7-2019

Social Media and Forecasting Stock Price Change

Joseph Coelho

Dawson D'Almeida

Scott Coyne

Nathan Gilkerson

Katelyn Mills

See next page for additional authors

Authors
Joseph Coelho, Dawson D'Almeida, Scott Coyne, Nathan Gilkerson, Katelyn Mills, and Praveen Madiraju

2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC), (July 15-19, 2019): 195-200. DOI.

Social Media and Forecasting Stock Price Change

Joseph Coelho
Department of MSCS, Marquette University, Milwaukee, WI
Dawson D'Almeida
Department of Computer Science, Carleton College, Northfield, WI
Scott Coyne
Department of MSCS, Marquette University, Milwaukee, WI
Nathan Gilkerson
Department of Strategic Communication, Marquette University, Milwaukee, WI
Katelyn Mills
Department of Strategic Communication, Marquette University, Milwaukee, WI
Praveen Madiraju
Department of MSCS, Marquette University, Milwaukee, WI

Abstract:

The Stock Market is a big influence on both national and international economies. Stock prices are driven by a number of factors: industry performance, company news and performance, investor confidence, micro and macro economic factors like employment rates, wage rates, etc. Stock pricing trends can be gauged from the factors that drive it as well as from the stock's historical performance. As fluctuations in stock prices become more volatile and unpredictable, forecasting models help reduce some of the randomness involved in investing and financial decision making. Users on social media platforms like twitter, StockTwits, and eToro discuss issues related to the stock market. Can the analysis of posts on StockTwits add value to the existing features of stock price predicting models? An existing model that uses twits as features was extended to include sentiment analysis of the text referenced by the URL in the twits to see if the model accuracy did improve. Initial results indicate that the addition of sentiment analysis of the text referenced by the URL does not improve the performance of the model when all twits for a given day are taken into account since the model only identifies the direction of change and not the degree of change. The stock prediction model achieves 65% accuracy compared to the base case accuracy of 44% and augmenting the model with sentiment analysis did not change the accuracy. The study highlights some interesting observations regarding users on the StockTwits social media platform and proposes the need for a domain specific sentiment analyzer in future work.

SECTION I. Introduction

Last year, more
than 77 *trillion in stocks were traded worldwide, in which more than half of the total value came from United States stocks. The U.S.* 2016 *GDP* was 19.4 trillion, around half of the $40 trillion from stocks. The impact that stock trading has on the world economy requires those involved to make well educated decisions when buying and selling stocks. Consequentially, models to help predict stock movements are widely sought. Unfortunately, stock prices are not simply determined by company statistical performance which makes forecasting a more difficult task.

Recently, Apple was the first company whose total market value surpassed $1 trillion. This monumental economic mark was in part, however, pushed by the public’s appetite for the stock to reach such levels. If we were to obtain the general sentiment toward this stock in terms of achievement, then we could have better predicted this outcome. Fortunately, with advancements in communication technology, this sort of data is readily available through social media platforms.

Social media has become one of the largest growing sources of public data. It is an outlet where people can share their thoughts and opinions without restriction. As such, with an estimated 54% of the U.S. population involved in stock investment, it is expected that a forecasting system could benefit from data supplied by online users of StockTwits and other social media platforms. StockTwits is a social media platform catering to the discussion of company stocks and values, for users ranging from entrepreneurs to personal investors.

Current research, including [3], [6], [7], [13], [14] shows a correlation and even predictive relationship between sentiment and stock prices. Furthermore, additional research has been done showing relationships between public worry and stock market values, further validating research of public micro blog sentiment and the stock market.

This paper explores the relationship between StockTwits and stock market closing prices through URL mining and sentiment analysis to check whether this information can enhance stock price prediction accuracy. [3] showed a strong correlation between twits text and stock market prices. The current study uses that model to factor in sentiment analysis and determine whether there is any bearing on its prediction accuracy. While it is true that the trading process is increasingly getting automated [2], social media sentiment analysis can add value to that process. Not all social media users are insightful in their posts. Identifying the "smart users" is crucial to mining relevant texts. Finally, words in a specific domain may carry a different meaning than that when used in a general context. This study seeks to answer the following research questions (RQ):

- **RQ 1:** Does financial social media content sentiment correlate with stock price change?
- **RQ 2:** Do StockTwits users with large numbers of followers predict stock price change with greater accuracy?
- **RQ 3:** Does sentiment of texts referenced by URLs in social media content help improve prediction accuracy?
- **RQ 4:** Do generic text sentiment analyzers perform well (accuracy above 80%) in the financial domain?

This paper is organized as follows: Section 2 covers current and recent research in the fields of sentiment analysis and stock forecasting models. Section 3 details the different tools and data used in this project. Models and their results are explained in Section 4. Section 5 explores the need for a domain specific sentiment analyzer. The final section has the evaluation and future work that can be done to better the predictive accuracy of the model.

SECTION II. Related Work

Due to the demand for powerful stock forecasting models, there have been many published works relating to this topic. Different forms of social media as well as other economic data have been tested under many forecasting models. StockTwits' predictive power has been debated among researchers.

This project was inspired by and is a continuation of [3], who investigated twits under a variety of models. Originally, their work was influenced by the disagreement between [9] and [10]. Three stock forecasting models were developed using social media data from StockTwits using tools such as linear regression, Naive Bayes classification and support vector machines. Using market sentiment scores extracted from the users' twits, a prediction model was created that worked with up to 80% accuracy over a given time frame. StockTwits is, however, not the only source of social micro-blogging data being used in stock forecasting models.

As mentioned above, [6] used Twitter data with similar Naive Bayes classification and support vector machines for sentiment analysis. They found a significant relationship between the stock market and tweet sentiment and found greater success in the application of Support Vector Machines. [1] also found that Twitter data is at least correlated if not predictive of DJIA values using Granger causality analysis. They also looked into different moods expressed in tweets juxtaposed against a simple positive or negative score provided by OpinionFinder, a tool for sentiment extraction.

[13] made progress using Microblog sentiment on the Chinese stock market, using similar NLP and sentiment analysis techniques despite the difficulty of tokenizing (splitting text in tokens – words, n-grams,etc.) the Chinese language. [16] used social media mining technology in combination with factors such as technical and economical indexes found in [5]'s predictive model to come up with short term stock price trend predictions.

In general, the stock-market related dataset of social media posts and microblogs that contain public opinion has proven to contain sentiment that is useful in producing stock-price forecasting models. [4] investigated the relationship between concern or worry and stock market values.

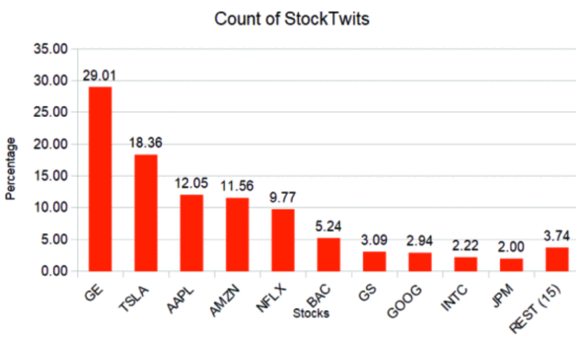

Fig. 1. Count of Twits (percent)

SECTION III. Data Collection

This section will discuss the different tools and datasets used in the project. We used StockTwits data for the period from May 2016 to April 2017. The contents discussed market sentiment regarding over 25 different companies.

StockTwits

Partner level access to StockTwits [12] provided us with a year's worth of data. All of the twits that discuss a specific company are cashtagged (marked with a dollar sign and up to four letters, i.e. $AAPL is the symbol for Apple Inc.). Information extracted for each twit included the author, text, date, etc. Unfortunately, as our model is only as good as our data, there were many smaller scale companies and stocks that received very little daily discussion to be included in our models. Only certain companies had sufficient twits to properly train and obtain results. Figure 1 indicates the proportion of twits for each company in relation to the entire twits data for the given period.

Stock Data

Quandl [11] supplied our project with economic and financial data. We gathered daily pricing data to investigate stock data.

Python

We coded in Python 3. This was adapted from the previous year's project written in Python 2.7. This language offers access to many textual analysis resources and machine learning tools. Among these tools, we predominantly used: Pandas, Scikit-learn, TextBlob, and Newsplease.

Pandas is a powerful data analyzing tool that allows for quick and efficient data storage and manipulation. Scikit-learn offers many machine learning and mathematical tools for building models. Text in twits was converted into features using the TF-IDF (Term Frequency - Inverse Document Frequency) statistic. TF-IDF is a statistic used in textual analysis in order to determine the importance of a term in a body of text. The basic idea is that the frequency of a specific term is weighted by how often that term appears in a corpus, so generally rarer terms would earn a higher weight. A model is then trained using these features to classify twits as bullish, bearish, or neutral.

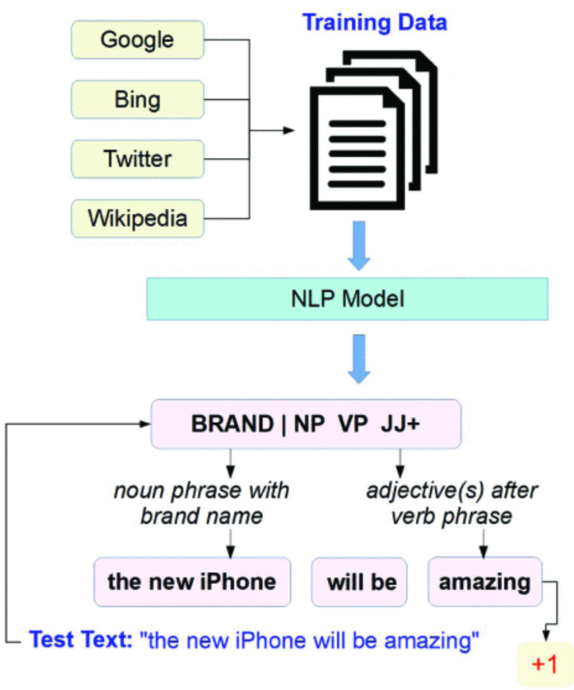

Fig. 2. TextBlob Pattern Schema

TextBlob is a sentiment analysis tool. The package uses the Brills algorithm to tag words, and then applies sentiment score based on a library with predetermined scores. We focused on the sentiment attribute, which had a polarity and a subjectivity score ranging from - 1 to +1 and 0 to +1. respectively. By default TextBlob uses PatternAnalyzer (see Figure 2) to identify parts-of-speech and construct a pattern graph which is then used to identify the sentiment. Newsplease is a web scraper that extracts the title, author and text from a given news article URL.

SECTION IV. Model and Results

The three forecasting models that [3] developed included classifiers based on Linear Regression and Neural Networks (Multi-Layer Perceptrons). These classifiers were used to identify the market sentiment of the twits and thereby the upward or downward movement of stock prices.

The base case was computed by predicting a favorable outcome every time since that is the likely outcome for the market in general. The average base case accuracy for the stocks under consideration was 44%.

A. Model 1: Twits as features for classifier

The twits are converted into TF-IDF scores and are used as features in this linear regression model. The TF-IDF produces a sparse vector representations of the twits text. The model is trained with 80% of the data and tested with the remaining 20%. This process is repeated 25 times to get the the average accuracy.

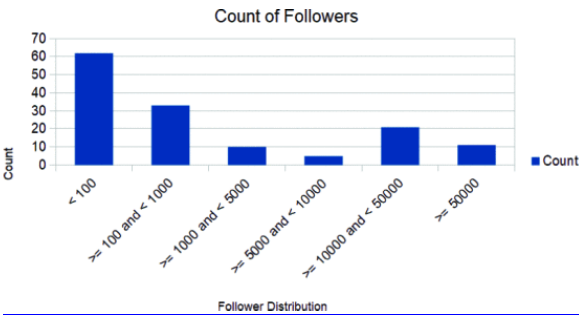

Fig. 3. Count of Followers

The model is based on the assumption that sentiment and price are correlated, so words that would receive heavily weighted sentiment scores translate to higher coefficients in the regression model. By predicting only the direction of the price and not the percent change, results from this model could suggest the existence of the correlation between price and sentiment.

The model was tested with three different approaches: same day, N-ahead, and N-aggregate. For same day and N-ahead, the twits read during a certain date would check how the stock changed that number of days ahead. N-aggregate would take all the twits over a period of N days, then compare them to the price change on day N + 1. Overall, this model's results were on average worse than the base case, but there were many stocks that did well which suggested underlying correlation between market sentiment and stock prices. **RQ 1:** Our study validated what [3], [6], [7], [13], [14] indicated: financial social media content does correspond to stock price change.

B. Model 2: Market Sentiment Analysis using Twits

This model investigates the market sentiment. Public mood has been shown to affect stock market prices and financial decision making [4].

A supervised learning approach, utilizing Naive Bayes [6], Support Vector Machines [3], and Neural Networks (Multi-Layer Perceptron) [5] was generated. Sentiment is generally marked as positive or negative. However, we found that many of the twits were not conveying any sentiment. In other words, they could be marked as "noise". The addition was a refinement to the market sentiment scores by adding a neutral score category to the positive and negative score categories in order to eliminate twits that could be considered noise. About 1800 twits were manually labeled, and then used to train the three machine learning models. The Multi-Layer Perceptron (MLP) classifier gave the highest accuracy.

C. Model 3: Market Sentiment Analysis based on Smart Users

While model 2 helped identify noisy twits, this did not help improve the accuracy very much. It was seen that twits of some users were mostly neutral in sentiment (noisy twits). Whereas some other users had twits that carried sentiment most of the time. Identifying these "smart users" was felt would add value. One of the ways to identify smart users is based on network features such as the number of likes their twits receive or the number of followers they have or a combination of such features. But we identified smart users as users who posted at least ‘x’ number of twits (the value of *x* is proportional to the popularity of the stock) in the given period and out of which the sentiment of at least 80% of them correlated to the actual stock price movement. Only twits posted by these users were collected and used for market sentiment analysis. Table I gives a count of users and smart users for the various stocks. **RQ 2:** Users of the StockTwis platform have attributes like number of ideas, following (friends), followers, liked, and watchlist. Based on the smart users that our system identified, the average values for these attributes were 19957 ideas, 156 following (friends), 12875 followers, 3839 likes and 44 items in the watchlist. As expected smart users did have a significant number of followers. However, there were many smart users who had less than 100 followers (see Figure 3). From tests, we found that number of followers by itself cannot be the criteria for being a smart user. The total number of StockTwits users for the given stocks was 10029 of which only 176 were classified as smart users. Table I shows the number and ratio of users to smart users. It is interesting to see that among users who tweet about TSLA (Tesla), the number of those whose tweets correlate to the market sentiment (i.e. smart users) is significantly low. This may be common occurrence for such "story" stocks whose value is driven by favorable press coverage rather than its assets and income. This needs to be investigated further.

TABLE I Count of Users and Smart-Users

Stock	Users	Smart Users
GOOG	556	6
NFLX	3319	36
AMZN	3061	37
AAPL	2579	36
TSLA	3627	23
BAC	1691	19
GS	1093	12
INTC	883	16
JPM	464	5

As a result of adding smart users into the dataset, the dataset shrank in size but the accuracy of the model got better. One problem, however, was that there were many unpredictable days due to the narrowed dataset. Only companies that were most frequently discussed could be predicted consistently using this model. Despite this, their average accuracy was 64%, with one out of nine predictions falling below the base case.

Filtering twits based on smart users significantly reduces the quantity of data available for analysis and might result in no data for some days. This means predictions cannot be made for those days. In addition, popular stocks have a good number of smart users while others have very few. However, the accuracy of the model for those stocks that had smart users was higher than the previous models.

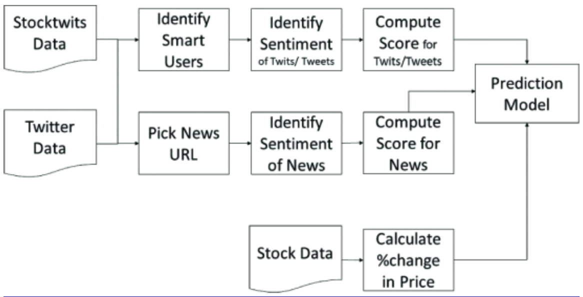

Fig. 4. Model Flow Chart

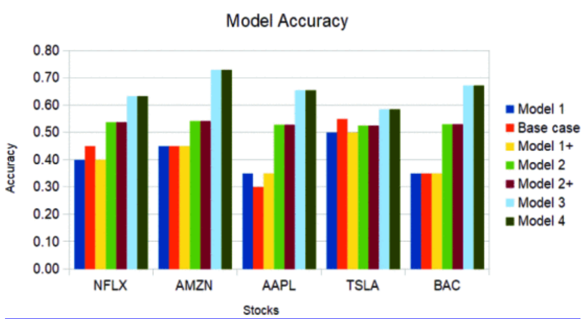

Fig. 5. Model Accuracy Chart

D. Model 4: URL Sentiment augmented Model 3

Figure 4 illustrates the work-flow for this model. In model 3, the smart users helped limit the noisy twits and thereby increased accuracy. Twits may contain URLs that reference news articles. This step seeks to find out whether factoring the sentiment of these articles might improve the accuracy of the stock price movement prediction. Apart from twits, data from Twitter can also be incorporated in this model. Twits can have at most 140 characters. URLs are a convenient way to reference text and comment on it in a twit. The news articles that the URL points to can be scraped and the sentiment of the article can be analyzed.

The Python NewsPlease package is used to extract the text of the news article and the text is passed on to the TextBlob object which in turn returns the sentiment score of the text in two variables: polarity and subjectivity. We used the following equation to come up with the sentiment score σ:

$$polarity * (1 - subjectivity) = \sigma,$$

where σ is the predictive score. This predictive score is then multiplied by that day's predicted score obtained from the classification model (model 3). Model 1 uses the entire dataset. Model 2 uses a reduced dataset containing only twits that have either a positive or negative market sentiment. Model 3 uses a further reduced dataset containing only twits by smart users. Since the dataset is reduced, the number of URL links in the twits are few. Therefore there is no significant change in the accuracy level of the model when the URL sentiment is augmented. This is seen in the Table II. Taking into account sentiment scores of all twits did not change the accuracy of the prediction.

TABLE II Model Accuracy Table

Sym	BC	M 1	M 1+	M 2	M 2+	M 3	M 4
NFLX	0.45	0.40	0.40	0.54	0.54	0.63	0.63

AMZN	0.45	0.45	0.45	0.54	0.54	0.73	0.73
AAPL	0.30	0.35	0.35	0.53	0.53	0.65	0.65
TSLA	0.55	0.50	0.50	0.53	0.53	0.59	0.59
BAC	0.35	0.35	0.35	0.53	0.53	0.67	0.67
GS	0.40	0.30	0.30	0.57	0.57	0.67	0.67
INTC	0.50	0.45	0.45	0.53	0.53	0.64	0.64
JPM	0.40	0.55	0.55	0.56	0.56	0.61	0.61

Model 1+ and Model 2+, extensions of Model 1 and Model 2 that include sentiment score of the text referenced by the URL in the twits, did not change the accuracy. One of the reasons for this outcome is there are many twits on a given day and the change for a day can either be bullish (positive) or bearish (negative). Multiple twits with sentiment corresponding to the already predicted market sentiment results in no net gain for the model in terms of price change direction.

While the model accuracy does not improve, it is interesting to note that adding sentiment scores of the text referenced by the URLs in the twits does not degrade it. Figures 5 indicates the accuracy levels of the models for some of the stocks. **RQ 3:** In general, though there is a correlation between text sentiment and stock price change, sentiment analysis of URLs referenced in the twits does not seem to enhance the accuracy of stock price prediction that is based on the twits themselves. However, the sentiment of the text corresponds to the sentiment of the twits themselves for the given period and given stock. The model used for stock price change only predicted the direction and not the degree of change.

SECTION V. Domain Specific Sentiment Analyzer

Even though sentiment analysis of the text referenced by URLs in the twits did not add to the model accuracy, there is scope to factor the degree of sentiment in future work.

TextBlob provides functionality for sentiment analysis on general text. However, words can have special meaning in specific domains. Developing a sentiment analyzer for text in the financial domain could add value.

For example, look at the text of the news article titled "Wirecard Plunges on Report Claiming Officials Knew of Book-Padding" [8] quoted below:

Wirecard AG fell as much as 20 percent after a new report claimed that two senior executives knew of alleged accounting fraud, reviving concerns about its business practices. Two officials in the Munich head office were aware of a round-tripping scheme in Singapore that may have been part of a pattern of book-padding across the companys Asian operations, the Financial Times reported Thursday. Wirecard denied the report in an emailed statement, saying that nothing about the article published today is true….

A cursory reading of the article indicates it has a negative sentiment. Sentiment analysis by TextBlob resulted in a polarity score of 0.089, indicating it has a positive or at best a neutral sentiment.

Generic machine learning based sentiment analyzers are trained on a large corpus of text. Domain specific sentiment analyzers need to be trained on text relevant to the given domain. Apart from machine learning (classification) based sentiment analyzers, there are lexicon based methods of sentiment analysis. Lexicon based methods define lists of positive and negative words. Each word is assigned a valence (score) to indicate its polarity. There are many word lists that have been curated for this purpose: AFINN, Bing, NRC, etc.

A simple lexicon based sentiment analyzer (bag-of-words model) counts the number of positive words and negative words in a given text and then by subtracting the net valence (score) of the negative words from that of the positive words the sentiment is determined.

[15] has curated a master list of words in the accounting and finance field. These words have been drawn from various journals pertaining to those fields. Words are tagged with a score indicating the following sentiment category: negative, positive, uncertainty, litigious, modal, and constraining. The site also has a curated list of stop words specific to the financial domain.

Using these word dictionaries a simple bag-of-words sentiment analyzer (SentAn) was developed. Positive and negative words in text were weighted with a score if they were preceded by a modal word. The net sum (positive-negative) of all the words in the text determines the sentiment. In a limited test of fifteen news articles we found that TextBlob had an accuracy of 47% while SentAn's accuracy was 87%. This may indicate that sentiment analysis needs to be domain specific if the actual sentiment is to be obtained.

The sentiment score given by SentAn for the text quoted above was -0.795, indicating the text has a negative sentiment (which corresponds to the actual sentiment).

RQ 4: TextBlob does a good job of sentiment analysis in general since it is trained on general text datasets. But financial news articles comprise words like share, bull, bear, etc. that have a special meaning in the financial domain. A general sentiment analyzer might mark the word 'bull' or 'bear' as neutral. However a financial domain specific analyzer will mark those words as 'positive' and 'negative' respectively. Therefore the sentiment analyzer needs to be trained on relevant texts for better results. The rudimentary bag-of-words model needs to be expanded to include lexical analysis for sentiment identification.

SECTION VI. Conclusion and Future Work

This study shows that stock related social media (and its sentiment) correlates to the stock price change. Mining additional features of social media and identifying the degree of sentiment could provide the extent of stock price change.

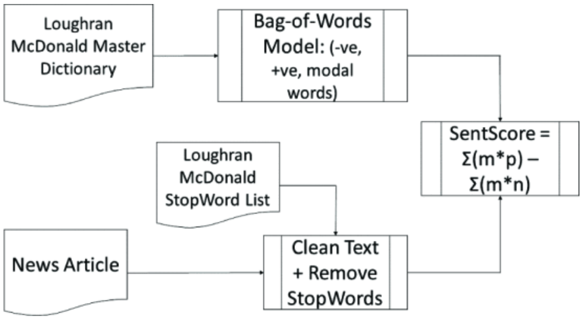

Fig. 6. SentAn: Sentiment Analyzer

The current study does not determine whether social media sentiment causes stock price change or is the result of the change. However, sentiment analysis on twits of a day (or n-prior-days) can be co-related with actual stock price change of the next day to determine the causal effect of social media on the change.

The volatility of the stock market makes it nearly impossible to predict, but studies have shown that public sentiment relates to stock prices [1], [4]. By harnessing various opinions and viewpoints towards individual stocks, we can enhance a forecasting model.

The main goal of this paper was to check whether factoring in text referenced by a URL in a twit would improve the accuracy of a stock price forecasting model. Model 3 provides the best accuracy scores on average and performing sentiment analysis on embedded URLs (Model 4) does not add value to it. However, social media may provide subtle insight into the direction of price change especially when conversation about stock is the cause for driving change.

Identifying smart users by keeping a minimum value for attributes like followers, ideas and likes might filter out users who are not really "smart".

The domain of the URL can be checked to identify the authenticity of the news article included in the twit. If the domain is of repute then the sentiment of the article could be weighted higher.

In addition to the incorporation of data from multiple platforms, recency bias could be taken into account whereby the URL sentiment scores could be balanced differently to provide different outcomes, both with the number of days ahead to predict and the relationship to the twits' sentiment score for that day.

Finally, a domain specific sentiment analysis tool would certainly help identify the correct sentiment of a financial or economic text. The bag-of-words model needs to take into consideration lexical analysis (parts of speech) to determine the sentiment degree and impact of modal words on the adjectives or adverbs. Domain specific text whose sentiment is identified may then be collected into a corpus for training a machine learning (classification) model which could then be used for sentiment analysis.

ACKNOWLEDGEMENTS

This work was sponsored in part by a National Science Foundation grant #ACI-14 61264 supporting research experience for undergraduates (REU).

REFERENCES

1. Johan Bollen, Huina Mao and Xiaojun Zeng, "Twitter mood predicts the stock market", *Journal of Computational Science*, vol. 2, no. 1, pp. 1-8, 2011.

2. Evelyn Cheng, "Just 10% of trading is regular stock picking", *JPMorgan estimates*, 2017.

3. Scott Coyne, Praveen Madiraju and Joseph Coelho, "Forecasting Stock Prices using Social Media Analysis", 2017.

4. Eric Gilbert and Karrie Karahalios, "Widespread Worry and the Stock Market", *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media*, pp. 58-65, 2010.

5. T. Kimoto, K. Asakawa, M. Yoda and M. Takeoka, "Stock market prediction system with modular neural networks", *1990 IJCNN International Joint Conference on Neural Networks*, vol. 1, pp. 1-6, 1990.

6.Man Li, Chi Yang, Jin Zhang, Deepak Puthal, Yun Luo and Jianxin Li, "Stock market analysis using social networks", *Proceedings of the Australasian Computer Science Week Multiconference on - ACSW '18*, pp. 1-10, 2018.
7.Yuexin Mao, Bing Wang, Wei Wei and Benyuan Liu, "Correlating S&P 500 Stocks with Twitter Data", *Proceedings of the First ACM International Workshop on Hot Topics on Interdisciplinary Social Networks Research*, pp. 69-72, 2012.
8.Stefan Nicola, "Wirecard Plunges on Report Claiming Officials Knew of Book-Padding", 2019.
9.Chong Oh and O Sheng, "Investigating Predictive Power of Stock Micro Blog Sentiment in Forecasting Future Stock Price Directional Movement", *Icis*, pp. 1-19, 2011.
10. Nuno Oliveira, Paulo Cortez and Nelson Areal, "On the predictability of stock market behavior using StockTwits sentiment and posting volume", *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8154 LNAI, pp. 355-365, 2013.
11."Financial Economic and Alternative Data".
12."StockTwits® - Share Ideas & Learn from Passionate Investors & Traders", 2017.
13. Tong Sun, Jia Wang, Pengfei Zhang, Yu Cao, Benyuan Liu and Degang Wang, "Predicting Stock Price Returns Using Microblog Sentiment for Chinese Stock Market", *Proceedings - 2017 3rd International Conference on Big Data Computing and Communications BigCom 2017*, pp. 87-96, 2017.
14. Wenting Tu, David W. Cheung, Nikos Mamoulis, Min Yang and Ziyu Lu, "Investment Recommendation using Investor Opinions in Social Media", *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval - SIGIR '16*, pp. 881-884, 2016.
15."Resources // Textual Analysis // Software Repository for Accounting and Finance // University of Notre Dame", 2016.
16. Yaojun Wang and Yaoqing Wang, "Using social media mining technology to assist in price prediction of stock market", *2016 IEEE International Conference on Big Data Analysis (ICBDA)*, pp. 1-4, 2016.