IIOAB JOURNAL
ISSN: 0976-3104

ARTICLE

ANALYZING COMPANY'S STOCK PRICE MOVEMENT USING PUBLIC SENTIMENT IN TWITTER DATA

Mythili Thirugnanam*, Smit Patel , Prakhar Vyas, Tamizharasi Thirugnanam

School of Computer Science and Engineering, VIT University, Vellore, INDIA

ABSTRACT

Recently some efforts have been made to use data from social media for prediction in various domains. Some studies also suggest public emotions in tweets may correlate with market trend. Using studies observations, public sentiments from social media helps to predict stock price of a particular company. The system can be effectively used to predict stock price movement of particular company or not. However, the current systems use to predict the overall market trend instead of predicting for individual company. Also, the public sentiment is not only the single factor which can affect the stock market so, there is need of combination of market and public sentiment to predict the stock price of the company. The system firstly mine the tweets of a particular company from different sources such as Twitter and Yahoo finance and news related to finance are also considered. The data is preprocessed to counter the noisy and missing data and sentiment of the public data of company is calculated using the natural language processing techniques. Financial data values of company will be fetched from yahoo finance. Some classifiers techniques such as SVM, Naïve Bayes and Maximum Entropy are also tested to find the best outcome. The public sentiment and market are sentiments are combined to obtain the desired outcome. The correlation between public's sentiment and company's stock price movements is observed. Granger causality test is used to determine whether sentiment polarity is able to predict the stock price in advance for a company.

KEY WORDS
Social web mining; Data mining; Sentiment analysis; Natural language processing; Market prediction.

Published: 12 Sept 2016

***Corresponding Author**
Email:
tmythili@vit.ac.in

INTRODUCTION

Twitter is platform for social interaction. It is an online application for sharing and gathering small messages. These messages can represent everything from a person's opinion of shoes, to the latest changes in the financial market or pictures from Olympics games. The gist of the Twitter is the tweets [1]. The Tweet is a message limited to the 140 character. Tweets let us interact with other people, share photos with people and post any relevant or non-relevant type of information. The small size of the tweets are not a barrier for the flow of information. The fast and rapidly gained worldwide popularity of Twitter, with more than 100 million users posting 340 million tweets a day in 2012 [2]. The service also handled on an average 1.6 billion search queries every day. In 2013, Twitter was one of the ten most-visited websites and has been described as "the SMS of the Internet"[3]. As of May 2015, Twitter acquired little more than 500 million users, 332 million users are frequently visitor of the service.

Economic growth plays a vital role in any country, as it increases job opportunities and income and in turn increases revenue. For the economic development of a country, it is necessary that the investments are in the right direction which yields high return. The capital market provides economic growth, it is made up of mechanism that encourages the formation of savings and are easily accessible to those wishing to invest. The stock market is good channel for the productive activities. However, the pricing of a share on the stock exchange has very dynamic behavior, driven often by the law of supply and demand for action. This dynamism attracts the attention of investors because it provides huge profits when investment is made in the best way at the right time. The objective in share market is to sell shares when price is high and buy shares when price is low. This is the way to generate profits and reduce risk and losses. The technique can also be referred to as prediction; can enhance the profitability of investments.

As the advancement in technologies take there had been developed an enabling environment for widespread use of social networks. Social network is a place where individual can freely express their desires, feelings and preferences, thus making the result of this process, more natural and close to reality. The use of individual may be the basis for establishing, certain preferences and calculating the mood for prediction. Twitter, one of the most popular social media, with over 200 million users who share their opinion through tweets.

Keeping in mind the large amount of available information, resulting from Twitter, the content of messages about certain company in order to establish a relationship between the collective mood of publications and their influence on the price of an asset in the financial market can be analyzed.

RELATED WORK

The generation of huge amount of online documents and text generated by user has led to a growth in research in the area of sentiment analysis and its association with financial segments. A broad analysis of some of the machine learning techniques used in sentiment analysis is described by B.Pang et al. [4]. There, authors presented methods in determining whether movie reviews are positive or negative and investigate some challenges in sentiment analysis. An overview of Support Vector Machine, Maximum Entropy, and Naive Bayes classification techniques in the movie review domain is provided.

IIOAB JOURNAL
ISSN: 0976-3104

Pak A. et al. [5] discusses the need of Twitter especially as a corpus for sentiment analysis. They confer the methods of gathering tweets and then processing tweets. The authors use specific emoticons to form a training set for sentiment classification, this technique helps in reducing manual tweet tagging. The training set was split into positive and negative samples on the basis of happy and sad emoticons. Additionally, a few accuracy improvement methods were analyzed.

Similarly, Albert B. et al. [6] also presented a discussion of streaming mining Tweet and sentiment extraction, while facilitating the discussion to include opinion mining. Bollen J. et al. [7] proposed one of the first hints that there might be a correlation between Twitter sentiment and the stock market. In their work, a sentiment value is correlated with the DJIA and then provided to a neural network to predict market movements. Opinion Finder tool was used by authors to track mood and to classify mood in 6 dimensions (Kind, Sure, Happy, Calm, Alert and Vital). Then, correlated the mood time series with DJIA closing values by using a Self-Organizing Fuzzy Neural Network. Their techniques measured an improvement on DJIA prediction accuracy.

A few other works in the area have been proposed. Eric Brown [8] suggests a method in which tweets are extracted according to stock symbols and then sentiment is gathered using Naive Bayes over a dataset already existing. Deng S. et al. [9] combines sentiment gathered from news with technical features of stock data. Sentiment is analyzed using SentiWordNet 3.0, a lexical resource used in opinion mining. The data was scraped from Engadget rather than Twitter. Multiple kernel learning is tested to find the best coefficients for the different parameters in their system for prediction of stock. The authors concluded that training a model based on more technical stock factors improved stock prediction performance.

Ronen Feldman [10] experimented 3 methods of sentiment analysis on tweets: aspect-based, sentence-level, and document-level. They found that aspect-based analysis, which finds the sentiment of context and also determine the aspect it refers to, is the fine-grained technique for the problem. Additionally, they provided various challenges of sentiment analysis on tweets.

Mittal A. et al. [11] achieves a 75% accurate prediction rate using Self Organizing Fuzzy Neural Networks on Twitter data and DJIA closing values. In their research, they created a custom dataset with words to calculate sentiment value for their tweets.

On a side note, E. Hurwitz [12] discusses some common issues associated in many of the works presented above. These include, but are not limited to: data not sufficient always, measures of performance are inaccurate and inappropriate scaling.
Go and L. Huang [13] implemented a sentiment analyzer for twitter data. For training data the tweets with emoticons were considered and able to divide tweets into two sections either positive or negative. Applied SVM, Naïve Bayes, and Max entropy techniques on the data.

Table 1. Study of different sentiment classifier techniques and stock market prediction

Authors	Description	Pros	Cons
Alexander Pak, Patrick Paroubek (2010) [5]	Automatic collection of a corpus that can be used to train a Sentiment classifier. Used TreeTagger for POS tagging and observed the difference in distributions among positive, negative and neutral sets. The classifier used the multinomial Naıve Bayes classifier that uses N-gram and POS-tags as features.	Classifier was able to determine positive, negative and neutral sentiments of documents.	Collected only datasets having emoticons. Not applicable for multilingual sentiment classifier. Limited dataset.
Go, Bhayani and L.Huang (2009)[13]	used emoticons as noisy labels for training data to perform distant supervised learning. Machine learning algorithms (Naive Bayes, maximum entropy classification, and support vector machines) method were used. Twitter messages have unique characteristics compared to other corpora, machine learning algorithms	Accuracy improved for Naive Bayes (81.3% from to 82.7%) and Max-Entropy (from 80.5 to 82.7) for using combination of Unigrams and Bigrams.	A set of 177 negative tweets and 182 positive tweets were manually marked. Not all the test data has emoticons. Semantics not considered in classifying tweets.

	are shown to classify tweet sentiment with similar performance		
Apoorv Agarwal, Boyi Xie, Ilia Vovsha, Owen Rambow and Rebecca Passonneau (2011) [14]	Used POS-specific prior polarity features and use of a tree kernel to obviate the need for tedious feature engineering. Tree kernel and feature based models were used.	An overall gain of over 4% for two classification tasks: a binary, positive versus negative and a 3-way positive versus negative versus neutral. POS-specific prior polarity features and tree kernel outperform the Naïve base.	Linguistic analysis, for example, parsing, semantic analysis and topic modeling were not considered. Feature based model applied with less data.
Paolo Fornacciari, Monica Mordonini and Michele Tomaiuolo (2014) [15]	Variations of Naive Bayes classifiers for detecting polarity of English tweets. Two different variants of Naive Bayes classifiers were built namely Baseline (trained to classify tweets as positive, negative and neutral), and Binary (makes use of a polarity lexicon and classifies as positive and negative. Neutral tweets neglected.	Inclusion of new features for classifier such as Lemmas (nouns, verbs, adjectives and adverbs), Polarity Lexicons, and Multiword from different sources and Valence Shifters.	Highlighted the typical problems of Sentiment Analysis (irony, sarcasm, lack of information, etc.). some peculiar problems of the considered channel were also detected
Luciano Barbosa and Junlan Feng (2010) [16]	A two phase automatic sentiment analysis method for classifying tweets. Classified tweets as objective or subjective and then in second phase, the subjective tweets were classified as positive or negative.	The new features used were retweets, hashtags, link, punctuation and exclamation marks in conjunction with features like prior polarity of words and POS. Creates a more abstract representation of tweets.	Failed to deal with the cases of sentences that contain antagonistic sentiments. Not able to perform a more fine grained analysis of sentences in order to identify its main focus.
Albert Bifet and Eibe Frank (2010)[6]	Used Twitter streaming data provided by Firehouse API, which gave all messages from every user which are publicly available in real-time. Experimented multinomial naive Bayes, stochastic gradient descent, and the Hoeffding tree.	Introduced SGD which provides an efficient means to learn some classifiers even if they are based on non-differentiable loss functions, such as the hinge loss used in support vector machines. SGD-based model, when used with an appropriate learning rate provides better result.	Not able to use methods in real time and not considered other features available in Twitter data streams, such as geographical place, the number of followers or the number of friends.
Dmitry Davidov, Ari Rappoport and Oren Tsur (2010) [23]	Proposed an approach to utilize Twitter user-defined hastags in tweets as a classification of sentiment type using punctuation, single words, n-grams and patterns as different feature types, which are then combined into a single feature vector for sentiment classification. Used of K-Nearest Neighbor strategy to assign sentiment labels by constructing a feature vector for each example in the training and test set.	The framework successfully identifies sentiment types of untagged sentences. The quality of the sentiment identification was also confirmed by human judges. Also dependencies and overlap between different sentiment types represented by smileys and Twitter hashtags were explored.	Predefined tags and smileys were used for the classification of tweets. Not enough dataset to apply over large scale. Not applicable for other domains.
Po-Wei Liang and Bi-Ru Dai (2013) [22]]	Used Twitter API to collect twitter data. The training data falls in three different categories (camera, movie , mobile). The data is labeled as positive, negative and non-opinions. Unigram Naive Bayes model was implemented and the Naive Bayes simplifying independence assumption was employed. Finally, the orientation of an tweet is predicted. i.e. positive or negative.	Tweets containing opinions were filtered. Eliminated useless features by using the Mutual Information and Chi square feature extraction method. Machine learning performed well in the classification of sentiments in tweets with good accuracy.	Not able to handle issues such as semantics of tweet, also not able to address issues regarding domain-specific tweets and the neutral tweets.
Peter D. Turney (2002)[21]	Used bag-of-words method for sentiment analysis in which the relationships between words was not at all considered and a document is represented as just a collection of words. To determine the sentiment for the whole document, sentiments of	The algorithm attains an average accuracy of 74%. For banks and automobiles, it seems that the whole is the sum of the parts, and the accuracy is 80% to 84%.Travel reviews	Only able to give recommendation whether it is considerable or not. The time required for queries was more and, for some applications, the level of accuracy that was

	every word was determined and those values are united with some aggregation functions.	are an intermediate case. For movie reviews accuracy is around 60%.	achieved was not good as compared to others.
Rui Xia, Chengqing Zong and Shoushan Li (2011)[20]	Used an ensemble framework for Sentiment Classification which is obtained by combining various feature sets and classification techniques. Used two types of feature sets (Part-of-speech information and Word-relations) and three base classifiers (Naive Bayes, Maximum Entropy and Support Vector Machines) Applied ensemble approaches like fixed combination, weighted combination and Meta-classifier combination for sentiment classification and obtained better accuracy.	Integrated different feature sets and classification algorithms to boost the overall performance. Also pursued hybrid generative/discriminative models that are suitable for sentiment classification.	Computational cost is increased. Also the computational time for the method is slightly higher than the baseline methods.
Zhiang Hu, Jian Jiao, Jialu Zhu (2014)[19]	Found the relationship between tweets of one important Twitter user and the corresponding one stock price behavior. The tweets of Elon Musk, who is the CEO of Tesla, and the change of Tesla stock price are used as data. They tried different sets of features using SVM model.	The accuracy and the confusion matrix of this set of features and labeling are reasonable. Around 60% accuracy can be reached if we leave 10% data to be testing set.	Limited to a single user. Also the method is not able to give up the overall scenario of the market.
Anshul Mittal and Arpit Goel (2011)[11]	Applied sentiment analysis and machine learning principles to find the correlation between "public sentiment" and "market sentiment". Used twitter data to predict public mood and use the predicted mood and previous days' DJIA values to predict the stock market movements	Concluded that public mood can indeed be captured from the large-scale Twitter feeds by means of simple natural language processing techniques, as indicated by the responses towards a variety of socio-cultural events during the year. a Self Organizing Fuzzy Neural Network performs very good in predicting the actual DJIA values when trained on the feature set consisting of the DJIA values, Calm mood values and Happiness dimension	Dataset doesn't really map the real public sentiment, it only considers the twitter using, English speaking people. Not able to obtain high percentage result of 87%, but obtained 75.56% result using k-fold sequential cross validation.
LI Bing, Keith C.C. Chan and Carol OU (2014)[18]	Proposed extracting ambiguous textual tweet data through NLP techniques to define public sentiment, then make use of a data mining technique to discover patterns between public sentiment and real stock price movements.	The proposed algorithms have a better prediction performance in some certain industries such as IT and media. On the other hand, our study indicates the proposed algorithms have a better performance in using current tweets sentiment to predict the stock price of three days later.	Considered the daily or weekly closing values of the stock price only. The dataset used is the Twitter data. Most Twitter messages are very short and some of them are actually meaningless so not able to give clear picture.
Tushar rao and Saket Srivastava (2012)[17]	Worked upon identifying relationships between Twitter based sentiment analysis of a particular company/index and its short-term market performance using large scale collection of tweet data.	Individual company stocks gave strong correlation values (upto 0.88 for returns) with twitter sentiment Features of that company. It is no surprise that this approach is far more robust and gives far better results (upto 91% directional accuracy) than any previous work.	Discussion based tracking of Twitter sentiments which may be biased or misleading. Also more data is required to make good predictions.

From the study of different works it is evident that there is need of a system which can predict and analyze the company's stock price. The study shows that the current systems can predict the overall market trend using the public sentiment. So, the system proposed should be such that it should be based on the individual company. There is a need of system which is capable of mining the company specific tweets from Twitter. The systems which take individual company data and apply all the classifier techniques and find out which one is more suitable is required. But the

IIOAB JOURNAL
ISSN: 0976-3104

technique used by the system must be fast and the accuracy of the technique should be high. The system should also be able to clearly distinguish between the positive, negative and neutral sentiments. The intended user of the system can be any one from stock analysts to company who want to use this as source of improvement. The system must be simple yet elegant and easily adaptable and easy to operate by mass group. Combination of the market and public sentiment must be taken into account as it can help in analyzing better.

PROPOSED FRAMEWORK

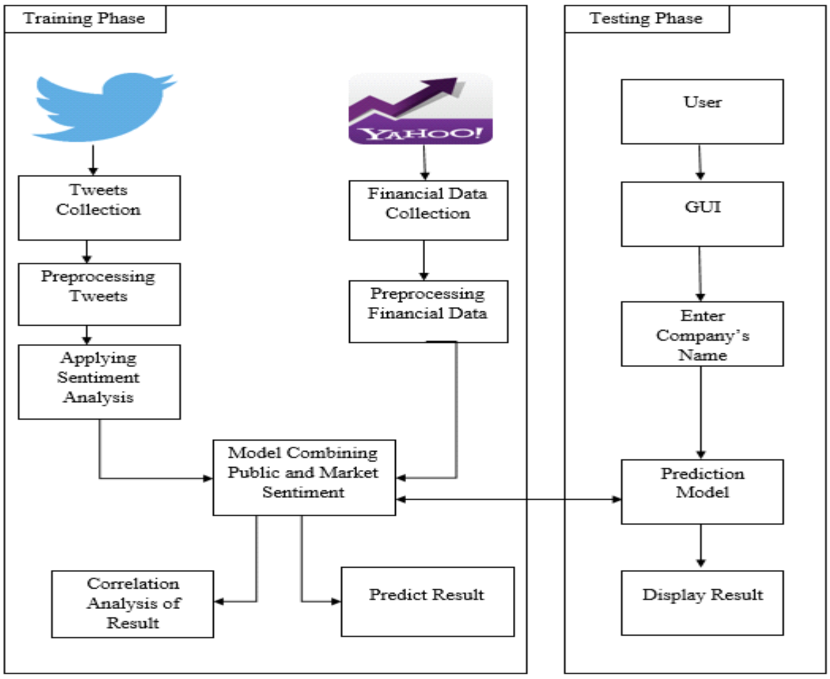

Fig. 1: Schematic view of developed system

In this system, first data is collected i.e. tweets from Twitter with the help of twitter API of Python and financial news and messages from Yahoo from Yahoo API of Python. Then preprocessing of collected data by removing noisy data, redundancy and other unwanted data. Then sentiment analysis is performed using natural language processing technique which first calculate polarity of individual word and then combine them to get polarity of whole sentence. This polarity value of each sentence is the averaged to get the whole sentiment value of the Tweet. The system then combines both sentiments of Twitter and Market. And as a consequence, the predicted value of the stock price of a particular company which the user has queried to the system is obtained. For the better visualization of system a smooth graph is drawn. To analyze the system and find the accuracy of the system correlation, regression, granger causality tests are performed. Also, the accuracy of the system is calculated and obtained the value of 89.4%. The design of the system and different requirements are explained below:

Dataset Collection

To train the system collects the dataset of tweets and make corpus of them. Also, the dictionary of words are needed which will be used to find the polarity of a given word.

So, the tweets on the basis of three different categories positive, negative or neutral were gathered. SentiWordNet will be used for the word corpus. It contains roughly around 8221 words with their polarity values according to the context of the word used. Thus this can be used for calculating the sentiment of the text provided.

Twitter API

For the collection of the tweet, Twitter contributes a rather powerful API. There are two ways to collect tweets: the Search API is a part of REST API which allows user to request the specific queries foe getting tweets. REST, simply uses HTTP methods to access the tweets.

Another one is Streaming API which allows user to obtain real time tweets just by entering the specific query. It can be done by establishing the connection between user and the Twitter to gain the stream of tweets. The rate limit is also not limited in Streaming API which is the case in Search API.

The requested tweets are returned in the JSON object format and contain tweets and the metadata associated with them. For our purpose we will store the time, location and the message part of tweet.
The Streaming API can be accessed by entering the unique API key and once it is authenticated the user can interact with Twitter and extract the required tweets.

Tweets Preprocessing

The text part of the tweet can contain many additional words which will not contribute to the sentiment of the tweet. They may contain URLs, retweets, other user's tags or symbols with no sense. To get the exact sentiment of the tweet we need to filter the noise from the original text.
The steps involved in the preprocessing are: -

- Tokenization: The first step is to form the collection of words from the text by splitting the text by spaces. The collection of words will we called as bags of word and used for future purpose.
- Removing stop words: Next step is to filter out the stop words. This can be done using the python library which has the list of stop words and each word is tested against it and if it is matched then removed from the text. These words will not contribute to the sentiment of the text.
- Twitter symbols: The final step is to symbols in tweets. The word immediately following @ will be removed as it will be username and have no sentiment value. The word after # will be considered as they can be used in filtering purpose. The URLs will also be removed from the text with the help of the regex.

Yahoo Finance

Yahoo finance which can be used to obtain every information about the finance. Historical data for the company's share can be obtained from it. The current stock price for company can also be extracted from this. Also, it is the great source for obtaining finance news and major developments going around in financial world. Yahoo finance board can also be used for the prediction of the stock price.

Ystockquote

It is used for obtaining the stock values of the company. It also provides options for giving the historical data.

Classifier Used

Naïve Bayes

A Naïve Bayes classifier is based on Bayes Rule and is type of probabilistic classifier and is one of the simplest form of the Bayesian Network. The classifier is very easy to implement and have wide applications. The underlying naïve assumption based on which classifier operates is conditional independence of each feature from its feature set.
The classifier is a classical application of Bayes Rule:
According to our view, we want a class "c" so we need to find the most probable class with given features "F". The value P(F) can be treated as constant and it will be independent of c and values provided by us. Now we need to calculate the value of P(F|c) and we can assume that a given class "c_j" the features are conditionally independent from each other. Therefore, From this we can find the label c* with maximum posterior decision rule which will take the most probable label from all label of C.

Sentiment Calculator

The sentiment for each and every tweet obtained will be done through this. The sentiment calculator is to identify the tweets in three type positive, negative or neutral tweets. The sentiment calculates the sentiment of each and every word and then it will move to the tweet level and finally the overall sentiment of company is obtained.

Prediction Model

The model is to combine the public and market sentiment. It will be used to merge the data obtained from the yahoo finance and the sentiment calculated of the tweets of a particular company. The model will give the prediction of the stock price of a company using twitter data and historical financial data.

RESULTS

Different tests were performed to evaluate the efficiency and performance of the system. Correlation test to find out whether the system predicted value and actual value are strongly related or not. Regression test was performed to find the relationship between the actual and predicted stock values.
To analyze result the confusion matrix is taken into account and discussed in Table 2. The confusion matrix is denoted by:

Table 2. Confusion Matrix

		Predicted Value	
		Positive	Negative
Actual Value	Positive	TP	FP
	Negative	FN	TN

Where TP is true positive, FP is false positive, FN is false negative and TN is true negative. The accuracy of the same is discussed in Table 3.

Table 3. Confusion Matrix with experimented value

		Predicted Value	
		Up	Down
Actual Value	Up	268	31
	Down	22	179

On applying the confusion matrix on the experiment performed, and applying the accuracy obtained is 89.4

The above data is obtained for the 500 sample points. The samples are the stock price of a particular company on a single day. The assumptions for calculating the accuracy is that if the stock price is moving up and the system predicts the upward movement then it is considered as true positive. Similarly, if the movement is downward and the system also predicts downward movement then it is true negative. If the movement is upward but prediction is downward then it is considered as false positive and in reverse case false negative is obtained.

The above values are obtained by analyzing the several companies over period of days shown in Figure 2 to Figure 7.

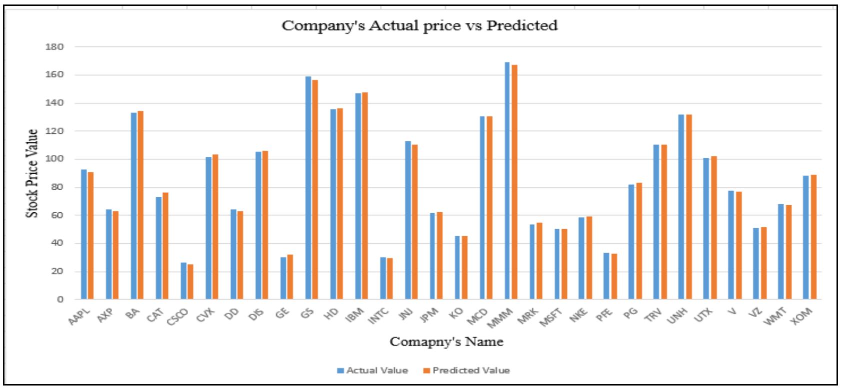

Fig. 2: Actual Vs Predicted price of different company's 1-day stock

The above graph depicts the stock price values of several companies. The graph contains the actual stock price value and also the previously predicted values of different companies.

Also, the more analysis is done with few companies.

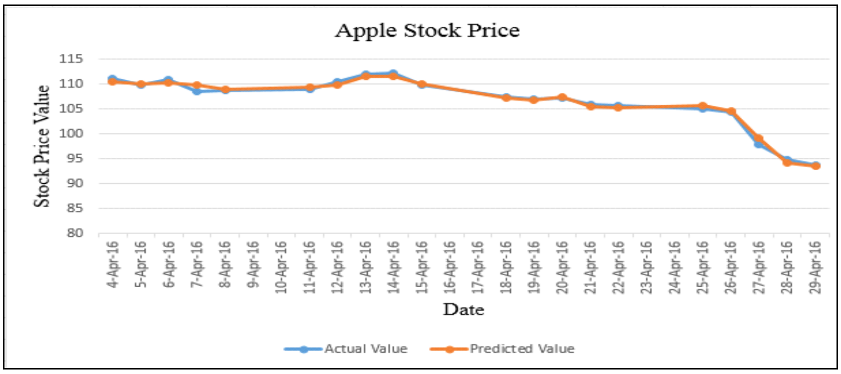

Fig. 3: Actual Vs Predicted price for Apple stock over 20 days

The above graph is obtained is showing the movement of Apple stock price over 20 days and the actual value of stock along with the predicted stock value is shown. The movement is captured and actual vs predicted values are displayed over span of days.

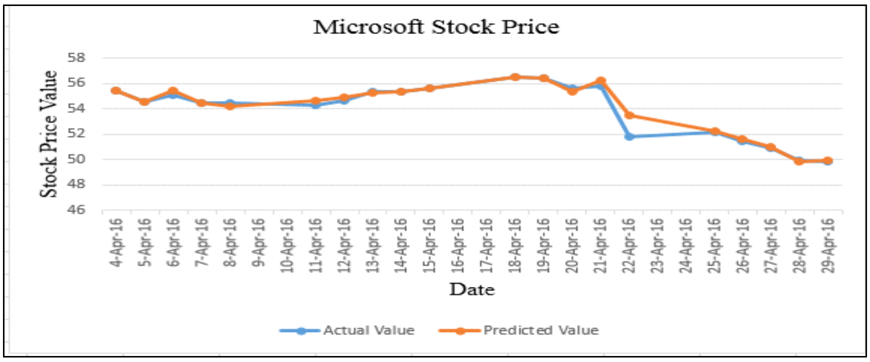

Fig.4: Actual Vs Predicted price for Microsoft stock over 20 days

The above graph is obtained is showing the movement of Microsoft stock price over 20 days and the actual value of stock along with the predicted stock value is shown. The movement is captured and actual vs predicted values are displayed over span of days.

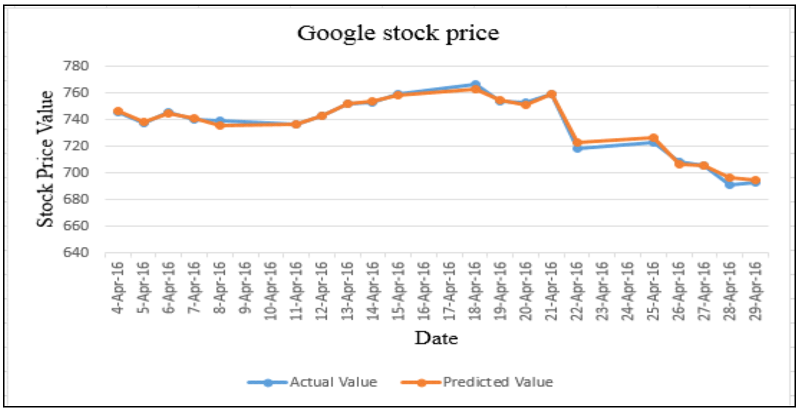

Fig. 5: Actual Vs Predicted price for Google stock over 20 days

The above graph is obtained is showing the movement of Google stock price over 20 days and the actual value of stock along with the predicted stock value is shown. The movement is captured and actual vs predicted values are displayed over span of days.

IIOAB JOURNAL ISSN: 0976-3104

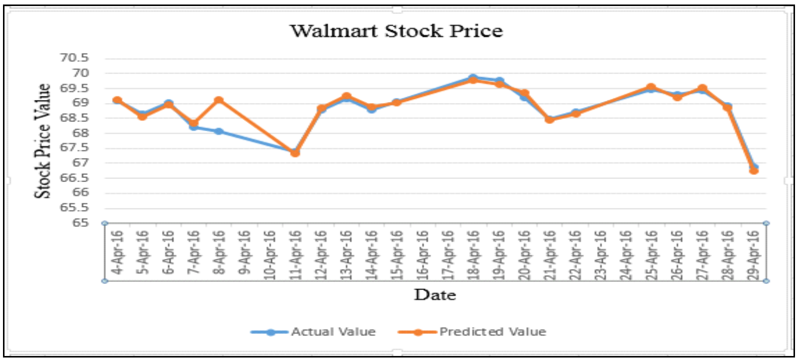

Fig. 6: Actual Vs Predicted price for Walmart stock over 20 days

The above graph is obtained is showing the movement of Walmart stock price over 20 days and the actual value of stock along with the predicted stock value is shown. The movement is captured and actual vs predicted values are displayed over span of days.

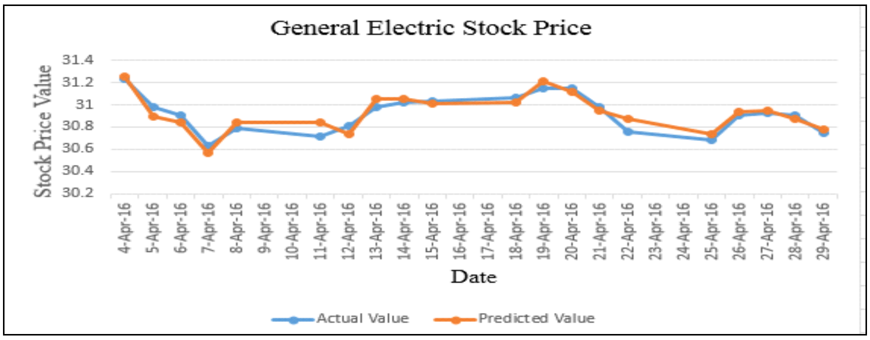

Fig. 7: Actual Vs Predicted price for General Electric stock over 20 days

The above graph is obtained is showing the movement of General Electric stock price over 20 days and the actual value of stock along with the predicted stock value is shown. The movement is captured and actual vs predicted values are displayed over span of days.

CONCLUSION

Camel The main essence of the project was sentiment analysis and applying on the public data obtained from Twitter and other financial news sources and combining with the market values to obtain the future price of the market. The Streaming API and mining of the tweets looks like a challenge at first, but after few initial try's tweets were obtained from the source efficiently. The data obtained from this is good for the analysis but it can be improved.

The use of Naïve Bayes classifier and other had produced better than the baseline methods and had very good accuracy and performed well with the data. The sentiment analysis of the text was a tedious job but was performed efficiently and time required was also less. The usage of the corpus with the polarity value helped a lot in calculating the overall sentiment value of the text. Combining the market and public sentiment to obtain the future price was used. The method was in many ways successful but the method can be improved in many ways for getting the better results. Thus, we were successful in analyzing stock price movement of company's stock and predict the future value.

CONFLICT OF INTEREST

There is no conflict of interest.

ACKNOWLEDGEMENTS

The authors are thankful to Hon'ble Dean and Management

FINANCIAL DISCLOSURE

No financial support was received to carry out this project..

IIOAB JOURNAL
ISSN: 0976-3104

REFERENCES

[1] https://twitter.com/about

[2] http://en.wikipedia.org/wiki/Twitter

[3] Leslie D'Monte, "Swine Flu's Tweet Tweet Causes Online Flutter," Business Standard, April 29, 2009, http://www.business-standard.com/india/news/ swine-flu%5Cs-tweet-tweet-causes-online-flutter/356604/

[4] B. Pang, L. Lee, and S. Vaithyanathan. "Thumbs up? *Sentiment classification using machine learning techniques*", 2002, pages 79-86.

[5] Alexander Pak and Patrick Paroubek. "Twitter as a corpus for sentiment analysis and opinion mining". 2010 *Proceedings of LREC*.

[6] Albert Bifet and Eibe Frank. *"Sentiment knowledge discovery in twitter streaming data"* 2011.

[7] J. Bollen and H. Mao. "*Twitter mood as a stock market predictor*". *IEEE Computer*, 2010, 44(10):91-94.

[8] Eric D. Brown. "Will twitter make you a better investor? *a look at sentiment, user reputation and their effect on the stock market."* 2012 *SAIS Proceedings*.

[9] Shangkun Deng, Takashi Mitsubuchi, Kei Shioda, Tatsuro Shimada, and Akito Sakurai. "*Combining technical analysis with sentiment analysis for stock price prediction." In Proceedings of the 2011 IEEE Ninth International Conference on Dependable, Autonomic and Secure Computing, pages 800-807, EEE Computer Society, Washington, DC, USA,. DASC '11 IEEE Computer Society.*

[10] Ronen Feldman. "*Techniques and applications for sentiment analysis." Communications of the ACM*, 2013, 56(4):82-89.

[11] Anshul Mittal and Arpit Goel. "*Stock prediction using twitter sentiment analysis*" 2012.

[12] E. Hurwitz and T. Marwala. "*Common mistakes when applying computational intelligence and machine learning to stock market modelling*. 2012

[13] Go, R. Bhayani, L.Huang. *"Twitter Sentiment Classification Using Distant Supervision". Stanford University, Technical Paper*. 2009

[14] Agarwal, A., Xie, B., Vovsha, I., Rambow, O. and Passonneau, R., 2011, June. Sentiment analysis of twitter data. In *Proceedings of the workshop on languages in social media* (pp. 30-38). Association for Computational Linguistics.

[15] Fornacciari P, Mordonini M, Tomaiuolo M. A Case-Study for Sentiment Analysis on Twitter.

[16] Barbosa, L. and Feng, J., 2010, August. Robust sentiment detection on twitter from biased and noisy data. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters* (pp. 36-44). Association for Computational Linguistics.

[17] Rao, T. and Srivastava, S., 2012, August. Analyzing stock market movements using twitter sentiment analysis. In *Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012)* (pp. 119-123). IEEE Computer Society.

[18] Bing, L., Chan, K.C. and Ou, C., 2014, November. Public sentiment analysis in Twitter data for prediction of a company's stock price movements. In *e-Business Engineering (ICEBE), 2014 IEEE 11th International Conference on* (pp. 232-239). IEEE.

[19] Hu, Z., Jiao, J. and Zhu, J., 2014. Using Tweets to Predict the Stock Market

[20] Xia, R., Zong, C. and Li, S., 2011. Ensemble of feature sets and classification algorithms for sentiment classification. *Information Sciences*,*181*(6), pp.1138-1152.

[21] Turney, P.D., 2002, July. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In*Proceedings of the 40th annual meeting on association for computational linguistics* (pp. 417-424). Association for Computational Linguistics.

[22] Liang, P.W. and Dai, B.R., 2013, June. Opinion mining on social media data. In *2013 IEEE 14th International Conference on Mobile Data Management* (Vol. 2, pp. 91-96). IEEE.

[23] Davidov, D., Tsur, O. and Rappoport, A., 2010, August. Enhanced sentiment learning using twitter hashtags and smileys. In *Proceedings of the 23rd international conference on computational linguistics: posters* (pp. 241-249). Association for Computational Linguistics.