public mood along with the mood dimensions that are relevant to assess particular socio-economic indicators.

VII. CONCLUSION

This study only focused on the stock market of UK to analyze the associating public mode and emotions. Information being fetched on those companies identified the factors and the outcomes reveals the correlation those exists among the people's emotion and market practice of a current organization.

VIII. ACKNOWLEDGMENTS

Our thanks to HTRC, a subsidiary organ of HMRC, London, UK, which has contributed towards completion of this research.

REFERENCES

[1] Li C.T., Wang C.Y., Tseng C.L. and Shou-De L. 2011. A sentiment-based audiovisual system for analyzing and displaying micro blog messages, Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Systems Demonstrations, p.32-37, June 21-21,Portland, Oregon.

[2] Guerra P.H.C., Adriano V., Wagner M.J. and Virgílio A. 2011. From bias to opinion: a transfer-learning approach to real-time sentiment analysis, Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, August 21-24, San Diego, California, USA.

[3] Agarwal A., Boyi X., Vovsha L., Owen R. and Rebecca P. 2011. Sentiment analysis of Twitter data, Proceedings of the Workshop on Languages in Social Media, p.30-38, June 23-23, Portland, Oregon.

[4] Bayir M. A., Toroslu I. H., Cosar A., and Fidan G. 2009. Smart miner: a new framework for mining large scale web usage data. In WWW, pages 161–170.

[5] Cao H., Jiang D., Pei J., He Q., Liao Z., Chen E., and Li H. 2008. Context-aware query suggestion by mining click-through and session data. In KDD, pages 875–883.

[6] Diakopoulos N. and Shamma D. A. 2010. Characterizing debate performance via aggregated twitter sentiment. In Conference on Human Factors in Computing Systems (CHI), April.

[7] Java, Song X., Finin T., and Tseng B. 2007. Why we twitter: understanding microblogging usage and communities. In Proceedings of the 9th WebKDD and1st SNA-KDD 2007 workshop on Web mining and social network analysis, pages 56–65. ACM.

[8] Jin W., Ho H. H., and Srihari R. K. 2009. Opinionminer: a novel machine learning system for web opinion mining and extraction. In KDD, pages 1195–1204.

[9] Kwak H., Lee C., Park H., and Moon S. B. 2010. What is twitter, a social network or a news media? In WWW, pages 591–600.

[10] Pang and Lee L. 2007. Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1-2):1–135.

[11] Chang C.C. and Lin C.J. 2011. LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2:27:1–27:27.

[12] Jansen J., Zhang M., Sobel K., and Chowdury A. 2009. Micro-blogging as online word of mouth branding. In CHI EA '09: Proceedings of the 27th international conference extended abstracts on Human factors in computing systems, pages 3859{3864, New York, NY, USA, ACM.

[13] Pang B. and Lee L. 2008. Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1-2):1.

[14] Dodds, P. S., and Danforth, C. M. 2009. Measuring the happiness of Large-Scale written expression: Songs, blogs, and presidents. Journal of Happiness Studies 116.

[15] Gilbert, E., and Karahalios, K. 2010. Widespread worry and the stock market. In Proceedings of the International Conference on Weblogs and Social Media.

[16] Hopkins, D., and King, G. 2010. A method of automated nonparametric content analysis for social science. American Journal of Political Science 54(1):229–247.

[17] Lindsay, R. 2008. Predicting polls with Lexicon. Available Online (accessed on: 22 January, 2012) http://languagewrong.tumblr.com/post/ 55722687/predicting-polls-with-lexicon.

[18] Manning, C. D., Raghavan, P. and Schutze, H. 2008. Introduction to Information Retrieval. Cambridge University Press, 1st edition.

[19] Mei, Q., Ling, X., Wondra, M., Su, H. and Zhai, C.X. 2007. Topic sentiment mixture: modeling facets and opinions in weblogs. In Proceedings of the 16th International conference on World Wide Web.

[20] Ounis, I., MacDonald, C. and Soboroff, I. 2008. On the TREC blog track. In Proceedings of the International Conference on Weblogs and Social Media.

[21] Pang, B. and Lee, L. 2008. Opinion Mining and Sentiment Analysis. Now Publishers Inc.

[22] Velikovich, L., Blair-Goldensohn, S., Hannan, K. and Mc- Donald, R. 2010. The viability of web-dervied polarity lexicons. In Proceedings of Human Language Technologies: The 11th Annual Conference of the North American Chapter of the Association for Computational Linguistics.

[23] Wilcox, J. 2007. Forecasting components of consumption with components of consumer sentiment. Business Economics 42(4):2232.

[24] Johan B., Huina M. and Xiaojun Z. 2011. Twitter mood predicts the stock market. Journal of Computational Science 2, p.1–8.

[25] Asur S., Huberman B.A. 2010, Predicting the Future with Social Media, http://www.arxiv.org arXiv:1003.5699 v.1.

[26] Bergsma S., Dekang L. and Goebel R. 2009. Web-scale N-gram models for lexical disambiguation, in: Proceedings of the Twenty-first International Joint Conference on Artificial Intelligence (IJCAI-09), Pasadena, CA, pp. 1507–1512.

[27] Choi H. and Varian H. 2009. Predicting the Present with Google Trends, Tech. rep., Google.

[28] Dodds P.S. and Danforth C.M. 2009. Measuring the happiness of large-scale written expression: songs, blogs, and presidents, Journal of Happiness (July).

[29] Edmans D. and García O.Y., 2007. Sports sentiment and stock returns, Journal of Finance 62 (4), 1967–1998.

[30] Frey B.S. 2008. Happiness: A Revolution in Economics, MIT Press Books, The MIT Press, June. Available Online (accessed on: 22 January, 2012) http://ideas.repec.org/b/mtp/titles/0262062771.html.

[31] Gilbert, K.K. 2010. Widespread worry and the stock market, in: Fourth International AAAI Conference on Weblogs and Social Media, Washington, DC, pp. 58–65.

[32] Mao H., Zeng X.J., Leng G., Zhai Y. and Keane J.A. 2009. Short and mid-term load forecasting using a bilevel optimization model, IEEE Transactions On Power Systems 24 (2), 1080–1090.

[33] O'Connor, Balasubramanyan R., Routledge B.R. and Smith N.A. 2010. From tweets to polls: linking text sentiment to public opinion time series, in: Proceedings of the International AAAI Conference on Weblogs and Social Media, Washington, DC, May.

[34] Pak, P.P. 2010. Twitter as a corpus for sentiment analysis and opinion mining, in: N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, D. Tapias (Eds.), Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10), European Language Resources Association (ELRA), Valletta, Malta, May, pp. 19–21.

[35] Pang and Lee L. 2008. Opinion mining and sentiment analysis, Foundations and Trends in Information Retrieval 2 (1–2), 1–135.

[36] Prechter R.R. and Parker W.D. 2012, The financial/economic dichotomy in social behavioral dynamics: the socioeconomic perspective, Journal of Behavioral Finance 8 (2) (2007) 84–108.

[37] Schumaker R.P. and Chen H. 2009. Textual analysis of stock market prediction using breaking financial news, ACM Transactions on Information Systems 27 (February (2)) (2009) 1–19.

[38] Zhang X., Fuehres H. and Gloor P.A. 2010, Predicting Stock Market Indicators Through Twitter I Hope It is Not as Bad as I Fear, Collaborative Innovation Networks (COINs), Savannah, GA.

[39] Zhu X., Wang H., Xu L. and Li H. 2008, Predicting stock index increments by neural networks: the role of trading volume under different horizons, Expert Systems with Applications 34 (4) 3043–3054.