European Research Studies Journal
Volume XXIV, Issue 4B, 2021

pp. 56-65

Financial Sentiment on Twitter's Community and it's Connection to Polish Stock Market Movements in Context of Behavior Modelling*

Submitted 21/09/21, 1st revision 11/10/21, 2nd revision 29/10/21, accepted 25/11/21

Eryka Probierz[1], Adam Gałuszka[2], Katarzyna Klimczak[3], Karol Jędrasiak[4], Tomasz Wiśniewski[5], Tomasz Dzida[6]

Abstract:

Purpose: *The paper analyzes the relationship between tweets published on social media related to the stock market and stock market movements. The research conducted aims to find patterns that may account for behavioral modelling patterns between social media and the stock market.*
Design/Methodology/Approach: *To conduct the research, 452,776 tweets containing a hashtag with the name of at least one company included in the WIG20 of the Polish stock market were analyzed. The data obtained also included information about the reach of the tweet and the popularity of the author. The analyzed text also contained a timestamp allowing the tweets to be linked to the behavior of the stock market. Additionally, the analysis implemented developed Polish financial sentiment analysis vocabularies allowing for better pattern retrieval.*
Findings: *The obtained results indicate that among the analyzed data, three patterns related to the sentiment of statements accompanying a given company in social media can be distinguished. A consistent sentiment present in many statements with a wide range is decisive. When different sentiments are present, the presence of a pattern is not identifiable. The implementation of Polish financial dictionaries allowed to distinguish also individual words characterizing positive and negative sentiment.*
Practical Implications: *The results obtained can form the basis for further, more in-depth analyses between sentiment published in social media and stock market movements.*
Originality value: *Studies analyzing the relationship between the analysis of tweets and stock markets are conducted but analyses related to the Polish stock market are not a popular field of analysis. Due to the unique nature of each stock exchange, it is indicated to be innovative and to be able to emerge some patterns, useful in the context of investing.*

****An earlier version of this article has been presented in the Digital ICABE, www.icabe.gr*
Keywords: *Sentiment analysis, stock market, twitter, social media, behavior modelling.*

---

[1]*Corresponding author, Silesian University of Technology, and University of Silesia, eryka.probierz@polsl.pl;*
[2]*Silesian University of Technology, adam.galuszka@posl.pl;*
[3]*SGH Warsaw School of Economics, katarzyna.klimczak@sgh.waw.pl;*
[4]*WSB Universities, k.jedrasiak@wsb.edu.pl;*
[5]*Warsaw Stock Exchange, tomasz.wisniewski@gpw.pl;*
[6]*Warsaw Stock Exchange, tomasz.dzida@gpw.pl;*

JEL codes: *G11, G17.*
Paper Type: *Research Paper.*

Acknowledgements: *The work of Adam Galuszka was supported by the SUT grant No 02/060/RGP20/0019. The work of Eryka Probierz was supported in part by the European Union through the European Social Fund as a scholarship under Grant POWR.03.02.00-00-I029, and in part by the Silesian University of Technology (SUT) through the subsidy for maintaining and developing the research potential grant in 2021 for young researchers in analysis (02/060/BKM21/0027).*

1. Introduction

The aim of our paper is to analyze tweets, which belong to the category of microblog information, from the perspective of behavioral modeling on the Polish stock market. This is an innovative approach, which assumes that changes in the indicators of companies are also influenced by social media and the general mood presented there in the context of a given company. The choice of tweets was dictated by certain aspects, which constitutes their undeniable advantage. One of them is the multiplicity of users, which allows to collect many opinions from different users, and thus the possible extension of the research sample. Due to the limited size of tweets, they usually contain simplified, condensed information, containing only the most important information. This is an attractive form of data for analysis, as it can be assumed that it does not contain as much noise as, for example, media reports in the form of press articles. The data is accessible and easily analysable, and in addition, because tweets are easy to add, they reflect users' almost immediate reactions to an issue.

In order to do so, the tweets were analyzed by means of sentiment analysis, which allowed for assigning specific categories to them. Then, for the stock market data, the closing price data was taken and the returns and trading volume were calculated. This data was then subjected to multiple regression analysis. According to the findings, there is a correlation between social media and stock market data. The compound coefficient, which includes both positive and negative reports, as well as the positive coefficient, were shown to be significant in the sentiment study. The examined stock market data had no correlation with neutral sentiment. The number of tweets examined was shown to be connected to the number of transactions on a given day, particularly for non-public firms. The findings show that the subject is important, particularly from the standpoint of individual investors.

2. Literature Review

The conducted research is a part of a series of analyses undertaken by previous researchers. It should be noted that the analyzed articles can be divided according to the sub-theme of the issue. The most represented group is the analysis of news and media reports. Tetlock (2007) analysed the content of the Wall Street Journal's "Abreast of the Market" column over the period 1984-1999 and constructed a measure of media pessimism. He estimated the intertemporal links between the measure of media pessimism and the stock market. He found that: a) high levels of media pessimism robustly predict downward pressure on market prices, followed by a reversion to fundamentals; b) unusually high or low values of media pessimism forecast high market trading volume; c) low market returns lead to high media pessimism.

The analysis performed by Tetlock was extended by Tetlock *et al.* (2008) by addressing the impact of negative words in all Wall Street Journal (WSJ) and Dow Jones News Service (DJNS) stories about individual S&P 500 firms from 1980 to 2004. They found that negative words in the financial press forecast low firm earnings. They also showed that negative words in stories about fundamentals are particularly useful predictors of earnings and returns. Another article (Ferguson *et al.*, 2015) investigates the link between news media content and stockmarket activity using information from daily firm-specific newspaper articles. Data sample consists of 264,647 firm-specific UK news media articles from The Financial Times, The Times, The Guardian, and Mirror covering FTSE 100 firms over the period 1981 to 2010.

Empirical test results show significant predictive power of firm-specific media content for stock returns. Positive as well as negative words in news stories convey valuable information about future returns. Similar analysis was pursued by Johnman *et al.* (2018), and Schumaker *et al.* (2012), where authors investigated the statistical and economic effect of positive and negative sentiment on daily excess returns and volatility in the FTSE 100 index, using business news articles published by the Guardian Media Group (The Guardian, The Observer and The Guardian Weekly) between 01/01/ 2000 and 01/06/2016. The analysis indicates that while business news sentiment derived from articles aimed at retail traders does not influence excess returns in the FTSE 100 index, it does affect volatility, with negative sentiment increasing volatility and positive sentiment reducing it.

Another subcategory is financial report analysis. Loughran and McDonald (2011) analyzed SEC's 10-K reports of more than 8 000 companies during the period from 1994 to 2008 and found that word lists developed for other disciplines (for example psychology) are not suitable to financial text and they may misclassify common words in financial language. They developed alternative word lists that better reflect tone in financial text that can be used to explore market reactions around the 10-K filing date. Feuerriegel and Prendinger (2016) used regulated announcements of companies were used from January 2004 to June 2011 and data from the CDAX financial index to develop negotiation strategies that use textual news for decision-making based on new information entering the market. Their study also presented a proposal to a newsbased investment decision support system.

Antweiler *et al.* (2004) examined 1.5 million messages posted on two forums for 45 companies in the Dow Jones Industrial Average (DJIA) and the Dow Jones Internet

Index and used naive Bayes text classification to obtain sentiments of messages. They demonstrated a strong positive correlation between the messages of the online boards, trading volume, and trading volatility and a minor correlation between message board posts and price activity on the next day. Das and Chen used message boards of 24 tech-sector stocks present in the Morgan Stanley High-Tech Index (145,110 messages). They developed a methodology for extracting small investor sentiment from stock message boards, and analyzed the performance of various algorithms. The algorithms were used to generate a sentiment index. Empirical applications evidence a relationship with stock values—tech-sector postings are related to stock index levels, and to volumes and volatility.

Tweets and microblogs are an equally popular category. Sprenger *et al.* (2014) studied 249,533 English-language, stock-related microblogging messages related to S&P 100 companies between 1 January and 30 June 2010. They find an association between market features (return, trading volume, volatility) and the corresponding tweet features (message sentiment, message volume, the level of agreement among postings). McGurk *et al.* (2020) collected 3,941,149 unique tweets over 296 days discussing 4,972 unique stocks to estimate investor sentiment index. They found strong empirical evidence that our estimated investor sentiment indexes are related to abnormal returns. Nofer and Hinz (2015) measured the sentiment in Twitter posts. The data source consisted of German Tweets from January 1, 2011, to March 17, 2012. The study showed that the correlation between tweets and the financial market exists when considering the number of Twitter followers in each publication. This topic has also been addressed in the context of predicting and forecasting stock market in the context of using microblogs (Oliveira *et al.*, 2017; Bollen *et al.*, 2011; Arias *et al.*, 2013).

A separate subcategory is earnings press releases. Among these, research by Davis *et al.* (2012) examined whether a measure of the language used in an earnings press release is related to future firm performance and generates market reaction. They analyzed a sample of approximately 23,000 quarterly earnings press releases published on PR Newswire between 1998 and 2003 and construct a measure of net optimistic language. They found that levels of net optimistic language in earnings press releases are predictive of firm performance in future quarters. They also found a significant market response to unexpected net optimistic language. Huang *et al.* (2014) examined earnings press releases (firm quarterly observations from 1997–2007, with average annual sample of 1300 firms) to test whether verbal tone management used in these documents influences investors' assessments about the value of the company. They found that investors are misinformed by tone management. That abnormal positive tone contains negative information about future firm fundamentals, that firms tend to engage in tone management particularly when incentives to manipulate investor perceptions are high, and that investors are misinformed by tone management.

Also noteworthy is the addressing of conference calls. Viana *et al.* (2019) analyzed the potential association between the linguistic sentiment of the quarterly earnings conference calls and the stock returns of Brazilian listed companies. A sample of 78 observations related to the conference calls of 24 companies listed in the B3 – Brasil Bolsa Balcao of the Novo Market segment, during the period 2016-2017 was used. They employed statistical tests, such as descriptive results, correlation matrix, and correspondence analysis. Evidence of a positive association between the optimism sentiment present in the conference calls, specifically related to the questions and answers sections, and abnormal company returns was found. This topic was also addressed by Price *et al.* (2012) and Blau *et al.* (2015). In total, more than 20 articles were analyzed identifying both common elements and differences in the studies.

3. Research Methodology:

3.1 Data Overview

Data was downloaded for the 20 largest companies on the Polish stock exchange, referred to as WIG20, based on the statement of 01.09.2021. Twenty companies included: ASSECOPOL, SANPL, CCC, KGHM, LOTOS, LPP, CDPROJEKT, PEKAO, PGNIG, PKNORLEN, PKOBP, ORANGEPL, CYFRPLSAT, PGE, PZU, TAURONPE, JSW, MERCATOR, DINOPL, ALLEGRO. Selected companies were analyzed due to higher activity associated with a given company on Twitter compared to smaller companies. Data was taken for each company from the time frame of 01.03.2021 to 01.09.2021, which covered 6 months. It is important to note that the data from twitter is downloaded and published around the clock regardless of the day of the week, while the data from the stock market only covers weekdays when the stock market is active. Figure 1 shows the number of tweets downloaded for each company, as well as information about the company's share of the WIG20. The tweets were retrieved using the Twitter REST API, and the reference was names and abbreviations for the 20 WIG20 companies preceded by a "#" or "$" sign. This yielded 452776 tweets in total.

3.2 Stock Market Data

Stock market company data for the analyzed time period was obtained courtesy of the Warsaw Stock Exchange. They allowed us to analyze the data in terms of closing price, returns and trading volume. Closing price is the price per share of a given company at the close of the stock exchange on a given day. Returns are calculated as the logarithm difference from base e between the closing price of a share on a given day and the closing price on the previous day (equation 1):

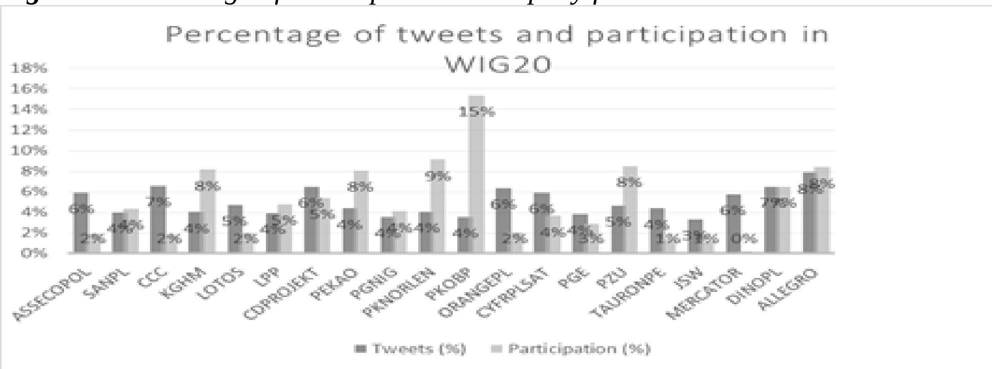
$$R_d = \{InClose_d - InClose_{(d-1)}\}x\ 100 \qquad (1)$$

Trading volume was calculated as the logarithm of the total number of shares. Daily volatility was adjusted based on intra-day highs and lows (equation 2).

$$\sigma = \sqrt{\frac{1}{n}\sum \frac{1}{2}\left[ln\frac{H_t}{L_t}\right]^2 - [2ln2 - 1]\left[ln\frac{C_t}{O_t}\right]^2} \tag{2}$$

Based on the analysis conducted for the stock market data, 3 variables were obtained i.e., closing price, returns and trading volume.

Figure 1. *Percentage of tweets per each company from WIG20*



Source: *Own study.*

3.3 Sentiment Analysis

In order to perform sentiment analysis, Vader Sentiment Analysis was used (Hutto *et al.*, 2014). It is a Valence Aware Dictionary and Sentiment Reasoner. It is based on a lexical solution combined with text analysis principles and was created to analyze data from social media. It provides 4 indicators, positive sentiment, negative sentiment, neutral sentiment, and a compound value that ranges from -1 to 1, where -1 indicates extremely negative text sentiment and +1 extremely positive text sentiment. The solution used during text analysis takes into account both punctuation elements, the use of capital letters and the gradation of adjectives. This allows for a thorough text analysis that also takes into account emoticons contained in tweets.

4. Research Results and Discussion

Analyses were conducted using Statistica 13.1 software, and data were prepared using Python. Multiple regression analysis was conducted for each variable derived from the stock market data. The results obtained were divided according to the variable under study. The first variable analyzed is the final price. The next is returns and the last is trading volume. For each analysis there is a table with R^2 for each company in relation to four variables obtained through sentiment analysis. For each analysis the Relative Absolute Error (RAE) is also provided.

In order to analyze the closing price, multiple regression analysis was conducted, the results of which are presented in Table 1. Based on the results obtained, it should be noted that for a number of companies no significant relationship was obtained between the sentiment of tweets and closing price, similarly for the following analysis returns were companies in which the state owns the majority of shares. Among the highest results it should be noted R^2=0.31 for CCC company for the compund indicator, for the same indicator the result R^2=0.37 for CDPROJEKT company. A significant result was also obtained for ALLEGRO company, which was R^2=0.44 for the positive sentiment of tweets in relation to the closing price.

Table 1. *R^2 values for closing price for each company for Vader components*

Components	**ASSECO -POL**	**SANPL**	**CCC**	**KGHM**	**LOTOS**	**LPP**	**CD-PROJEKT**	**PEKAO**	**PGNIG**	**PKN-ORLEN**
Positive	0.12	0.09	0.28	0.06	**0.05**	0.08	0.22	0.04	0.04	0.07
Negative	0.06	**0.11**	0.16	**0.09**	0.02	**0.1**	0.35	0.06	**0.07**	**0.08**
Neutral	0.02	0.04	0.02	0.05	**0.05**	0.02	0.03	**0.07**	0.02	0.03
Compound	**0.13**	0.1	**0.31**	**0.09**	0.02	**0.1**	**0.37**	0.06	**0.07**	**0.08**
best R^2	0.13	0.11	0.31	0.09	0.05	0.1	0.37	0.06	0.07	0.08
RAE	0.92	0.94	0.84	0.94	0.94	0.92	0.81	0.93	0.93	0.94
	PKOBP	**ORANGEPL**	**CYFR-PLSAT**	**PGE**	**PZU**	**TAURONPE**	**JSW**	**MERCATOR**	**DINOPL**	**ALLEGRO**
Positive	0.04	0.21	0.14	0.04	0.03	0.05	0.02	**0.13**	0.15	**0.44**
Negative	**0.08**	0.18	0.19	0.08	0.07	**0.05**	**0.11**	0.12	0.17	0.27
Neutral	0.02	0.01	0.03	0.04	**0.08**	0.05	0.02	0.02	0.06	0.03
Compound	0.07	**0.23**	**0.21**	**0.08**	0.07	**0.05**	0.09	**0.13**	**0.19**	0.38
best R^2	0.08	0.23	0.21	0.08	0.08	0.05	0.11	0.13	0.19	0.44
RAE	0.93	0.85	0.86	0.92	0.94	0.95	0.92	0.91	0.87	0.81

Source: *Own study.*

In order to analyze the returns, multiple regression analysis was conducted, the results of which are presented in Table 2. Based on the data contained therein, it can be noted that for many companies, no statistically significant relationship was shown between the indicators obtained from the sentiment analysis and daily returns. Particularly noteworthy is the fact that the companies that showed no relationship between tweet sentiment and daily returns were mostly state-owned companies. Particularly noteworthy are the companies that obtained the highest R^2 indications, i.e. CCC company obtained the value of R^2=0.37 for compound indicator, similar value was obtained by CDPROJEKT company, R^2=0.36, and the highest value was obtained by ALLEGRO company, R^2=0.41. It means that both positive and negative reports included in the analyzed tweets had a significant impact on daily returns for these companies.

The final element of the analysis is to conduct multiple regression for trading volume. The results obtained are presented in Table 3. For the trading volume analysis, no sentiment analysis indicators were used. Tweets volume was assigned to each trading volume value so as to examine the combination of these two variables. Due to the varying time of publication, which allows tweets to be published regardless of the day of the week or time of day, the data was adjusted to match the rules of the stock market. That is, tweets from Saturday and Sunday were excluded from the analysis, and tweets up to 4 p.m. were included in the same day for trading volume, while tweets after 4 p.m. were analyzed in the context of trading volume for the following day. Tweets from Friday from 4 pm to midnight. Monday tweets from midnight onwards were included in Monday's trading volume. The obtained results indicate significant relationships between the volume of tweets and trading volume for selected companies. The highest R2 values were obtained for the companies: ALLEGRO, CD-PROJEKT and CCC.

Table 2. *R^2 values for returns for each company for Vader components*

Components	**ASSECO -POL**	**SANPL**	**CCC**	**KGHM**	**LOTOS**	**LPP**	**CD-PROJEKT**	**PEKAO**	**PGNIG**	**PKN-ORLEN**
Positive	0.09	0.13	0.32	**0.05**	**0.07**	0.09	0.27	0.04	0.06	0.02
Negative	**0.14**	0.11	0.12	0.04	0.02	**0.11**	0.33	**0.11**	**0.07**	**0.04**
Neutral	0.02	0.03	0.01	0.02	0.02	0.01	0.03	0.05	0.04	0.02
Compound	0.13	**0.14**	**0.37**	**0.05**	0.05	**0.11**	**0.36**	0.1	**0.07**	**0.04**
best R^2	0.14	0.14	0.37	0.05	0.07	0.11	0.36	0.11	0.07	0.04
RAE	0.93	0.92	0.81	0.94	0.94	0.91	0.82	0.93	0.94	0.95
	PKOBP	**ORANGEPL**	**CYFR-PLSAT**	**PGE**	**PZU**	**TAURONPE**	**JSW**	**MERCATOR**	**DINOPL**	**ALLEGRO**
Positive	0.03	0.14	0.18	0.06	0.08	0.05	0.02	0.09	0.12	0.37
Negative	**0.08**	0.12	0.13	0.11	0.03	**0.06**	**0.08**	0.08	**0.14**	0.21
Neutral	0.03	0.01	0.02	0.02	0.05	0.04	0.02	0.02	0.03	0.01
Compound	**0.08**	**0.15**	**0.21**	**0.12**	**0.09**	**0.06**	0.07	**0.1**	**0.14**	**0.41**
best R^2	0.08	0.15	0.21	0.12	0.09	0.06	0.08	0.1	0.14	0.41
RAE	0.93	0.91	0.89	0.92	0.93	0.95	0.94	0.92	0.91	0.81

Source: *Own study.*

Table 3. *R^2 values for trading volume and tweets volume*

Volume	**ASSECO -POL**	**SANPL**	**CCC**	**KGHM**	**LOTOS**	**LPP**	**CD-PROJEKT**	**PEKAO**	**PGNIG**	**PKN-ORLEN**
R^2	0.22	0.26	0.38	0.18	0.14	0.19	0.37	0.12	0.17	0.2
RAE	0.88	0.86	0.81	0.9	0.92	0.9	0.79	0.92	0.89	0.9
	PKOBP	**ORANGEPL**	**CYFR-PLSAT**	**PGE**	**PZU**	**TAURONPE**	**JSW**	**MERCATOR**	**DINOPL**	**ALLEGRO**
R^2	0.25	0.29	0.3	0.18	0.19	0.22	0.17	0.27	0.22	0.41
RAE	0.88	0.83	0.82	0.89	0.89	0.89	0.9	0.84	0.9	0.74

Source: *Own study.*

A number of analyses indicate that there is a correlation between social media activity, analyzed in terms of sentiment and volume of tweets, and the value of the final price, returns and trading volume. It should be noted, however, that these relationships are extremely different for different companies, and the analyzed relationships were almost non-existent in companies owned by the state.

5. Conclusions, Proposals, Recommendations:

Sentiment analysis was conducted for the developed set of tweets, which was then combined with data from the Warsaw Stock Exchange. The analysis shows that there is a regression relation between social media and stock market data. Based on the sentiment analysis, the compound coefficient, which combines both positive and negative reports, as well as the positive coefficient proved to be significant. Neutral sentiment showed no relationship with the analyzed stock market data. The volume of tweets analysed proved to be related to the volume of transactions on a given day, especially for non-stock companies. The analyses conducted indicate the relevance of the topic, especially from the perspective of individual investors. The inclusion of social media analysis is likely to increase the effectiveness of conducted transactions and, consequently, increase traders' profits. Further research is planned to deepen the analysis and expand the dataset.

References:

Antweiler, W., Frank, M.Z. 2004. Is all that talk just noise? The information content of internet stock message boards. The Journal of Finance, 59, 1259-1294.
Arias, M., Arratia, A., Xuriguera, R. 2013. Forecasting with twitter data. ACM Transactions on Intelligent Systems and Technology, 5(1), 1-24. doi:10.1145/2542182.
Blau, B.M., DeLisle, J.R., Price, S.M. 2015. Do sophisticated investors interpret earnings conference call tone differently than investors at large? Evidence from short sales. Journal of Corporate Finance, 31(1), 203-219.
Bollen, J., Mao, H., Zeng, X. 2011. Twitter mood predicts the stock market. Journal of Computational Science, 2, 1-8.
Das, S.R., Chen, M.Y. 2007. Yahoo! for Amazon: sentiment Extraction from small talk on the web. Management Science, 53, 1375-1388.
Davies, A.K., Piger, J.M., Sedor, L.M. 2012. Beyond the Numbers: Measuring the Information Content of Earnings Press Release Language. Contemporary Accounting Research, 29(3), 845-868.
Ferguson, N.J., Philip, D., Lam, T.Y.H, Guo, M.J. 2015. Media content and stock returns: the predictive power of press. Multinational Finance Journal, 19, 1-31.
Feuerriegel, S., Prendinger, H. 2016. News-based trading strategies. Decision Support Systems, 90, 65-74. doi:10.1016/j.dss.2016.06.020.
Heston, S.L., Sinha, N.R. 2017. News vs. Sentiment: Predicting Stock Returns from News Stories. Financial Analysts Journal, 73(3), 67-83. A Publication of CFA Institute.
Huang, X., Teoh, S.H, Zhang, Y. 2014. Tone Management. The Accounting Review, 89(3),1083-1113.
Hutto, C., Gilbert, E. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In Proceedings of the International AAAI Conference on Web and Social Media, 8(1).
Johnman, M., Vanstone, B.J., Gepp, A. 2018. Predicting FTSE 100 returns and volatility using sentiment analysis. Accounting & Finance, 58, 253-274.
Loughran, T., McDonald, B. 2011. When Is A Liability Not A Liability? Textual Analysis, Dictionaries. And 10-Ks, 35-65.
McGurk, Z., Nowak, A., Hall, C.J. 2020. Stock returns and investor sentiment: textual analysis and social media. Journal of Economics and Finance, 44, 458-485.
Nofer, M., Hinz, O. 2015. Using Twitter to predict the stock market. Business & Information Systems Engineering, 57(4), 229-242. doi:10.1007/s12599-015-0390-4.
Oliveira, N., Cortez, P., Areal, N. 2017. The impact of microblogging data for stock market prediction: Using twitter to predict returns, volatility, trading volume and survey sentiment indices. Expert Systems with Applications, 73, 125-144.
Price, S.M., Doran, J.S., Peterson, D.R., Bliss, B.A. 2012. Earnings conference calls and stock returns: The incremental informativeness of textual tone. Journal of Banking & Finance, 36(4), 992-1011.
Schumaker, R.P., Zhang, Y., Huang, N.C., Chen, H. 2012. Evaluating sentiment in financial news articles. Decision Support Systems, 53(3), 458-464. doi:10.1016/j.dss.2012.03.001.
Sprenger, T.O., Tumasjan, A., Sander, P.G., Welp, I.M. 2014. Tweets and trades: The

information content of stock microblogs. European Financial Management, 20(5), 926-957.
Tetlock, P.C. 2007. Giving content to investor sentiment: the role of media in the stock market. The Journal of Finance, 62, 1139-1168.
Tetlock, P.C., Saar-Tsechansky, M., Macskassy, S. 2008. More than words: quantifying language to measure firms' fundamentals. The Journal of Finance, 63, 1437-1467.
Viana Junior, D.B.C., Castro, L.A., Ponte, V.M.R., Lima, M.C. 2019. Do the Words Matter? An Analysis of the Linguistic Sentiment of Earnings Conference Calls and Abnormal Stock Returns in Brazilian Companies. Advances in Scientific and Applied Accounting, 12(3), 144-158.
Yan, D., Zhou, G., Zhao, X., Tian, Y., Yang, F. 2016. Predicting stock using microblog moods. China Communications, 13(8), 244-257. doi:10.1109/CC.2016.7563727.