A SYSTEMATIC REVIEW ON THE RELATIONSHIP BETWEEN STOCK MARKET PREDICTION MODEL USING SENTIMENT ANALYSIS ON TWITTER BASED ON MACHINE LEARNING METHOD AND FEATURES SELECTION

[1]GHAITH ABDULSATTAR A.JABBAR ALKUBAISI, [2]SITI SAKIRA KAMARUDDIN, [3]HUSNIZA HUSNI

[1,2,3]School of Computing, Universiti Utara Malaysia, Sintok 06010, Malaysia

E-mail: [1]ghaith.alkubaisi@outlook.com

ABSTRACT

This study is mainly a systematic review analysis which discusses studies related to the role of sentiment analysis, Twitter data, and features in predicting stock market returns. Studies show that it is not only the historical financial data of firms or stock markets that can predict the returns of the stock market, but sentiments and emotions of people can also help in predicting stock market returns. One primary source of information that is available now to everyone is Twitter, and tweets made by significant personals affect the emotions of people and which will ultimately affect their investment decisions. If the news is positive, then most probably it will affect people positively, and they will invest more in stocks of that firm. If the news is negative, the reaction of people is expected to be opposite. Besides the sentiments of people, there are features like spatial and temporal that can also affect the stock market returns. The spatial feature is a geographical division, it can either be different emotions of different people from different geographical regions, or they can be other stock markets which can affect the home stock market by any relation. Similarly, temporal effect shows the change in something over a span of time. People might have different opinions at different times, and they can behave differently according to their sentiments at that specific time. Finally, all these factors help us in predicting the future stock market returns.

Keywords: *Sentiment Analysis, Features, Spatial, Temporal, Stock Market*

1. INTRODUCTION

Nowadays, with a global increment of the stock exchanges has raised the need for decision making using a stock market prediction model [6]. A stock market prediction model using sentiment analysis on Twitter needs an accurate classification model to measure the tweets sentiment analysis [50]. The required prediction model works as an assistant to the decision-makers in the field of the stock market to make right decisions [37].

Current stock market prediction models are still suffering low accuracy in classification [59, 26, 21, 4]. The low accuracy in classification has a direct effect on the reality and the reliability of stock market indicators like a series of statistical figures and financial reports which explain the stock returns in existing stock market [9, 32, 33]. There are different factors like features [11, 24, 54] labeling technique [23] and classification methods [43] that can affect the accuracy of the stock market prediction model.

This study focuses on the features selection by defining two essential features for the stock market prediction model; they are geographical location and timestamp. At the same time, this study explains the role of these two features to increase the accuracy of the prediction model. Data without timestamp and geographical location cannot support the decision makers in the field of stock market exchange [44, 18, 52] and the sentiment analysis will not achieve the fine-grained level [9, 52, 34].

On the other hand, before moving to the next section we should explain four essential concepts, they are sentiment analysis, machine learning method, sentiments of people from Twitter and, spatial and temporal feature in section 1.1, 1.2, 1.3 and 1.4 respectively.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

1.1 Sentiment Analysis and its Uses Related to Stock Market

Sentiment analysis described as the study of people's emotions, their attitude or returns towards specific news or events and their opinions towards these events. Sentiment analysis is customarily considered to be the result of procedure; this procedure starts from the identification of a person's sentiment or emotion towards an event, then the selections of feature and classifying that sentiment and at the end, we get the polarity of that sentiment.

Sentiment analysis has three major classifications: the first one is called document-level classification it just finds that whether a sentiment or emotion expressed by a person is either positive or negative. The second one is called sentence level, it studies the sentiments in sentences, whether these sentences are of subjective or objective nature. Moreover, after that, they also classify their opinions to be negative or positive. The third one is aspect level classification, and it studies different aspects of opinions or emotions expressed by different persons. Some aspects can have positive opinions, some can have negative opinions, and some can have both positive and negative opinions [45, 58].

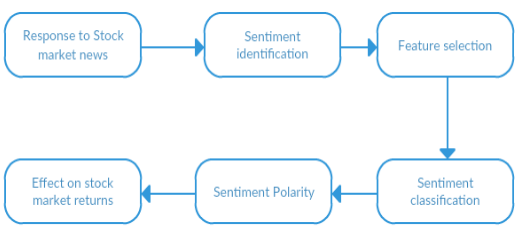

Figure 1: The process of sentiment analysis on market news and effect on stock market

Figure 1 exhibits the whole process when a person gives his opinion about news, and after that, his sentiment is observed from that news whether it is positive or negative, and after that, his response or reaction on the stock market observed. It is essential to keep in mind that a single person cannot affect stock market it is the result of mutual actions of groups of people.

After conducting studies on sentiments of people, company's officials know that what kind of news can negatively affect the public, so they can avoid that kind of news in future. It will also help the firm in launching a new product as they already know about the sentiment of people and can predict their reactions. So, the firm will act accordingly, and it will have a positive effect on stock returns. A firm can have a comparative advantage over rival firms as firm's management knows their shareholder's psychology [2, 12].

1.2 Machine Learning Technique

Machine learning is programming computers to optimize a performance criterion using example data or experience [3]. The goal of machine learning is to develop methods that can automatically detect patterns in data and then to use the uncovered patterns to predict future data or other outcomes of interest [36].

Step 4 in figure 1 has explained in figure 2 in which classification techniques have discussed. In which different techniques like machine learning and lexicon and further classifications of these two techniques have discussed in detail.

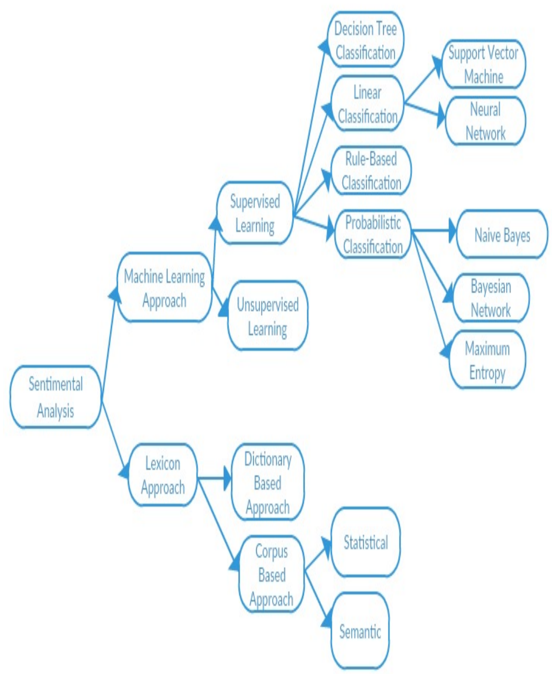

Figure 2: Sentiment classification techniques

From figure 2, a machine learning approaches divided into two parts. First one is supervised and second one is unsupervised. Supervised learning has further parts like a decision tree, linear and probabilistic classification and so forth. The probabilistic classification has divided into byes types, and linear has divided into neutral network and network machine. Overall this whole process helps us in learning the process of sentiment analysis [45-42].

1.3 Sentiments of People from Twitter

Social media has become the center of attraction and news since last decade; previously people used to do sentiment analysis by using news from different sources. However today it is more comfortable. Data from Twitter is readily available to everyone, and we can predict several things from that news. For example, if the Chief Executive Officer (CEO) of a company who has an account on Twitter announce that their company has acquired shares of another company or their company has announced a profit, so it will ultimately affect the stock returns of their company positively as news is positive. On the other hand, a negative news announcement like declaring loss will affect negatively. Similarly, announcements made by government officials also can affect stock markets of that country. Twitter has its significance as most of the government official, company's management personals, and the public, and so on are on Twitter. At the same time, Twitter allows a post limited to 140 words only, so people are precise about what they want to say and extraction of their thinking and mood is easy to find from their statements [10, 22, 48, 28, 57, 17, 40].

While talking about sentiments of people from Twitter data, it is also important to talk about their location and time of expressing their views and thoughts. Spatial features explain the geographical location of people; it is entirely possible that people who are expressing their thoughts about an event have nothing to do with event directly or indirectly. Alternatively, it is also possible that some groups of people have a different sentiment about an event and another group has a different opinion. For example, Apple has launched a new product some people will like it, and some will not. Temporal feature mean studying the timing of tweets, as it is also possible that people have different opinions at different times [5, 8, 41].

1.4 Spatial and Temporal Feature

Spatial and temporal sentiment analysis used geotagged Twitter data, which allow a sentiment polarity analysis at a fine-grained level. This method has applied in many areas, such as polls [55], consumer opinions concerning brands [25], stock market performance [10], crime prediction [56] and tourism information [47].

However, it is essential to understand how opinions grow on Twitter over location and timestamp across communities of users. So, what is a spatial and temporal feature? The answer in the following two examples: (a.) spatial, Tweets posted in different locations are various in sentiments. For example, Tweets posted in Entertainment and marketing area may be more positive while Tweets posted in study areas may be more negative. (b.) Temporal, [1] researched student stress levels at the beginning and the end of the semester and found the stress level of university students varies throughout a semester. Students felt more stressful at the end of the semester than at the beginning of the semester.

The rest of this study discusses and compares different studies related to these issues.

2. RELATED WORK

This study primarily follows the technique of Kitchenham and Charters [27] which includes three steps for the study

1. Planning the review
2. Selection of relevant and crucial studies
3. Reporting the final review

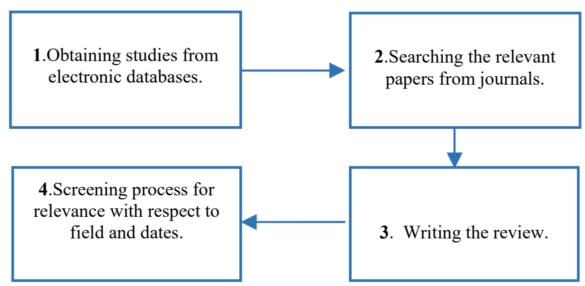

Figure 3: Planning and review process

Figure 3 explains the process of planning the review level.

This study aims to define the relationship between stock market prediction accuracy and features selection by addressing specific studies are using sentiment analysis on financial news and reports based on machine learning classifiers, and studies are using sentiment analysis on Twitter based on machine learning classifiers. So, an essential part of this study is to clarify and define the study questions and using literature from relevant fields:

This study mainly focuses on a review of those studies which address the following study questions:

RQ1: Which studies are studying the sentiment analysis?

RQ2: Which techniques have currently used for this analysis?

RQ3: Which study is using machine learning methods for sentiment analysis?

RQ4: Which studies is using Twitter as a dataset for their study in sentiment analysis?

RQ5: Which studies are addressing features in sentiment analysis?

RQ6: Which studies are using these sentiment techniques to predict stock markets returns?

Stock markets are directly related to sentiments of people especially to those people who are engaged in this business. Nowadays, different institutions use to read news and articles related to stock markets and try to identify the aspects of those news whether they will affect stock market positively or negatively, ultimately trying to predict the trends or returns of stocks. Identification of these trends helps the people in making their investment decisions.

Table 1 shows some recent essential studies related to stock market returns and sentiment analysis. These studies have done different countries, and every one of them has used different techniques and tools to find that emotions and sentiments of people also play an important role in predicting stock returns and returns. Yu et al. [58] performed a study on sentiments related to stock market returns predictions. They found that contextual distributions of words in a sentence are significant in the determination of positive or negative aspect. They also found that besides positive or negative aspect of news, the intensity of that news is also significant. Intensity measure the impact of that aspect on stock returns. The new proposed model by them was called a contextual entropy model. This model combines the aspect of news (whether it is positive or negative) with the intensity of news so that predictions made in trends of stock markets should be more precise and accurate. It the data set they collected 6,888 news related to stock markets. Moreover, applied their proposed contextual entropy model and found that predictions of stock market trends were improved.

Chan and Franklin [15] argued that although there are many traditional financial methods in the market which use financial data to predict trends in stock markets. However, they say that other than that there are many other methods as well like predicting a trend following news related to stock market. For this purpose, they introduced a new model called decision support system which studies the texts of news and helps in forecasting trends. They collected data from 2000 different reposts and studied the texts of those reports. They found 28,000 sentences and classified that type so that predictions have made. They found that their model predicts better results as compared to other models.

Schumaker et al. [45] also studied the question related to the impact of sentiments associated with different news on the prediction of stock markets. They studied articles from texts of "Arizona financial text system." However, their results were entirely different as they found that people were selling stock when there was good news and buying when there was bad news. However, their results were significant.

The three previous studies have adopted financial reports and news for sentiment analysis. In upcoming sections following essential issues are discussed in this study:

1. Role of Twitter in predicting Stock markets returns.
2. Spatial and temporal studies in stock market
3. Spatial and temporal methods using Twitter as database
4. Benefits of using sentiment analysis in stock markets

2.1 Role of Twitter in Predicting Stock Markets

Table 1. Shows studies in which data from Twitter has been collected to measure the sentiments of people in predicting the stock returns. Ahuja et al. [2] performed a study to find the relation between news from Twitter and its effect on stock returns. They argued that emotions of people have attached with different news annulments, so it affects the stocks of those firms or the whole stock market. They did their study on data of Bombay stock market. They used Sentiment analysis technique in their study. First, they asked questions form people to inquire about their mood or sentiments attached to different news using a questionnaire, after that, they compared that data with Twitter news and financial data of stock returns. They found a significant relationship between sentiments of people resulting from Twitter news and its impact on stock market returns.

Bartov et al. [6] also conducted a similar study, but his dataset was quite large as he used data from more than 99,000 tweets and more than

34,000 companies. They classified their data into positive and negative news related to companies and their effect on sentiments of people. He found that opinion of people following news from Twitter can not only predict the return on stocks, but it can also predict the earnings announcements related things of the company as well.

Bollen et al. [10] also argued that sentiments of a personal effect their decision-making power, the used the data of over 9 million tweets and they studied the text content of these tweets to find out the sentiments of public and the mood of people. They used opinion finder technique to study the mood of people. They also studied the effect of mood on prices of DIJA prices. They found that changes in the mood of people affect the prices of DIJA after 3 to 4 days.

Hamed et. Al [22] also did a similar study to find a relation between stocks returns and sentiments of people (using data from Twitter) in KSA. Their data included more than 3300 tweets covering more than 50 days. They used opinion mining techniques to find out the mood of people and correlation analysis between mood and Saudi stock index. They found that mood of the public has correlated with change in the index of the stock market.

Skuza and Romanowski [49] studied the relation between tweets and market returns. They used machine learning approach in their model to find the relation between these two variables. They used manual labeling technique to study the sentiments of people and studied its relationship with a return in the stock market with different intervals, i.e., after 1 hour, 30 hours and five mins, and so forth. They found correlations between these two variables. Concluding that mood of the public can help us in predicting returns on stocks.

Şimşek and Özdemir [48] did a straightforward study by finding frequencies of only two words happiness and unhappiness. They used data of Turkish stock market and tweets. They made simple graphs, and frequency distributions see the relation between tweets (mood) and return on the stock market index and found that there is a relation between these two variables. Good mood increases the investment and hence increase the returns while lousy mood does the opposite.

Cakra and Trisedya [12] did a study on Indonesian stock markets using data from Twitter by finding sentiments of people and tried to predict the market returns. They used sentiment analysis techniques and simple linear regression. They also found that there is a relation exists between these two variables.

So, these studies show that it is besides using statistical measures related to financial data (historical data) to predict stock markets, emotions of people also matter in this case. Sentiments of people and their opinions related to an event or in a time frame can also predict the stock market returns. If people are happy they will more like to invest in stocks and returns will grow and opposite will happen if they are not happy. To measure/estimate the mood of people simple sentiment analysis methods have used, and data is readily available form Twitter.

Table1: Relation Between Sentiments and Stock Returns

Authors	Tools / model	DATA used	Using Timestamp and geographical location	Relation
Yu et al. [58]	Contextual entropy model	Stock market news articles	No	yes
Chan and Franklin [15]	Novel text-based decision support system	2000 financial reports with 28,000 sentences	No	yes
Schumaker et al. [45]	Sentiment analysis tool	Arizona Financial Text (AZFinText) system	No	yes
Ahuja et al. [2]	Sentiment analysis	Bombay stock exchange and Twitter	No	yes
Bartov et al. [6]	Regression model /Correlations	Twitter,	No	yes
Bollen et al. [10]	OpinionFinder/ Granger causality analysis	Twitter, Dow Jones industrial average	No	yes
Hamed et al. [22]	Opinion mining/ Correlations	KSA Twitter and stock market data	No	yes
Skuza and Romanowski [49]	Manual Labeling/ SentiWordnet	Twitter data, stock market returns	No	yes
Şimşek and Özdemir [48]	Frequency tables, Graphs	Turkish Twitter data, stock market returns	No	yes
Cakra and Trisedya [12]	Sentiment analysis	Twitter, Indonesian stock market	No	yes

Journal of Theoretical and Applied Information Technology
31st December 2017. Vol.95. No 24
© 2005 – ongoing JATIT & LLS

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 1 shows that all models come with low accuracy in classification and they do not use timestamp and geographical location as a classification features.

Table2: The Latest Five Studies Have Used Sentiment Analysis on Twitter Based on Machine Learning Method

Authors	Tools / model	DATA used	Using timestamp and geographical location	Relation
Catal and Nangir [14]	Sentiment Analysis	Twitter	No	Yes
M.Vadivukarass [35]	Sentiment Analysis	Twitter	No	Yes
Garcia-Diaz et. al [19]	Sentiment Analysis	Twitter	No	Yes
Gupta et al. [20]	Sentiment Analysis	Twitter	No	Yes
Song et al. [51]	Sentiment Analysis	Twitter	No	Yes

Table 2 lists the latest five studies which are used sentiment analysis on Twitter based on machine learning classifiers like Naïve Bayes. All studies in table 2 have done in 2017, and we can observe that the latest studies also have not used timestamp and geographical location as features for classification which lead to affect the sentiment classification accuracy.

2.2 Spatial and Temporal Studies in Stock Markets

A lot of studies related to sentiment analysis involving spatial and temporal features have already done by many researchers [38, 13, 46, 7]. However, studies on these effects on stock markets are not very common.

Tam [53] did a study on stock markets of East Asian countries; this study was mainly of spatial and temporal nature. For the spatial feature, East Asia is a geographical region, and it has many essential countries like Japan, Hong Kong, Singapore, and so on. He used data from all countries located in this region and used the spatial-temporal model in his study. For the case of spatial analysis, he found that stock markets of neighbor countries also affect the stock market of the home country. i.e., returns in one market affect the other one. For the case of temporal analysis, he also found that crises in a country which affect the stock market of the home country also affect the stock markets of neighbor countries. Overall these markets were found to be interrelated with each other. Asgharian et al. [5] did a spatial study on an extensive data set of 16 years form 1995-2011. They studied different aspects of one country which can affect the stock markets of other countries as well. It has fundamentally based on the distribution of different countries based on their geographical locations. They found that geographical distance is a significant variable in determining the effect on the market of another country. Moreover, it trades between two countries is more than their stock markets will also affect each other more.

Ouyang et al. [39] did a study on four stock markets to see their co-movements over time and their effect on each other. These stock markets were from China, Taiwan, and Hong Kong. Their model based on spatial and temporal methods. They found that the main changes in one stock market affect the other stock market and the trend has changed over time. They found that stock markets of Hong Kong and Taiwan were exhibiting a leveraging effect. While two major stock markets in China were showing a positive correlation before 2000 and after that, it has changed to negative. So, this study shows that movement in one stock market can predict the movement in other stock markets which are geographically close to each other.

Lin et al. [31] did a study on correlations between Chinese and U.S stock exchanges. They used DFA and DCCA methods in their study. Three major stock indicators of U.S (S&P 500, NYA and DJI) have compared with three major stock indicators of the Chinese market. They used data from 2002-2009. They found that there was cross-correlation between these two markets.

2.3 Spatial and Temporal Features of Using Twitter (Sentiment Analysis Approach)

Kucuktunc et al. [29] did a study on sentiments of people by using data from yahoo answers, they also spatial and temporal features in their study. They studied whether sentiments of people change due to their different age groups or due to their level of education and so on. They also studied these sentiments during different days of the week. They used correlations, graphical techniques, and tables on their data and found that generally on weekend's people use to exhibit more positive returns as compared to regular days. Usually, people who were younger express great opinions about a question. They used data of 1-year from Oct 2009 to Sep 2010 in their study.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A sentiment analysis study using data from Twitter and their region of focus was New York City. Studying the spatial part of their study, they found that intensity in sentiments of people is more in areas like medical centers and jails, and so forth. They also found the temporal effect and their results show that positive effect of tweets was more observed in weekends as compared to weekdays. They used graphical techniques in their study to observe these patterns.

Li et al. [30] did a study by collecting data from Twitter and flicker; they studied the tweets patterns of different types of people by their level of education and by the areas in which they live. They used graphical techniques and partial least square method in their study. Their primary focus was on spatial and temporal patterns of these tweets. They found that photos taken at tourist places show different spatial and time patterns. For example, in weekends photos posted on social media were much more as compared to a regular weekday. They also found that people who are more educated are more involved in georeferenced tweets.

Cheng and Wicks [16] argued that data from tweets can be helpful in detecting an event happening somewhere. As they argued that different kinds of events happen every day and millions of people talk about those events through Twitter. They applied space-time scan statistics method in their study and argued that this method can help in detecting an event. They used helicopter crash in London as a case study so flowing geographical pattern of tweets they found that it is easy to get information from a cluster of data at that time because most of the people are talking about that event at that time and expressing their sentiments. Using the model of space-time scan statistics, it is also possible to detect other events like some matches and information related to flight delays. This method uses the keywords at a specific time and from those clusters of tweets event detection become easy.

Song and Xia [52] conducted a study using Twitter data of university students, and they divided their sample into different groups and different time periods. Different groups' means classifying them by their departments and different times means classifying them at different times of the semester, i.e., beginning, mid and end of the semester. Their findings were fascinating; they found that students who were in the social science department were posting positive things while applied science students were posting negative statements. Moreover, at the closing times of semester number of negative tweets was increasing. So, this study shows that is beside studying sentiment analysis it is also imperative to study the geographical location and timing of the tweet.

For example, if we take the case of an American firm the tweets posted in other countries or by other nationals may be irrelevant, and they may not affect the stock returns pf that company. While if we talk about the timing people may have different opinions before and after earnings announcements. So, it is also possible that sentiments of people can change from time to time and from location to location. If we talk about the stock market, a rumor spread during trading hours may affect the market more as compare to rumor spread after closing hours.

However, studies using the Twitter data and sorting that data into a geographical and timely basis to predict stock market returns are infrequent as it is a very complicated task to do and justifying steps will be complicated. However, on the other hand, it is also an opportunity for future work.

2.4 Benefits of Using Sentiment Analysis in Stock Markets

While talking about advantages, it has observed from above discussions that understanding of people's sentiments leads to better prediction of stock market, notably spatial and temporal features. Control over stock markets become easier after studying the statements of people related to issues. They give a new dimension in the field of financial research and will help in future to generate new research ideas. It makes easy for companies to plan their business projects according to sentiments of people.

While if we discuss the disadvantages, we can see that sometimes data is not available for every region (Some countries do not have proper records) so this will not work there. Discarding and identification of irrelevant data from millions of statements is very time taking and critical job and misunderstanding of statements can lead to the wrong prediction of stock markets.

3. RECOMMENDATIONS AND FUTURE WORKS

We recommend that, adopt spatial and temporal features is necessary for a stock market classification model because they will lead to increase the importance and value of generated information besides raising the reliability of the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

produced reports and indicators. At the same time, they will be reducing the risk of the decision taken in the domain of stock market exchange because it will command to increase the classification accuracy.

For future work, doing a spatial and temporal analysis of Twitter data to predict stock returns will be very productive, but this is a very tough task as separating tweets based on a geographical basis and justifying the relevant tweets will be an important and challenging at the same time. Doing the same study to find out changes in sentiments of people across different times will also contribute a lot to this field. Spatial and temporal segregation of tweets and applying sentiment analysis on that to predict returns in stock can be very useful and helpful study in future in this field.

4. CONCLUSION

Sentiments of people related to tweets play an important role in predicting stock market returns. However, to measure the sentiment from statements using sentiment analysis is challenging and necessary tasks because it involves analysis of statements and finding out positive negative and neutral emotions from those statements. Recently, due to the presence of social media especially Twitter most of the people express their thoughts using this platform. So, a collection of information has become more comfortable, and studies have found that emotions of people play an essential role in the prediction of stock market returns. The spatial and temporal analysis also helps in the prediction of these returns. Most of the stock exchanges are affected by other movements in related stock exchanges. It can be due to the closeness of their geographical presence or due to other factors like trade between two countries. If two countries have a massive trade, then movements in the stock exchange of one country can quickly affect the movement is the stock of another country.

This study could be the first study that explains the importance of spatial and temporal features for stock market prediction model and their role in classification accuracy of the Twitter dataset by using sentiment analysis based on machine learning classifiers.

REFERENCES:

[1] Adnan, N., Murat, Z. H., Kadir, R. S. S. A., & Yunos, N. H. M. (2012). University students stress level and brainwave balancing index: Comparison between early and end of study semester. Paper presented at the Research and Development (SCOReD), 2012 IEEE Student Conference on.

[2] Ahuja, R., Rastogi, H., Choudhuri, A., & Garg, B. (2015). Stock market forecast using sentiment analysis. Paper presented at the Computing for Sustainable Global Development (INDIACom), 2015 2nd International Conference on.

[3] Alpaydin, E. (2014). Introduction to machine learning: MIT press.

[4] Arvanitis, K., & Bassiliades, N. (2017). Real-Time Investors' Sentiment Analysis from Newspaper Articles Advances in Combining Intelligent Methods (pp. 1-23): Springer.

[5] Asgharian, H., Hess, W., & Liu, L. (2013). A spatial analysis of international stock market linkages. Journal of Banking & Finance, 37(12), 4738-4754.

[6] Bartov, E., Faurel, L., & Mohanram, P. S. (2015). Can Twitter Help Predict Firm-Level Earnings and Stock Returns? Available at SSRN 2631421.

[7] Beauchamp, N. (2017). Predicting and Interpolating State-Level Polls Using Twitter Textual Data. American Journal of Political Science, 61(2), 490-503.

[8] Bertrand, K. Z., Bialik, M., Virdee, K., Gros, A., & Bar-Yam, Y. (2013). Sentiment in new york city: A high resolution spatial and temporal view. arXiv preprint arXiv:1308.5010.

[9] Bollen, J., Mao, H., & Pepe, A. (2011). Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. ICWSM, 11, 450-453.

[10] Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. Journal of Computational Science, 2(1), 1-8.

[11] Bonde, G., & Khaled, R. (2012). Extracting the best features for predicting stock prices using machine learning. Paper presented at the Proceedings on the International Conference on Artificial Intelligence (ICAI).

[12] Cakra, Y. E., & Trisedya, B. D. (2015). Stock price prediction using linear regression based on sentiment analysis. Paper presented at the 2015 International Conference on Advanced Computer Science and Information Systems (ICACSIS).

[13] Caragea, C., Squicciarini, A. C., Stehle, S., Neppalli, K., & Tapia, A. H. (2014). Mapping

moods: Geo-mapped sentiment analysis during hurricane sandy. Paper presented at the ISCRAM.

[14] Catal, C., & Nangir, M. (2017). A sentiment classification model based on multiple classifiers. Applied Soft Computing, 50, 135-141.

[15] Chan, S. W., & Franklin, J. (2011). A text-based decision support system for financial sequence prediction. Decision Support Systems, 52(1), 189-198.

[16] Cheng, T., & Wicks, T. (2014). Event detection using Twitter: a spatio-temporal approach. PloS one, 9(6), e97807.

[17] Dickinson, B., & Hu, W. (2015). Sentiment analysis of investor opinions on twitter. Social Networking, 4(03), 62.

[18] Fernández-Avilés, G., Montero, J.-M., & Orlov, A. G. (2012). Spatial modeling of stock market comovements. Finance Research Letters, 9(4), 202-212.

[19] García-Díaz, V., Espada, J. P., Crespo, R. G., G-Bustelo, B. C. P., & Lovelle, J. M. C. (2017). An approach to improve the accuracy of probabilistic classifiers for decision support systems in sentiment analysis. Applied Soft Computing.

[20] Gupta, B., Negi, M., Vishwakarma, K., Rawat, G., & Badhani, P. (2017). Study of Twitter Sentiment Analysis using Machine Learning Algorithms on Python. International Journal of Computer Applications, 165(9).

[21] Hajli, M. N. (2014). The role of social support on relationship quality and social commerce. Technological Forecasting and Social Change, 87, 17-27.

[22] Hamed, A.-R., Qiu, R., & Li, D. (2015). Analysis of the relationship between Saudi twitter posts and the Saudi stock market. Paper presented at the 2015 IEEE Seventh International Conference on Intelligent Computing and Information Systems (ICICIS).

[23] Hasbullah, S. S., Maynard, D., Chik, R. Z. W., Mohd, F., & Noor, M. (2016). Automated Content Analysis: A Sentiment Analysis on Malaysian Government Social Media. Paper presented at the Proceedings of the 10th International Conference on Ubiquitous Information Management and Communication.

[24] Iacomin, R. (2016). Feature optimization on stock market predictor. Paper presented at the Development and Application Systems (DAS), 2016 International Conference on.

[25] Jansen, B. J., Zhang, M., Sobel, K., & Chowdury, A. (2009). Twitter power: Tweets as electronic word of mouth. Journal of the American Society for Information Science and Technology, 60(11), 2169-2188.

[26] Kim, S., & Park, H. (2013). Effects of various characteristics of social commerce (s-commerce) on consumers' trust and trust performance. International Journal of Information Management, 33(2), 318-332.

[27] Kitchenham, B., Budgen, D., Brereton, P., Turner, M., Charters, S., & Linkman, S. (2007). Large-scale software engineering questions–expert opinion or empirical evidence? IET Software, 1(5), 161-171.

[28] Kouloumpis, E., Wilson, T., & Moore, J. D. (2011). Twitter sentiment analysis: The good the bad and the omg! ICWSM, 11(538-541), 164.

[29] Kucuktunc, O., Cambazoglu, B. B., Weber, I., & Ferhatosmanoglu, H. (2012). A large-scale sentiment analysis for Yahoo! answers. Paper presented at the Proceedings of the fifth ACM international conference on Web search and data mining.

[30] Li, L., Goodchild, M. F., & Xu, B. (2013). Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr. cartography and geographic information science, 40(2), 61-77.

[31] Lin, A., Shang, P., & Zhao, X. (2012). The cross-correlations of stock markets based on DCCA and time-delay DCCA. Nonlinear Dynamics, 67(1), 425-435.

[32] Lin, J., & Ryaboy, D. (2013). Scaling big data mining infrastructure: the twitter experience. ACM SIGKDD Explorations Newsletter, 14(2), 6-19.

[33] Ludwig, S., De Ruyter, K., Friedman, M., Brüggen, E. C., Wetzels, M., & Pfann, G. (2013). More than words: The influence of affective content and linguistic style matches in online reviews on conversion rates. Journal of Marketing, 77(1), 87-103.

[34] Luo, X., Dong, L., Dou, Y., Zhang, N., Ren, J., Li, Y., . . . Yao, S. (2017). Analysis on spatial-temporal features of taxis' emissions from big data informed travel patterns: a case of Shanghai, China. Journal of Cleaner Production, 142, 926-935.

[35] M. Vadivukarassi, N. P. a. P. A. (2017). Sentimental Analysis of Tweets Using Naive Bayes Algorithm. World Applied Sciences Journal, 54-59. doi: 10.5829/idosi.wasj.2017.54.59

[36] Murphy, K. P. (2012). Machine learning: a probabilistic perspective: MIT press.

[37] Nguyen, T. H., Shirai, K., & Velcin, J. (2015). Sentiment analysis on social media for stock movement prediction. Expert Systems with Applications, 42(24), 9603-9611.

[38] Nie, S., Das, A., Gabrilovich, E., Lu, W.-L., Mazniker, B., & Schilling, C. (2016). STEPS: Predicting place attributes via spatio-temporal analysis. arXiv preprint arXiv:1610.07090.

[39] Ouyang, F., Zheng, B., & Jiang, X. (2014). Spatial and temporal structures of four financial markets in Greater China. Physica A: Statistical Mechanics and its Applications, 402, 236-244.

[40] Pagolu, V. S., Challa, K. N. R., Panda, G., & Majhi, B. (2016). Sentiment analysis of Twitter data for predicting stock market movements. arXiv preprint arXiv:1610.09225.

[41] Pak, A., & Paroubek, P. (2010). Twitter as a Corpus for Sentiment Analysis and Opinion Mining. Paper presented at the LREc.

[42] Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and trends in information retrieval, 2(1-2), 1-135.

[43] Prusa, J. D., Khoshgoftaar, T. M., & Dittman, D. J. (2015). Impact of Feature Selection Techniques for Tweet Sentiment Classification. Paper presented at the FLAIRS Conference.

[44] Ruth, M., & Hannon, B. (2012). Positive Feedback in the Economy Modeling Dynamic Economic Systems (pp. 71-76): Springer.

[45] Schumaker, R. P., Zhang, Y., Huang, C.-N., & Chen, H. (2012). Evaluating sentiment in financial news articles. Decision Support Systems, 53(3), 458-464.

[46] Shah, R. R., Yu, Y., Verma, A., Tang, S., Shaikh, A. D., & Zimmermann, R. (2016). Leveraging multimodal information for event summarization and concept-level sentiment analysis. Knowledge-Based Systems, 108, 102-109.

[47] Shimada, K., Inoue, S., Maeda, H., & Endo, T. (2011). Analyzing tourism information on Twitter for a local city. Paper presented at the Software and Network Engineering (SSNE), 2011 First ACIS International Symposium on.

[48] Şimşek, M. U., & Özdemir, S. (2012). Analysis of the relation between Turkish twitter messages and stock market index. Paper presented at the Application of Information and Communication Technologies (AICT), 2012 6th International Conference on.

[49] Skuza, M., & Romanowski, A. (2015). Sentiment analysis of Twitter data within big data distributed environment for stock prediction. Paper presented at the Computer Science and Information Systems (FedCSIS), 2015 Federated Conference on.

[50] Smedt, T. D., & Daelemans, W. (2012). Pattern for python. Journal of Machine Learning Research, 13(Jun), 2063-2067.

[51] Song, J., Kim, K. T., Lee, B., Kim, S., & Youn, H. Y. (2017). A novel classification approach based on Naïve Bayes for Twitter sentiment analysis. KSII Transactions on Internet and Information Systems (TIIS), 11(6), 2996-3011.

[52] Song, Z., & Xia, J. C. (2016). Spatial and Temporal Sentiment Analysis of Twitter data. European Handbook of Crowdsourced Geographic Information, 205.

[53] Tam, P. S. (2014). A spatial–temporal analysis of East Asian equity market linkages. Journal of Comparative Economics, 42(2), 304-327.

[54] Tang, B., He, H., Baggenstoss, P. M., & Kay, S. (2016). A Bayesian classification approach using class-specific features for text categorization. IEEE Transactions on Knowledge and Data Engineering, 28(6), 1602-1606.

[55] Wang, H., Can, D., Kazemzadeh, A., Bar, F., & Narayanan, S. (2012). A system for real-time Twitter sentiment analysis of 2012 us presidential election cycle. Paper presented at the Proceedings of the ACL 2012 System Demonstrations.

[56] Wang, X., Gerber, M. S., & Brown, D. E. (2012). Automatic crime prediction using events extracted from twitter posts. Paper presented at the International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction.

[57] Wied, D. (2013). CUSUM-type testing for changing parameters in a spatial autoregressive model for stock returns. Journal of Time Series Analysis, 34(2), 221-229.

[58] Yu, L.-C., Wu, J.-L., Chang, P.-C., & Chu, H.-S. (2013). Using a contextual entropy model to expand emotion words and their intensity for the sentiment classification of stock market news. Knowledge-Based Systems, 41, 89-97.

[59] Zhang, L. (2013). Sentiment analysis on Twitter with stock price and significant keyword correlation.