Journal of Engineering and Applied Sciences 13 (Special Issue 1): 2388-2402, 2018
ISSN: 1816-949X

Stock Market Prediction Using Sentiment Analysis Based on Social Network: Analytical Study

[1]Salam Al-Augby, [2]Noor Al-musawi and [1]Alaa Abdul Hussein Mezher
[1]Department of Computer Science,
[2]Department of Mathematics, Faculty of Computer Sciences and Mathematics,
University of Kufa, Kufa, Iraq

Abstract: Social media such as “Twitter and Facebook” is carefully thought as one of the effective factors in many fields of study today such as stock market prediction. This study reviews new methods of stock market prediction that used the effect of social media networks. In the last decade, the social media has an effective impact on the stock market prediction. This research is managed and done conducted by reviewing 25 articles that studied the social media effect on stock market prediction and discussed their effect on different stock markets. One of the main end results is that there is a significant effect of the stock market prediction as shown in this study.

Key words: Stock market prediction, social network, sentiment analysis, Twitter, Facebook, effect

INTRODUCTION

One of the prominent studies today is the effect of social media in stock market movement. The stock market has many definitions such as it is the mechanism in which one can lend and borrow money (Weston *et al*., 1996). Anthony Sunders defines stock markets as the structures through which funding flows (Saunders and Cornett, 2003). The stock market is an intermediary financial institution that is run by an elected body that mobilizes the savings and financing projects through the provision of legal and technical frameworks that regulates the process of trading medium and long-term securities issued by companies and government in order to obtain funding (Abbood, 2016). The stock market is a complex and difficult system and its complexity come from the influence of many different and uncertain factors and it can be affected and interacts with different other political, economic and social factors. The ambiguous about stock market nature if it is “linear” or “non-linear” system make it difficult (Luo *et al*., 2010). In spite of its complexity stock market still one of the important economic sectors (Al-Augby, 2015).

The stock market prediction is considered as one of the most challenging task because of the stock volatility and dynamic (Duong *et al*., 2016). The social media platforms like Twitter and Facebook offer the ability to express, post, share and publish the people’s opinions on it that make it popular to use. Recently, many researchers have studied the effect of social media in stock market movement and prediction. According to Alexa ranking Facebook is of the top ten most-visited websites in the world while Twitter in the rank 12th of 2017 ranking (Alexa Internet Inc., “Alexa Top 500 Global Sites” (Anonymous, 2017). Social media has a significant role in our society as a new phenomenon through which one can communicate and share his ideas through the internet in the virtual community. Although, the Facebook and Twitter have a high number of active users but there are some other social media websites have a high active number of users in different countries such as Guba.com.cn which considered the most used social media website in Chine (Chen and Du, 2013). Prediction in one of its definitions refers to a process of forecasting is not with standing, most often referred to the forecast of future accepted numerical values or to increase/decrease trends in time-related data by using large amount historical data (Han *et al*., 2011).

SENTIMENT ANALYSIS AND STOCK MARKET PREDICTION

Sentiment or opinion analysis can be defined as a process of opinion determination from people’s emotion and feelings (Cakra and Trisedya, 2015). Sentiment analysis has strong, relevant with Natural Language Processing (NLP), hence, sentiment analysis may be able to be known as the process of opinions and subjective

Corresponding Author: Salam Al-Augby, Department of Computer Science, Faculty of Computer Sciences and Mathematics, University of Kufa, Ramadi, Iraq

expression extraction from a text by using NLP techniques. That is performed by the polarity classification of text into positive, negative or neutral sentiments (Aqel and Vadera, 2013). The people's opinions are significant in the decision-making process (Daniel *et al*., 2017). Recently, sentiment analysis widely used in the microblogging websites analysis. Thanks to technology that make the interaction between people easier. The social media contents (posts, Tweets, photos, etc.) are mostly analyzed by using sentiment analysis techniques by different peoples such as political, marketers and companie's analysts (Taboada, 2016). The ability of sentiment analysis to classify the document, the sentence or even part of the sentence (phrase) according to its polarity (positive, negative and sometimes neutral) (Pang and Lee, 2004) make it significant.

Nowadays, the high stock market value considered as a measure of the highest economies (Kumar and Ravi, 2016), therefore, the investing in the stock market has an important role in the economy. The people try to invest their money in the stock market to earn money in easy way but there is still a high chance to lose their money because of the volatility of stock market nature. The stock market prediction is significance for both common and specialist people (Nayak *et al*., 2016). Even, though the complexity of studying stock market prediction it has been studied by many researchers. There is no one methodology can give accurate prediction results, thus, there were a lot of trials to get good results. The social media websites are considered important platforms to provide opinions or feelings, sharing among investors (Li *et al*., 2016). Using the sentiment analysis as an assistant factor in stock market prediction is a prominent field of studies and below is a presentation of the analytical study of 25 recent studies of using sentiment analysis that's based on social media in stock market prediction.

One of the first studies that illustrated Twitter mood effect on predicting stock market is the article written by Bollen *et al*. (2011) in which they used two tools to estimate the changes of the people's mood that got from Tweets on Twitter from February 28, 2008-December 19, 2008 for 9,853,498 Tweets posted by approximately 2.7 M users Fig. 1 shows the outline of the used methodology. The first tool was opinion finder (a scale measurement of positivity and negativity of mood) that has no significant effect on prediction accuracy (according to their conclusions) in comparison to using only the historical DJIA data. The second tool was Google-Profile Of Mood States (GPOMS) in which mood measurement depends on six dimensions. The second tool has highest prediction accuracy. Behavioral finance prominent role was the cornerstone of their study.

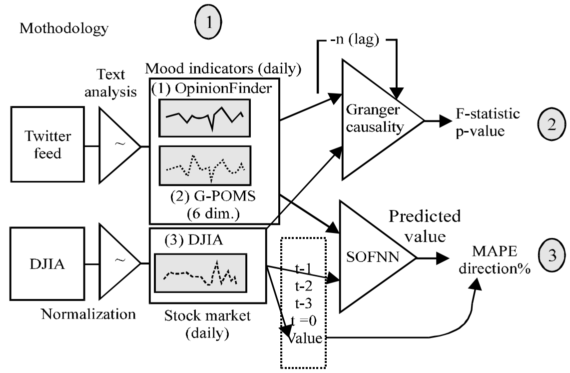

Fig. 1: Diagram of the used methodology that divided in three parts

Zhang *et al*. (2011) considered one of the early attempted to predict stock market indicators like Dow Jones, S and P 500 and NASDAQ by taking the effect of Twitter Tweets analysis. This study is based on taking the correlation between the hope and fear indicators (the number of Tweets ranging from 8100-43040 Tweets per day lasted 6 months starting from March 30, 2009-Sept 7, 2009) with the stock market indicators for the same period of time. There was significant positive correlation with VIX while there was significant negative correlation with Dow Jones, NASDAQ, S and P 500 where the indicators of dow went down when there were the person's emotions express a lot of hope, fear and worry and the indicators of Dow went up when there were less hope, fear and worry (Table 1).

Different classifying methods were used in different study Vu *et al*. (2012) such as decision tree classifier with hybrid methods in its training phase to classify and to predict Apple (AAPL), Google (GOOG), Microsoft (MSFT) and Amazon (AMZN) stocks, respectively for changes in their daily up and down in a 41 market day sample the prediction model is illustrated in Fig. 2. These companies are taken because they have an adequate number of related Tweets (5,001,460 daily Tweets were crawled by using Twitter online streaming API from 1st April 2011-31st May 2011). The accuracies of 82.93, 80.49, 75.61 and 75.00% in predicting the daily up and down changes was tested by cross validation method.

Chen and Du (2013) in their study, they concluded that the social network (Guba.com.cn which considered one of the most popular social media in chine if not the only) can give moderate prediction result 56.28% in 3 months and made an increment 1.17% in stock price. They studied the effect of social media and calculated key characteristic of social behavior graph that based on human's online behavior as the correlated factor

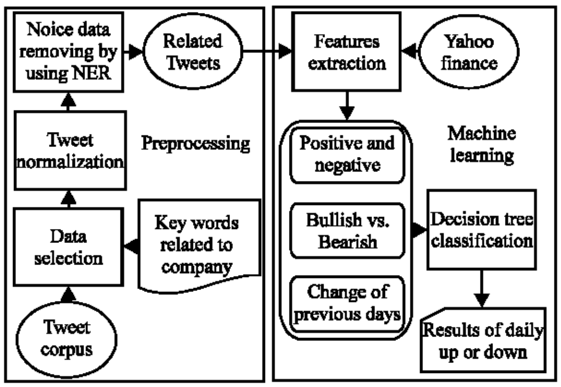

Fig. 2: Prediction model

with trading volume/price Shanghai/Shenzhen stock exchange. They used in their study BP-neural network method which gives some rough results. They thing that the investment decision making process will depend on the social media information. They don't depend on emotion analysis because they believe that the stock gambler will not express their emotions.

Sentiment analysis was used as an improvement factor in the stock market prediction accuracy in Porshnev *et al*. (2013). In this research, the researchers analyzed the psychological states of Twitter users and used this information to increase the prediction accuracy but the best average accuracy was 64.10% that means these information doesn't have a significant effect on increasing accuracy based on these results. Their data were taken from Yahoo finance website (http://finance.yahoo.com) for a 61 days period starting from 13th, February to 29th, April, 2013 of DJIA, S and P500 indicators and the used Tweets were 755'000 101 Tweets. The used prediction algorithms were support vectors machine and neural networks algorithms.

One of the contributing studies that discussed the role of social media (Twitter as an example in this study) in stock market prediction is Qiu *et al.* (2013) where they concluded that the information acquired from social media can improve the accuracy of the network-embedded prediction where the threshold strategy in the Bayes-Nash equilibrium was adopted by the participants. They suggested market mechanisms that can collect the information that has an effect on stock market prediction and hence, it could increase the prediction accuracy.

As an answer to the problem statement of Yigitcan Karabulut research they deduced that the investor sentiment as in Facebook's Gross National Happiness (GNH) can be used as an indication to predict the changes in both daily and weekly returns and trading volume for the US stock market in their study. The data used is from Thomson Reuters Datastream. From January 1, 2008-April 27, 2012 and includes the 1,090 trading days after taking into consideration that the weekends and holidays were not taken.

Xu and Keelj (2014) proposed a method to predict the change of stock market of the next day depending on the collective sentiment analysis for the period from March 13th, 2012-May 31st, 2012. Their method based on two stages, firstly the natural language processing and secondly the machine learning algorithms (Naive Bayes, Decision Tree (J48 in Weka) and support vector machine (SMOin Weka). In natural language processing algorithms (Hand-labeled classification with the assistance of the Weka toolkit for sentiment detection) they tried to classify the Tweets into positive and negative Tweets after checking their polarization. The best classification results at the two stages were 71.84 and 74.3% where these results are considered satisfactory according to their conclusions. Their prediction power on stock price (9 out of 15 stocks) for the next day was shown after hours based on the collective sentiments.

Utilize social media for stock market prediction with Factorization Machine (FM) can probably enhance the prediction accuracy (Chen *et al*., 2014). Chen *et al.* (2014) study there are three main steps had done the first step was the analyzing the relationship of FM to Generalized Linear Model (GLM) and Support Vector Model (SVM) to illustrate the reason of using FM. The second step was to use social media in process of prediction and at last step shows the benefits of using this method in predicting. The accuracy of using FM was 81% that means this method has a credit over other highest development models. The data used in this method were Sina Weibo (is the most popular Chinese micro-blogging service) for 361 trading days starting from 29, September, 2011-29, March, 2013.

The goal of Nguyen and Shirai (2015) research is to implement a model to predict stock price movement using sentiments on social media. Topic Sentiment Latent Dirichlet Allocation (TSLDA) Model was proposed to specify the topic and the sentiment of it at the same time. This model gave better accuracy than Latent Dirichlet Allocation (LDA) and Joint Sentiment/Topic (JST) based methods by 6.43 and 6.07%. This research tried to extract opinions from a text which is the gist of sentiment analysis. Only the historical data (extracted from Yahoo finance for 5 stocks) was used with accuracy in about 6.07%. Finally, this research showed the importance of social media in stock market prediction.

Prediction of the stock market was studied by Liu *et al.* (2015) they concluded that firm-specific microblogging metrics were used to check the possibility to identify homogeneous stock groups with their proposed prediction method. The used clustering algorithm method was K-means to distribute the firms into different groups depending on their similarity attributes. The case study was NYSE and NASDAQ stock exchanges where the program lasted 5 days to crawl the data starting from June 30, 2013. Their results showed that it is possible to use some metrics (social media data) to predict co-movement of stocks as well as increase the co-movement prediction accuracy.

Hamed *et al*. (2015) had studied the relationship between Twitter messages and the Saudi Stock Market Index and found that there is a significant relationship between them. The Tweets (3335 Tweets company websites that are chosen from Mubasher for over a 53 days period between 17-3-2015 and 10-5-2015) were formed in three categories: positive, negative and neutral using Naive-Bayes, SVM and KNN classifying methods and the accuracies were 69.86+/-2.71% (micro: 69.86%), Accuracy: 96.60+/-2.16% (micro: 96.60%) and 96.45+/-1.66% (micro: 96.45%), respectively.

Prediction of the Indonesian stock market using simple sentiment analysis was discussed by Chakra and Trisedya (2015). In this research sentiment analysis was used to predict companie's stock prices of 13 companies of the Indonesian stock market. The sentiment analysis was a result of gathering Tweets using Twitter REST API. After Tweets collection there were a word weights counting, Part of Speech (POS) tagging and finally sentiment shifters. In order to classify the Tweets and the calculate the sentiment Naive Bayes and Random Forest algorithm were used where these algorithms give 56.50 and 60.39% accuracy, respectively. To supervise the classification the Support Vector Machine (SVM), Naive Bayes, decision tree, random forest and neural network (with single layer perceptron) were used. The prediction model was built by using linear regression methods. The best prediction of the combination of sentiment and stock prices can be with 0.9989 and 0.9983 coefficient of determination.

Based on behavioral finance researchers the participant's emotion can be considered as driving force to the stock market. In an attempted to reproduce the findings by measuring the mood states on Twitter Nofer and Hinz (2015) analyzed approximately 100 million German Tweets between January, 2011 and November, 2013, the process of data collection was divided into two stages. Their research consisted of two analysis stages. In a first analysis which lasted from January 1, 2011 to March 17, 2012, they discovered that there was no significant relationship between collecting Twitter mood states and the stock market. On the other hand, taking into consideration the Twitter followers in addition to the main Tweets can give better results and that one of their main conclusions where this process started from December 1, 2012 to May 31, 2013 and this was the second analysis stage. As a result of their study there was an up to 36% portfolio increment within a 6 month, period after the consideration of transaction costs. Social Mood Index (SMI) and follower-Weighted Social Mood (WSMI) used in this research.

There is a significant correlation between Twitter Tweets and technology company stock changes as proved by Li *et al*.(2016) in their study "Can Twitter Posts Predict Stock Behavior?" where there is a change in the Twitter's mood when sharing Tweets in accordance to the company's stock price trends. The Tweets chosen between November 9-November 20, 2015. They depend on the words in Tweets such as happy, sad, anger, fear, disgust surprise, etc. to define the people's mood from their Tweets and they retrieved the NASDAQ market closing price during the same period of collecting the Tweets in their research.

Future stock price prediction using the analyzed data that collected from social network was discussed by Skuza and Romanowski (2015) as well. Their sentiment analysis results taken from Titter microblogging for 3 months starting from 2nd January, 2013-31st March 2013. This study took the Apple stock (historical data for the same period) as their case study by studying the Apple hashtags (#AAPL) of Apple stock symbol in Twitter for the same data. Figure 3 illustrated the design of the system. The classifying method of Twitter information was a machine learning algorithm. The map reduce programming model was used to calculate the prediction results. In evaluation stage the prediction results were tested for different time of prediction 5, 15 and 30 min and 1 h before forecasting moment. The number of Twitter messages was little that led to making 5 min forecasting difficult to perform. This study shows there is a correlation between sentiment analysis results and the stock market that is to say there is good chance to use the information comes from social media to help in stock market prediction.

As a continue with the series of study that studied the prediction of stock market using the effect of social network (Twitter platform). That comes from the importance of the sentiment analysis that studies the effect of participant's mood on their decision. Calingo *et al*. (2016) concluded that Filipino public Tweets could effect on the closing index of the Philippine Stock Exchange movement (Quandl.com was the source of the stock market closing index collected data). The Tweets of Filipino public Tweets were classified into positive and

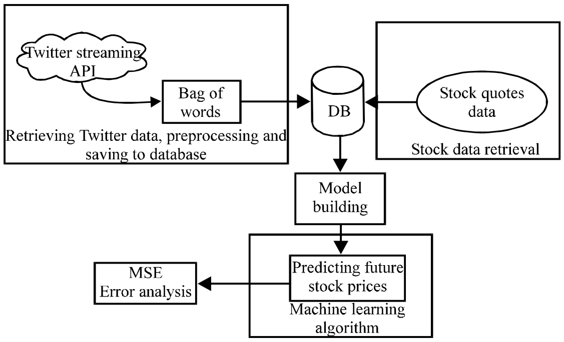

Fig. 3: The design of the system

negative polarity (the neutral excluded because the researcher thought that it is difficult to specify neutrality). These Tweets (800,000-850,000 Tweets or 9,300 Tweets per day) collected for 91 days, approximately 3 months, from June 2 to August 31, 2015. This classifying based on the extracted features and the pre-processing phase that led to sentiment analysis. Garner causality analysis was performed and finally the prediction. P-values and regression algorithms were the used prediction models. This study concluded that the geo-location and local news which are the basis of Tweets collection might cause the future change in Philippine Stock Exchange closing index. Figure 4 depicts the operational framework of the studied system.

The role of sentiment analysis that results from analyzing Tweets is one of the thrilling subjects in financial aspects this subject was discussed by Kordonis *et al*. (2016). In this study, they developed a system that collected the Tweets (extracted using Twitter's Search API for approximately, 1.5 million messages). Natural language processing algorithms used in sentiment analysis and in order to get the opinion from the Tweets Naive Bayes Bernoulli and support vector machine techniques were applied to classify the Tweets into positive and negative. The same algorithm used to check for how long this analysis correlated with stock market movement (Yahoo! finance API was the source of open, close, high and low values for each day). Finally, next day's actual close price prediction was examined where the prediction accuracy 75%. Figure 5 the flow diagram of the system is shown below. Starting from the importance of stock market prediction there are many studies that tried to enhance the prediction accuracy. In this context the study of prominent role in sentiment analysis of the people's Tweets and how it correlated with stock market movements is increased recently. Kordonis *et al*. (2016) in their study they used a developed system to predict and classified the Tweets sentiment by using Naive Bayes and SVM into positive

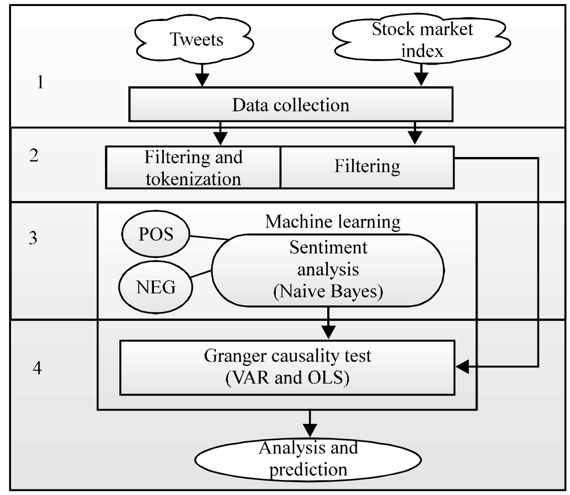

Fig. 4: The framework of the system

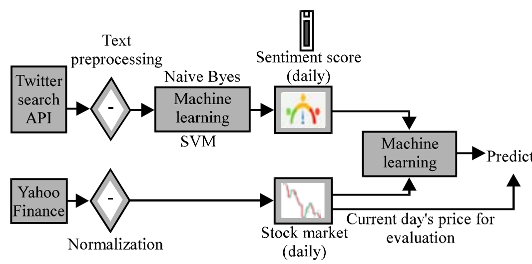

Fig. 5: The flow diagram of the system

and negative sentiment and in the next step the used the same classifying algorithm to predict the next day close price (these stocks taken from Yahoo! finance API for 16 companies) after checking the correlation between sentiment and stock market behavior. Lastly, they check the accuracy of their prediction system. One of their main findings is that there is a significant role of people sentiment to predict stock's movement and there is a possibility to create a stock lexicon of common used words.

Nayak *et al*. (2016) research is an attempt is made for prediction of the stock market trend. Two models are built one for daily prediction and the other one is for monthly prediction. The daily prediction model is illustrated in Fig. 6. Supervised machine learning algorithms are used to build the models. As part of the daily prediction model, historical prices are combined with sentiments. Up to 70% of accuracy is observed using supervised machine learning algorithms on a daily prediction model. Monthly prediction model tries to evaluate whether there is any similarity between any 2 months trends. Evaluation proves that trend of one month is least correlated with the trend of another month.

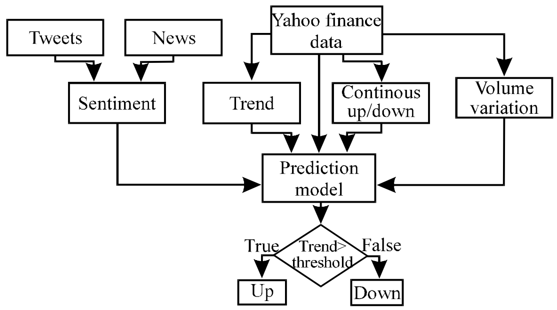

Fig. 6: Daily prediction model

The difficulty of stock market prediction comes from the stock prices are not static. This subject was discussed in Wang and Wang (2016). In this research, they considered social media as one of the assistance factor that could be used in stock market prediction. The historical stock market data were collected from Yahoo and Google financial API. Sina Weibo, Tong Hua Shun Network and Dong Fang Cai Fu network are the social media websites sources of this study. Hadoop Distributed File System (HDFS) was used to store social media data. The stock comments were fetched after closing the trading day. The training stock market data was performed in the period starting from January 1st, 2009 on October 1st, 2015 while the experimental period was from November 13th to December 18th, 2015.

A combination of news and social media was used in the process of performance Karachi Stock Exchange (KSE) prediction process by Usmani *et al*. (2016). Many other rates used besides news and social media such as oil rates (which found as the most relevant to market performance), gold and silver rates, interest rate and Foreign Exchange (FEX) rate. Simple Moving Average (SMA) and Autoregressive Integrated Moving Average (ARIMA) statistical methods were used as another type of input in the prediction model. Many machine learning algorithms was compared such as Single Layer Perceptron (SLP), Multilayer Perceptron (MLP), Radial Basis Function (RBF) and Support Vector Machine (SVM). As a result, KSE-100 index can be predicted using ML algorithms with 77% prediction accuracy. The period of study lasted over 4 months started from September, 2015 to January, 2016 with 100 records with 30 (news and Twitter feed) instances to test while the remaining instances for training. Figure 7 shows the prediction model.

Zhang *et al*. (2016) in their study proposed a framework in which a Degree of Social Attention (DSA) measurement. The DSA is based on collecting the social media researcher's opinion besides their personal influence in a social network. In other words, they tried to suggest a method to calculate the effect of social media

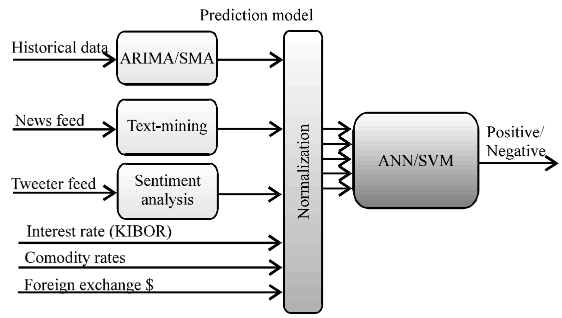

Fig. 7: Prediction model

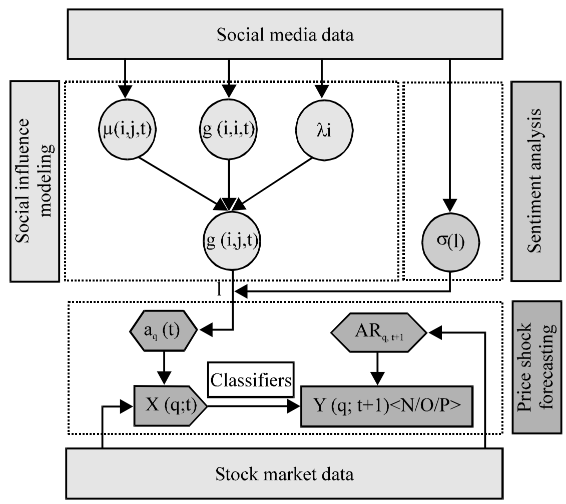

Fig. 8: DSA-based forecasting model

on the stock market and use these calculations in price shocks forecasting of Chinese stock market Shenzhen Stock Exchange (SZSE) and Shanghai Stock Exchange (SSE) as illustrated in Fig. 8. The social media data were taken from Weibo.com, the largest social media network in China. Their research was divided into three main stages calculating DSA which used to enhance the performance of forecasting stock price shocks then selecting the best results of the five tested classifiers where the SVM showed the worst performance. Finally, was the comparison that made to check which of the positive or negative social attention gives better forecasting results.

Ming *et al*. (2014) Model, Sparse Matrix Factorization (SMF) Model, used by Sun *et al*. (2016) to map both text and financial data in a joint latent factor space of dimensionality. They used data were from StockTwits® (Founded in 2008), Yahoo! finance, CNN Money, Reuters, TheStreet.com, Bing.com and The Globe and Mail. They created a dictionary that contains the economic terms. These terms aligned with 420 stocks after

Table 1: Explanation of the prediction algoriyhms

Research	Prediction Model	Data	Social network type	Period (long term, short term, medium term)	Time frame	Sentiment analysis classifier	Main conclusions	Strength	Weakness
Bollen *et al*. (2011)	Self-Organizing Fuzzy Neural Network (SOFNN) Model	9,853,498 Tweets posted by approximately 2.7M users and DJIA closing-values from Yahoo! finance from	Twitter	February 28, 2008 to December 19, 2008	Short term	GPOMS and opinon finder	There is a correlation between 4 of 7 factors of mood dimensions of Granger scale between the collected Tweets with DJIA index	This system is not limited to specified geographical area because it is designed to analyze English Tweets	It is very difficult to validate the mood results ofthecollected Tweets that were analyzed by the two models are truth or not
Zhang *et al*. (2011)	Not mentioned	the number of Tweets ranging from 8100-43040 Tweets per day and Dow Jones, S&P 500, and NASDAQ indices	Twitter	6 months starting from March 30 ,2009, to Sept 7, 2009	Short term	Chicago board options exchange Volatility Index (VIX) was used for fear mood	The main conclusion is that there is a correlation between the person's mood when they feel very happy, fear and worry the Dow Jones goes down and if they feel less happy, fear and worry the Dow Jones index goes up	-	The study need to present more results and to verify these results
Chang *et al*.	Machine learning (decision Tree (C4.5)) classifier for training	5,001,460 daily Tweets was crawled by using Twitter online streaming API and Apple (AAPL), Google (GOOG), Microsoft (MSFT) and Amazon (AMZN) stock, respectively, days) in a 41 market day sample	Twitter	1st April 2011 to 31st May 2011	Short term	Online sentiment classifier named Twitter Sentiment Tool (TST)	In this study there was a trail to predict stock market movement based on sentiment analysis of Tweets (Positive, negative and consumer confidence). The accuracies of their predicting were 82.93, 80.49,75.61 and 75.00%	Using NER to reduce the noise that enhances the prediction accuracy	The Tweets classification is just for positive and negative withouttaking neutral class and beside that the period of study was short (41days)
Chen and Du (2013)	Back propagation-Neural network	PetroChina volume/price	Chinese stock forum Guba.com.cn.	Jan 4, 2013 to Jan 25, 2013	Short term	Social behavior graph and the hamming distance for calculating the similarity of arbitrarily two topics and post. The spider was used for collecting social post	The main conclusion is that there is a significant interaction between social media and stock market movement	This work proves the interaction between social media and stock market movement other than Facebook and Twitter	It is directed for Chinese people only because this system used a social media by Chinese almost
Porshnev *et al*. (2013)	Support Vectors Machine and Neural Networks algorithms	DJIA and S and P500 from Yahoo Finance and more than 755 million Tweets (3483642 Tweets per day)	Twitter	February 2nd to April 29th, 2013	Short term	They used a dictionary, Naïve Bayes algorithm and Brief Mood Introspection Scale	Their main conclusion stated that the sentiment analysis of Tweets can be used as an additional factor that can increase the prediction accuracy	Using more than one method of classification of Twitter messages (Dictionary and Naive Bayes algorithm)	The training period is short and the sentiment analysis algorithms need to be improved

Table 1: Continue

Research	Prediction Model	Data	Social network type	Period (long term, short term, medium term)	Time frame	Sentiment analysis classifier	Main conclusions	Strength	Weakness
Qiu *et al*. (2013)	The network-embedded prediction market	Not mentioned	Twitter	Not mentioned	Short term for movie box office revenue	Collecting dispersed economic information	The aggregation of dispersed information can improve prediction results according to wisdom of crowds	The prediction accuracy error is small and the prediction result will be better	The system depends on the amount information provided by people in addition to the nature of information more private the better results
Karabulut	Regression return for Seasonal Affective Disorder (SAD) Variable Vector Autoregressive (VAR) was used to check the relationship between GNH and stock market	1,090 trading days for Chicago Board Options Exchange (CBOE) Market Volatility Index (VIX) and daily open-to-close returns of Dow Jones US, composite index of NYSE and S & P 500	Facebook	1st January, 2008 to 27th April, 2012 from Facebook	Short and Long term	Facebook's Gross National Happiness (GNH)	There is some evidence that there is a relationship between GNH (as a measurement of sentiment) and future trading volume. The GNH can be used for both Short term and reversely for the long term	It uses the online information from social network to help in predicting the stock market	There are two groups of mood by GNH only positive or negative there is no mention of neutral and using basic prediction model (auto-regressive model)
Xu and Keelj (2014)	Machine learning algorithms (Naive Bayes, Decision Tree (J48 in Weka) and support vector machine (SMOin Weka)) with granger causality tests for validation	Collect about 100,000 Tweets with hand-labeled 2380 Tweets for daily 16 stocks	Twitter	From March 13th, 2012 to May 31st, 2012	Short term	Hand-labeled analysis and Weka toolkit for sentiment detection of Tweets	There is a significant role of social interaction in investment decision making	The researcher took in their consideration the neutral class of their sentiment classification of Tweets	The data used in this work were small and the period was short
Chen *et al*. (2014)	Factorization Machine (FM) and analyzing FM to Support Vector Model (SVM) and Generalized Linear Model (GLM)	361 trading days of the Shanghai composite index from 29, September 2011 to 29, March, 2013 and 256,691 messages collected in 361 days	Sina Weibo (is the most popular Chinese micro-blogging service)	From September 1st, 2010 to Match 29th, 2013 for data from Sina Weibo and 361 trading days for the period mentioned in the data	Short term	Stanford word segmenter to split Chinese text and Supervised Latent Dirichlet Allocation (sLDA) for labeling messages and Logistic Regression (LR) for classification	The factorization machine can be used for improving stock market prediction beside the significant impact of textual text in the prediction of stocks	This social media platform presents not only financial news but useful discussion as a well that gives clearer view	It is dedicated for the Chinese speakers in addition to the limited number of search keywords that used in this model
Nguyen and Shirai (2015)	Support Vector Machine (SVM)	Open, close, high, low, and adjusted close prices for 5 stocks from Yahoo finance and 5 message board from Yahoo finance as well	Message Board from Yahoo finance	From July 23rd, 2012 to July 19th, 2013	Short term	Topic Sentiment Latent Dirichlet Allocation (TSLDA)	This research tried to prove that the role of sentiments result from topics in stock movement prediction. The TSLDA model result was better than Joint Sentiment /Topic Model (JST)	The researchers tried to classify the sentimental mood into true positive and true negative besides false positive and false negative	This model discussed specialized kind of social media and not used Twitter or Facebook The prediction accuracy was 56% and it is not so high

Table 1: Continue

Research	Prediction Model	Data	Social network type	Period (long term, short term, medium term)	Time frame	Sentiment analysis classifier	Main conclusions	Strength	Weakness
Liu *et al*. (2015)	Correlation test	NYSE and NASDAQ stock exchanges (daily returns) for 6 months. 293 American firms account	Twitter (Official firms account) running from June 30, 2013 for 5 days.	From July 1, 2013 to December 31, 2013	Not mentioned	K-means clustering algorithm	The main conclusion was that there is a significant influence of clustering firm-specific microblogging metrics (social media has better results than industry ones) in choosing the proper prediction model by different types of users	This study studied the comovement which has an important role in selection and risk premium in addition to the cost of capital	This study did not take into their account the effect of other social media effect such asFacebook, blogs,etc and they studied the posts only without studying the followers. Theutilizing of text and sentiment analys is could be useful for this type of studies
Hamed *et al*. (2015)	Correlation test	3335 Tweets from Mubasher (company website in Saudi Arabia). The financial data from Saudi Stock Market Index of closing values	Twitter	March 17th to May 10th,2015 for 53 days	Short term	Rapidminer software was used in the preprocessing stage. SVM, KNN and Naive Bayes algorithms were used as classification algorithms	There is a significant effect of Twitter messages on Saudi Stock Market Index movement	This study studied the Arabic Tweets and that consider a good attempt	Thestrength could be considered asaweakness pointbecause this model is dedicated for Arabic speakers only
Cakra and Trisedya (2015)	Linear regression	Several companies (13 companies) of Indonesian Stock Market collected from Yahoo finance API	Twitter	April 14th 2015- April 30th 2015 for two weeks	Short term	Naive Bayes and random forest algorithms	One of the main conclusions random forest algorithm have better classification results than a Naïve Bayes algorithm	This model studied more than one case in their prediction model such as price fluctuation prediction, margin percentage prediction and stock price prediction	The sentiment analysis modelneeds more improvement tohavemore significant features. The period of this prediction modelneeds to be expanded and it may result better prediction modelincase of using another method
Nofer and Hinz (2015)	Regression	100 million German Tweets between January 1st, 2011 and November 30th, 2013, the financial data taken from the DAX intraday return	Twitter	January 1st, 2011 and March 17th, 2012 for 310 trading days		German version of the Profile of Mood States (POMS) that called ''Aktuelle Stimmung-sskala'' (ASTS) in mood analysis	This study proved that follower-weighted social Mood can give good prediction results of share returns	In their social media analysis they took into account the followers of the twitter accounts	The fake messagesdid notrecognize besidessome Tweetsfeatures didn'tcontain inthemodel dictionary asemoticons and slang language

Table 1: Continue

Research	Prediction Model	Data	Social network type	Period (long term, short term, medium term)	Time frame	Sentiment analysis classifier	Main conclusions	Strength	Weakness
Li *et al*. (2016)	Correlation test	Tweets were collected for 2 weeks The financial data (closing price) collected for the same period from the NASDAQ market for five companies	Twitter	From November 9th to November 20th, 2015	Short term	Their own classifier that defines peoples mood after collecting the words from Twitter by using Python module called Tweepy	There is a certain degree of correlation twitter mood and stock market trend but not as expectations	-	The period of this study was needed to be expanded (for both types of data financial and sentiment analysis) to give better results and clearer view besides there are many procedural steps needs to be illustrated such as sentiment classifier
Skuza and Romanowski (2015)	Naive Bayes algorithm	Tweets collected for 3 months with Apple stock (historical data for the same period)	Twitter	From 2nd January 2013 to 31st March 2013	Short term	SentiWord Net and manual sentiment	This study concluded that the prediction process is highly depending on choosing the training dataset, the period of study and preparation method but still the main conclusions were there is a correlation between social media and stock market	The researcher tried to present the effect of social media (Twitter messages)on specific type of companies (Apple)	The period of study was not adequate to give better results Adding of information such as blogs articles and newspaper can be used to enrich this study
Calingo *et al*. (2016)	p-values and regression algorithms	Closing Index of the Philippine Stock Exchange and 800,000-850,000 Tweets were collected for 91 days	Twitter	From June 2 to August 31, 2015.	Short term	A Naive Bayes algorithm for classifying and Granger the different causality testing used for checking the usefulness of data for predicting	As a conclusion of this research helped in methods of sentiment analysis, classification method can give different results of predicting the stock market movement. The Tweets that use geo location and news sources gave better prediction results	Granger causality technique day can give determining the best predicting model	The more number of Tweets per better results. There is an absence of Neutral sentiment classification type. Using a specific dictionary to improve the text processing in sentiment analysis

Table 1: Continue

Research	Prediction Model	Data	Social network type	Period (long term, short term, medium term)	Time frame	Sentiment analysis classifier	Main conclusions	Strength	Weakness
Kordonis *et al*. (2016)	SVM algorithm	1.5 million Twitter messages. daily open, close, high and low values from Yahoo finance	Twitter	Not mentioned	Short term	Naive Bayes Bernoulli and SVM algorithms	Their main conclusion was the public mood change has a notable influence on the stock market movement	Applying the same algorithm for sentiment analysis and prediction process. There is a significant correlation between Twitter messages and stock market	Longer period, different sentiment classification algorithms, different languages, and different social media platforms may show better results
Nayak *et al*. (2016)	Support machine learning and decision Boosted Tree	Historical prices from Indian Stock Market from Yahoo finance	Twitter and News	From 2002 to July 2015	Short term	To determine the polarity of Tweets there was a comparison between the collections of Tweet words with a dictionary using the Hadoop software	The predictive model used in this model proved that decision boosted tree had better results than SVM and Logistic Regression	Using two sources of sentiment, mood analysis could give better indications	There are two types of sentiment classification positive and negative there is mention to neutral or even degree of the positive and negative Tweets
Wang and Wang (2016)	Correlation test and SVM	Historical data from Yahoo and Google API from 1st, 2009 to December 18th, 2015 (January 1st, 2009 to October 1st, 2015 for training and was from November 13th to December 18th, 2015 for testing) Sentiment I Index (SI)got from social networks	Sina Weibo, Tong Hua Shun Network and Dong Fang Cai Fu social network for 23 trading days from November 13th - December 18th, 2015	1st, 2009 to December 18th, 2015	Short term	Chinese segmentation tool ICTCLAS was used in word segmentation and How Net+ sentiment dictionary in words' labeling	They concluded that the change in sentiment, mood could change the stock market movement and the SVM algorithm gave higher prediction accuracy	Combining than one source of financial, social information to get the sentiment analysis	This study analyzed the Chinese messages only
Usmani *et al*. (2016)	Different machine learning techniques (Single Layer Perceptron (SLP), Multilayer Perceptron (MLP), Radial Basis Function (RBF) and Support Vector Machine (SVM)). The Simple Moving Average (SMA)and Autoreg -RessiveInte grated Moving Average (ARIMA) as input	The historical closing index of Karachi Stock Market KSE-100 for four months	Twitter	From September, 2015 to January, 2016	Short term	Text Mining techniques the Opinion Finder Library with giving the alternate words of non-English words	MLP technique gave the better prediction results among the used techniques Using	They used two different sources of sentiment information (News + Social media) multiple machine learning techniques gives the opportunity to check which one gives better prediction results	The period can be extended to give better results

Table 1: Continue

Research	Prediction Model	Data	Social network type	Period (long term, short term, medium term)	Time frame	Sentiment analysis classifier	Main conclusions	Strength	Weakness
Zhang *et al*. (2016)	Naive Bayes, Decision Tree (J48), Random Forest, Logistic and LibSVM algorithms	Chinese stock market-Shenzhen Stock Exchange (SZSE) and Shanghai Stock Exchange (SSE) data for 6 months	Weibo 8th, 2013 to March 31st, 2014	From October	Short term	A new Degree of Social Attention (DSA)	Their main conclusions were first the DSA sentiment can be used to improve the stock market shocks prediction and secondly Random Forest algorithm and SVM had the best and worst performance respectively and the last but not the least Negative social attention results gave better results than positive	This research combined the social media authors effect and personal mood in addition to their effect on social media	It was performed for Chinese speakers in addition toexpanding the period of study may give betterresults
Sun *et al*. (2016)	Sparse matrix factorization	Major listed stocks in S and P (SMF) 500 index (1,173 trading days) and for 420 component stocks with approximately 45 million messages from StockTwits streams for the same period	StockTwits .com.	From January 1st, 2011 to August 31st, 2015	Short term	Text mining techniques (tf-idf, normal--ization and Euclidian norm to divide the text vector)	The study main conclusions were that the proposed system with SMF and Stock Twits gave the researchers better than other basic models The second conclusion was the prediction accuracy will not be improved by increasing predictions accuracy	They divided their prediction process into four divisions and compare their SMF prediction results with Autoregressive (AR) and Rondom	The cleaned emoticons which may give some indications about their mood.Using other social mediaplatforms and online information suchasblogs Twitter, Facebook may give better prediction results
Oliveira *et al*. (2017)	Multiple Regression (MR), Neural Network (NN), Support Vector Machine (SVM), Random Forest (RF) and an Ensemble Averaging (EA) method	About 31 Million Tweets For financial data Standard & Poor's 500 (SP500), Russell 2000 (RSL) Dow Jones Industrial Average (DJIA), Nasdaq 100 (NDQ), excess return on the market (RMRF), Small Minus Big (SMB), High Minus Low (HML), Momentum Factor (MOM), CBOE Volatility Index (VIX), Portfolios formed on size (Psize) and 10 industry portfolios (Pind)	Twitter	From 22nd of December 2012 to 29th of October 2015	Short term	Twitter sentiment measures and text mining techniques	The main conclusion was there is a strong effect of microblogging sentiment data and stock market movement	This study studied many financial data such as volatility, trading volume and portfolio besides many indexes. Using different prediction algorithms	They don't take into their consideration thefollowers of twitter accounts, blogs, other information sourcesthat can give betterviews

Table 1: Continue

Research	Prediction Model	Data	Social network type	Period (long term, short term, medium term)	Time frame	Sentiment analysis classifier	Main conclusions	Strength	Weakness
Li *et al.* (2017)	Long Short-Term Memory (LSTM) neural network	18 million posts and 1263 trading days for daily Chinese stock market CSI300 index (open-values)	Stock forums of CSI300 index	January 1st, 2009 to October 31st, 2014 Short term	Naive Bayes classifier	They concluded that this method can be applied in different stock market prediction with different variables and at entirely different time-scales		This model shows the strength of using sentiment analysis in learning stock market time series as a strong noise	The increasing number of factors used in prediction can improve the accuracy results

taking the top words for the studied years. This dictionary contains some words such as: “sell”’ “buy”, “short”, etc. The term-document matrix was created where the rows represent the terms while the colomns represnet the documents. The system acquired its data for the period from January 1, 2011 to August 31, 2015 on the major listed stocks of S and P 500 index (1,173 trading days) and took approximately, 45 million messages from StockTwits streams for the same period. Their main conclusion was using SMF Model performs good prediction results while the increase of the predictions frequency will not increase the prediction accuracy.

One of the research that studied the role of sentiment and attention indicators to forecast stock market is Oliveira *et al*. (2017). They suggested a method to predict the stock market variables (returns, volatility and trading volume of diverse indices and portfolios) In this study, they tested five different regression methods. These models are: Multiple Regression (MR), Neural Network (NN), Support Vector Machine (SVM), Random Forest (RF) and an Ensemble Averaging (EA) method. The micblogging data were acquired from Twitter and in order to filter the indicators Kalman filter was used. In order to validate the results several machine learning methods. The results give an indication that the microblogging has a significant effect on helping financial experts in stock market prediction.

Li *et al*. (2017) suggest a model a new stock market prediction model that uses Long Short-Term Memory (LSTM) neural network that has only 4 layers and 30 nodes with the assistance of market factors and sentiment analysis of the investors. Naive Bayes was used in sentiment analysis to classify the forum posts into positive, negative or neutral for more than 18 million posts collected from constituent stocks of CSI 300 index for the period January 1, 2009 and October 31, 2014 (1263 trading days). The stock market data was collected for the daily CSI300 index (open-values) from wind database. The training classified posts was classified manually (500 posts) and the results were 1586 positive, 1765 negative and 1649 neutral posts. They used LSTM neural network in order to enhance the prediction process results. The prediction accuracy of the entire training data set was 87.86%.

CONCLUSION

The stock market prediction is a complex and difficult process because of its ambiguous nature and because of the stock market affected by many and different types of information. Through this study one can conclude that the sentiment analysis can be considered as an assistance factor in the prediction process. Through this review one can conclude that the social media effect can be used as an effective factor for short term prediction (>1 year) and Twitter as a social media has the first rank of studying social media but that does not deny the existence of other social networks such as Facebook to effect on stock market movements. Different social media can impact on a different stock market in different countries such as weibo which has the most popular microblogging social network in Chine where it is just like a hybrid of Twitter and Facebook. Many of these researches uses complicated NLP techniques in their researches in order to extract the useful (sentiment) information. Table 1 illustrated the reviewed study with an explanation of the prediction algorithms, the data, the sentiment information sources, sentiment classifiers, time frame, main conclusions, strength and weaknesses.

REFERENCES

Abbood, M.O., 2016. Investment of securities in the light macro risks: An economic study of Iraqi stock exchange in the period (2005-2014). Ph.D Thesis, University of Kufa, Kufa, Iraq.

Al-Augby, S.H., 2015. Text mining method in evaluation of media’s impact on market value ratios. Ph.D Thesis, University of Szczecin, Szczecin, Poland.

Anonymous, 2017. The top 500 sites on the web. Alexa Internet, San Francisco, California, USA. https://www.alexa.com/topsites.

Aqel, D. and S. Vadera, 2013. A Framework for Employee Appraisals Based on Inductive Logic Programming and Data Mining Methods. In: Natural Language Processing and Information Systems, Metais, E., F. Meziane, M. Saraee, V. Sugumaran and S. Vadera (Eds.). Springer, Berlin, Germany, ISBN:978-3-642-38823-1, pp: 404-407.

Bollen, J., H. Mao and X. Zeng, 2011. Twitter mood predicts the stock market. J. Comput. Sci., 2: 1-8.

Cakra, Y.E. and B.D. Trisedya, 2015. Stock price prediction using linear regression based on sentiment analysis. Proceedings of the 2015 International Conference on Advanced Computer Science and Information Systems (ICACSIS'15), October 10-11, 2015, IEEE, Depok, Indonesia, ISBN: 978-1-5090-0363-1, pp: 147-154.

Calingo, A.R., A.M. Sison and B.T. Tanguilig III, 2016. Prediction model of the stock market index using Twitter sentiment analysis. Intl. J. Inf. Technol. Comput. Sci., 8: 11-21.

Chen, C., W. Dongxing, H. Chunyan and Y. Xiaojie, 2014. Exploiting social media for stock market prediction with factorization machine. Proceedings of the 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT) V ol. 2, August 11-14, 2014, IEEE, Warsaw, Poland, ISBN:978-1-4799-4143-8, pp: 142-149.

Chen, Z. and X. Du, 2013. Study of stock prediction based on social network. Proceedings of the 2013 International Conference on Social Computing (SocialCom'13), September 8-14, 2013, IEEE, Alexandria, Virginia, USA., ISBN:978-0-7695-5137-1, pp: 913-916.

Daniel, M., R.F. Neves and N. Horta, 2017. Company event popularity for financial markets using Twitter and sentiment analysis. Expert Syst. Appl., 71: 111-124.

Duong, D., T. Nguyen and M. Dang, 2016. Stock market prediction using financial news articles on Ho Chi Minh stock exchange. Proceedings of the 10th International Conference on Ubiquitous Information Management and Communication, January 04-06, 2016, ACM, Danang, VietNam, ISBN:978-1-4503-4142-4, pp: 71:1-71:6.

Hamed, A.R., R. Qiu and D. Li, 2015. Analysis of the relationship between Saudi Twitter posts and the Saudi stock market. Proceedings of the 2015 IEEE 7th International Conference on Intelligent Computing and Information Systems (ICICIS'15), December 12-14, 2015, IEEE, Cairo, Egypt, ISBN:978-1-5090-1950-2, pp: 660-665.

Han, J., J. Pei and M. Kamber, 2011. Data Mining: Concepts and Techniques. Elsevier Publishing company, Amsterdam, Netherlands, ISBN:978-0-12-381479-1, Pages: 702.

Kordonis, J., S. Symeonidis and A. Arampatzis, 2016. Stock price forecasting via sentiment analysis on Twitter. Proceedings of the 20th Pan-Hellenic Conference on Informatics, November 10-12, 2016, ACM, Patras, Greece, ISBN:978-1-4503-4789-1, pp: 36-41.

Kumar, B.S. and V. Ravi, 2016. A survey of the applications of text mining in financial domain. Knowl. Based Syst., 114: 128-147.

Li, J., H. Bu and J. Wu, 2017. Sentiment-aware stock market prediction: A deep learning method. Proceedings of the 2017 International Conference on Service Systems and Service Management (ICSSSM'17), June 16-18, 2017, IEEE, Dalian, China, ISBN:978-1-5090-6371-0, pp: 1-6.

Li, Q., B. Zhou and Q. Liu, 2016. Can Twitter posts predict stock behavior?: A study of stock market with Twitter social emotion. Proceedings of the 2016 IEEE International Conference on Cloud Computing and Big Data Analysis (ICCCBDA'16), July 5-7, 2016, IEEE, Chengdu, China, ISBN:978-1-5090-2595-4, pp: 359-364.

Li, Q., Y. Jiang, L.L. Jiang, P. Li and H. Chen, 2016. A tensor-based information framework for predicting the stock market. ACM. Trans. Inf. Syst., 34: 11:1-11:30.

Liu, L., J. Wu, P. Li and Q. Li, 2015. A social-media-based approach to predicting stock comovement. Expert Syst. Appl., 42: 3893-3901.

Luo, F., J. Wu and K. Yan, 2010. A novel nonlinear combination model based on support vector machine for stock market prediction. Proceedings of the 8th World Congress on Intelligent Control and Automation (WCICA'10), IEEE, Jinan, China, ISBN: 978-1-4244-6710-5, pp: 5048 5053-10.1109/WCICA.2010.5554607.

Ming, F., F. Wong, Z. Liu and M. Chiang, 2014. Stock market prediction from WSJ: Text mining via sparse matrix factorization. Proceedings of the 2014 IEEE International Conference on Data Mining (ICDM'14), December 14-17, 2014, IEEE, Shenzhen, China, ISBN:978-1-4799-4302-9, pp: 430-439.

Nayak, A., M.M. Pai and R.M. Pai, 2016. Prediction models for Indian stock market. Procedia Comput. Sci., 89: 441-449.

Nguyen, T.H. and K. Shirai, 2015. Topic modeling based sentiment analysis on social media for stock market prediction. Proceedings of the 7th International Joint Conference on Natural Language Processing and 53rd Annual Meeting on Association for Computational Linguistics Vol. 1, July 26-31, 2015, Association for Computational Linguistics, Vancouver, Canada, pp: 1354-1364.

Nofer, M. and O. Hinz, 2015. Using Twitter to predict the stock market. Bus. Inf. Syst. Eng., 57: 229-242.

Oliveira, N., P. Cortez and N. Areal, 2017. The impact of microblogging data for stock market prediction: Using Twitter to predict returns, volatility, trading volume and survey sentiment indices. Expert Syst. Appl., 73: 125-144.

Pang, B. and A.L. Lee, 2004. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. Proceedings of the 42nd Annual Meeting of the Association for Computational, November 16-24, 2004, Kimberly Patch, pp: 271-278.

Porshnev, A., I. Redkin and A. Shevchenko, 2013. Machine learning in prediction of stock market indicators based on historical data and data from Twitter sentiment analysis. Proceedings of the 2013 IEEE 13th International Conference on Data Mining Workshops (ICDMW'13), December 7-10, 2013, IEEE, Dallas, Texas, ISBN: 978-1-4799-3142-2, pp: 440-444.

Qiu, L., H. Rui and A. Whinston, 2013. Social network-embedded prediction markets: The effects of information acquisition and communication on predictions. Decis. Support Syst., 55: 978-987.

Saunders, A. and M.M. Cornett, 2003. Financial Institutions Management: A Risk Management Approach. Irwin-McGraw-Hill, New York, USA.,.

Skuza, M. and A. Romanowski, 2015. Sentiment analysis of Twitter data within big data distributed environment for stock prediction. Proceedings of the 2015 Federated Conference on Computer Science and Information Systems (FedCSIS'15), September 13-16, 2015, IEEE, Lodz, Poland, ISBN:978-8-3608-1065-1, pp: 1349-1354.

Sun, A., M. Lachanski and F.J. Fabozzi, 2016. Trade the tweet: Social media text mining and sparse matrix factorization for stock market prediction. Intl. Rev. Financial Anal., 48: 272-281.

Taboada, M., 2016. Sentiment analysis: An overview from linguistics. Ann. Rev. Ling., 2: 325-347.

Usmani, M., S.H. Adil, K. Raza and S.S.A. Ali, 2016. Stock market prediction using machine learning techniques. Proceedings of the 3rd International Conference on Computer and Information Sciences (ICCOINS'16), August 15-17, 2016, IEEE, Kuala Lumpur, Malaysia, ISBN:978-1-5090-2550-3, pp: 322-327.

Vu, T.T., S. Chang, Q.T. Ha and N. Collier, 2012. An experiment in integrating sentiment features for tech stock prediction in Twitter. Proceedings of the Workshop on Information Extraction and Entity Analytics on Social Media Data, December 9, 2012, Institute of Distance and Open Learning, Mumbai, India, pp: 23-38.

Wang, Y. and Y. Wang, 2016. Using social media mining technology to assist in price prediction of stock market. Proceedings of the 2016 IEEE International Conference on Big Data Analysis (ICBDA'16), March 12-14, 2016, IEEE, Hangzhou, China, ISBN:978-1-4673-9592-2, pp: 1-4.

Weston, J.F., E.F. Brigham and S. Besley, 1996. Essentials of Managerial Finance. 11th Edn., Dryden Press, New York, USA., ISBN: 9780030168895, Pages: 788.

Xu, F. and V. Keelj, 2014. Collective sentiment mining of microblogs in 24-hour stock price movement prediction. Proceedings of the 2014 IEEE 16th Conference on Business Informatics (CBI'14) Vol. 2, July 14-17, 2014, IEEE, Geneva, Switzerland, ISBN:978-1-4799-5779-8, pp: 60-67.

Zhang, L., L. Zhang, K. Xiao and Q. Liu, 2016. Forecasting price shocks with social attention and sentiment analysis. Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM'16), August 18-21, 2016, IEEE, San Francisco, California, ISBN:978-1-5090-2847-4, pp: 559-566.

Zhang, X., H. Fuehres and P.A. Gloor, 2011. Predicting stock market indicators through twitter i hope it is not as bad as i fear. Procedia Soc. Behav. Sci., 26: 55-62.