

Analyzing Microblogging Activity and Stock Market Behavior through Artificial Neural Networks

Piñeiro-Chousa Juan ¹, López-Cabarcos M Ángeles ¹, Pérez-Pico Ada M ¹, Vizcaíno-González Marcos ^{2,*}

¹ Business Administration and Management School, Universidade de Santiago de Compostela,
Avda. Alfonso X El Sabio s/n., Lugo E27002, Spain

² Financial Economics and Accounting, Facultad de Economía y Empresa, Universidade da Coruña,
Campus de Elviña s/n, A Coruña, E15071, Spain

* Corresponding author: marcos.vizcaino@udc.es; Tel.: +34-981-563100

Submitted: 27 August 2019, accepted: 25 March 2020, published: 28 March 2020

Abstract: This paper attempts to analyze the relationship between social network activity (message sentiment) and stock market (trading volume and risk premium). We used Artificial Neural Networks to analyze 87,511 stock-related microblogging messages related to S&P500 Index posted between October 2009 and October 2014. The results obtained suggest that there is a direct relationship between trading volume and negative sentiment, and between risk premium and negative sentiment. The paper concludes with several directions for future research.

Keywords: microblog sentiment; S&P500; multi-layer perceptron neural network; Stanford CoreNLP Natural Language Processing; financial markets

How to cite: Piñeiro-Chousa Juan, López-Cabarcos M Ángeles, Pérez-Pico Ada M, Vizcaíno-González Marcos. Analyzing Microblogging Activity and Stock Market Behavior through Artificial Neural Networks. *J. Bus. Account. Financ. Perspect.*, 2020, 2(2): 10; doi:[10.35995/jbafp2020010](https://doi.org/10.35995/jbafp2020010).

© 2020 Copyright by the authors. Licensed as an open access article using a [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) license.



Introduction

The recent advances in technology have made it possible to access more information almost instantaneously. In a world where information has a great importance, as in the case of the financial world and especially in financial markets, these advances have reached a great relevance. When we talk about information, it is necessary to talk about social media, and particularly about social networks. The number of social network users is constantly growing. The amount of information shared through social networks is also increasing, and financial markets are not indifferent to this phenomenon. Furthermore, when the new language analysis software appeared, it

opened new avenues for researchers. The sentiment analysis of sentences allows measuring what people think and relating it to different events. Along this line, several studies analyzed the relationship between social network sentiment and the stock market (Oh and Sheng, 2011; Rao and Srivastava, 2012; Bissattini and Christodoulou, 2013; Oliveira et al., 2013), and found that sentiment can help in predicting stock market variables such as trading volume, returns, or market movements. Nevertheless, in order to obtain a better prediction, it is necessary to take many more variables into account, such as market risk (Ribeiro-Soriano and Urbano, 2010), or Tobin's Q (Piñeiro-Chousa et al., 2016), as well as other information from social networks (e.g., message volume, certain user profile characteristics) (Piñeiro-Chousa et al., 2017). Still, we observe a gap in this area, since most of the papers published about social networks and the stock market analyze the influence that social network sentiment has in the stock market. However, there are few studies about the inverse relationship—that is, how stock market variables influence social network sentiment (Piñeiro-Chousa et al., 2016).

In this study we analyzed the relationship between social network sentiment and the financial markets, focusing on the influence of stock market variables on social network sentiment. For this purpose, we considered trading volume and other financial variables like a measure of market risk through the risk premium, calculated as the difference between the price of BAA Yields and AAA Yields. Additionally, we included message-related variables, like the number of tweets and user experience. Among all the data-mining techniques, we decided to use an artificial neural network (ANN) for the analysis since it is more adaptive to real-world situations and does not suffer from statistical model constraints (Aktan, 2011).

Literature Review

Information plays a relevant role in financial markets. At present, there is an enormous amount of readily accessible information that investors can benefit from to make their investment strategies. At the same time, researchers can use this information to improve the quality of their studies about market behavior. In this sense, it is necessary to talk about an increasingly relevant source of information and research data: social media. In recent years, advances in technology facilitated social media to gain popularity in society through blogs, microblogging sites, forums, and social networks. Investors use these sources of information to make their decisions whilst researchers use them to analyze financial market activity (Wysocki, 1998; Antweiler and Frank, 2004). Thus, data from blogs like RagingBull.com (Tumarkin and Whitelaw, 2001), from news sites like Yahoo! Finance (Antweiler and Frank, 2004) or LiveJournal (Gilbert and Karahalios, 2010), or from social networks like Twitter (Asur and Huberman, 2010; Sprenger et al., 2014) or StockTwits (Piñeiro-Chousa et al., 2017) can help to explain stock market activity, and even to improve stock market predictions (Oh and Sheng, 2011). Among these new studies, there are a group of papers that analyze the sentiment of messages posted in social networks. For example, Asur and Huberman (2010) extracted sentiment from Twitter with the aim of predicting market behavior. Sprenger et al. (2014) showed that the sentiment extracted from tweets was associated with the stock abnormal performance. Zhang et al. (2011) analyzed the mood derived from Twitter posts to predict the Dow Jones, S&P 500, and NASDAQ performance. Bollen et al. (2011) improved the accuracy of DJIA predictions using public mood dimensions extracted from Twitter. Furthermore, it has been shown that social network sentiment can help in predicting stock market movements (Oh and Sheng, 2011; Makrehchi et al., 2013; Bissattini and Christodoulou, 2013).

Most of the papers are focused on the effect that social media sentiment has on stock markets, but not many studies analyze market influences on social networks. In this sense, some papers have shown that trading volume relates with the volume of posted messages (Tumarkin and Whitelaw, 2001; Sprenger et al., 2014). Regarding the influence of stock market variables in social networks, it has been shown that certain variables such as Tobin's Q, the P/E ratio, or market capitalization influence social network sentiment (Piñeiro-Chousa et al., 2016). For this reason, it seems reasonable to suppose that other market variables could also have influence. For this purpose, we propose the following hypotheses:

Hypothesis 1: *Risk premium and trading volume influence social network sentiment.*

It has been shown that social network variables such as experience or the number of followers influence the stock market (Sprengrer et al., 2014), and are related to social network sentiment (Piñeiro-Chousa et al., 2016). For example, an experienced investor would react differently to a situation of increased risk compared to an inexperienced investor. In this sense, we propose the following hypothesis:

Hypothesis 2: *Users' experience and number of followers influence social network sentiment.*

Methodology

Neural networks consist of a layered, feed-forward, completely connected network of artificial nodes, which can be used for classification or estimation (Larose, 2006). Neural networks can have different uses, including descriptive and predictive data mining (Alan et al., 2014). Artificial neural networks (ANN) are a machine learning tool based on computational models inspired by biological neural networks (i.e., the human central nervous system). The multi-layer perceptron (MLP) is the most extended ANN algorithm for financial prediction. The MLP is composed of three layers: (a) the input layer, containing the predictors (i.e., attributes); (b) the hidden layer, containing the unobservable nodes; and (c) the output layer, containing the responses (Aktan, 2011). The most frequently used algorithm for learning MLP is the back-propagation algorithm (BPA), which uses a gradient descent that can find local minimum (Aktan, 2011).

According to Alan et al. (2014), an MLP has three characteristics: (a) the model of each node in the network usually includes a nonlinear activation function, either sigmoidal or hyperbolic; (b) the network includes one or more layers with hidden nodes that do not belong to the network input or output. They allow the network to learn complex and nonlinear tasks through the progressive extraction of meaningful characteristics from the patterns of the input; and (c) the network shows a high degree of connectivity from one layer to the next.

In the financial context, ANNs have been used for bankruptcy (Odom and Sharda, 1990) and bank failure prediction (Tam and Kiang, 1992), to estimate the future financial health of firms (Coats and Fant, 1993), to predict bankruptcy filing (Boritz and Kennedy, 1995), or to examine financial distress (Etherige and Sriram, 1997; Hu and Ansell, 2006).

We used the WEKA (Waikato Environment for Knowledge Analysis) program, version 3.7, developed in Waikato University. WEKA is open-source software that supports many categorizing, piling, and coupling rules algorithms.

Data and Variables

We chose StockTwits.com as our source for social network data, since it is only used by the financial community. As in Twitter, StockTwits.com users can label stock-related messages by a dollar sign followed by the related ticker symbol (e.g., \$SPX), making it easy to collect the information. We collected 87,511 stock-related microblogging messages holding the ticker symbol of S&P500 Index during the analyzed period (i.e., from October 2009 to October 2014). Our study focused on the S&P500 Index and on the companies mentioned in S&P500 Index messages, representing a total of 490.

We analyzed the sentiment of each message posted through the sentiment analysis software Stanford CoreNLP Natural Language Processing Toolkit, developed by Manning et al. (2014). We obtained a sentiment value for each message. This value fluctuates between -2 (which refers to a very negative sentiment) and 2 (which means a very

positive sentiment). A neutral sentiment has a value of 0. Then, we aggregated the data on a monthly basis, and calculated the average sentiment for each month and for each company as follows:

$$sent_{it} = \frac{\sum_{j=0}^m S_{jt} * m_{jt}}{\sum_{j=0}^m m_{jt}},$$

where S_{jt} is the sentiment j about ticker i in moment t , and m_{jt} is the number of messages with sentiment j in moment t .

We turned the sentiment variable into a dummy variable, being 0 for negative values and 1 for positive and neutral (equal to 0) values.

We collected financial data in daily intervals for the S&P500 Index from Nasdaq.com and Morningstar.com, which allows Microsoft Excel® files to be downloaded with daily data of the companies that belong to the index. We also downloaded the monthly price of AAA and BAA Yields from the Federal Reserve web to calculate the risk premium.

The class attribute was the sentiment. This variable was nominal, with values of 0 and 1.

The variables for the ANN were trading volume (vol), user experience (exp), message volume (not), and risk premium (premrisk).

The variable *vol* is calculated as the amount of the trading volume for each company and for each month. We only used the trading volume data of days when the company was mentioned in messages.

$$vol_{it} = \sum_{j=0}^m V_{jt},$$

where V_{jt} is the trading volume j about company i in moment t , $j = \{1, 2, \dots, m\}$.

The variable *exp* is the average experience of users that mention a company in a given month. Experience is rated in a scale from 1 (novice) to 3 (professional). This classification is given by the users when they fill their profile in StockTwits.com.

The variable *not* is the count of the total number of microblog messages about a ticker (company) for each month.

The variable *premrisk* is the difference between the price of BAA Yields (considered as bad yields) and AAA Yields (considered as good yields):

$$premrisk = Yield(BAA) - Yield(AAA).$$

Additionally, we considered time variables, specifically annual (year) and quarterly (trim) variables.

Our final dataset was composed of 5281 instances and 7 attributes (sentiment, year, trim, vol, exp, not, and premrisk).

Results

We applied a multilayer perceptron (MLP) classifier, which uses a back-propagation algorithm, for classification. Results are shown in Table 1. The Kappa statistic was 0, which means that there was no correlation between attributes. The fit is remarkable, with a true positives rate over 95% and precision over 90%.

Correctly Classified Instances	5089
Kappa statistic	0
Mean absolute error	0.0494
Root mean squared error	0.1835
TP Rate	96.40%
Precision	92.9%

Table 1 ANN results.

Figure 1 shows the neural network, where the output layers were negative and positive sentiment. Table 2 shows that one node (node 4) clearly stood out above the others, with a weight of 2.79. The second node with higher weight was node 3 (1.73). This figure shows that node 4 was the most prominent, since its weight was approximately 61% higher than the weight of the second. The weight of node 4 was positive, which means that this node had a positive influence in node 0, which represents the negative sentiment.

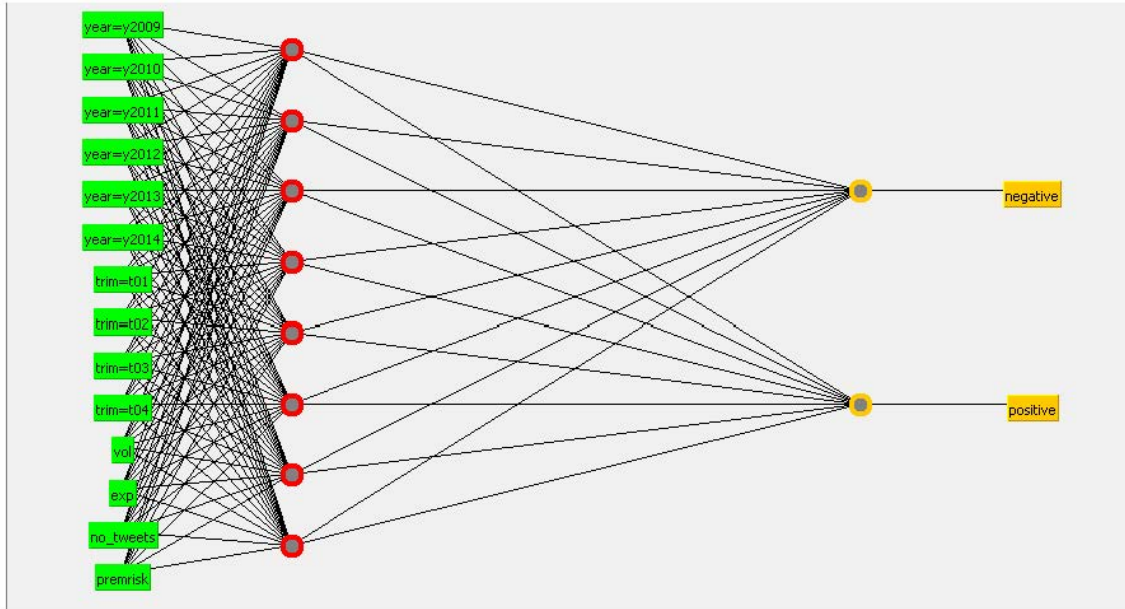


Figure 1 Artificial Neural Network output.

Sigmoid Node 0	
Inputs	Weights
Threshold	-0.25199827
Node 2	1.38692511
Node 3	-1.73919065
Node 4	2.79783564
Node 5	1.41229572
Node 6	1.48754123
Node 7	1.55641441
Node 8	-0.32991562
Node 9	0.49017774

Table 2 Node 0 weights.

The weights of attributes in node 4 are shown in Table 3. When trading volume increased, negative sentiment increased as well. This could relate to an increase of trading volume leading to an increase of share sales, due to a panic situation in the stock market. Additionally, when risk premium increased, negative sentiment increased as well. These results confirm Hypothesis 1. Additionally, when a post was written by a professional user, negative sentiment arose, partially confirming Hypothesis 2. This may relate to professional users posting more messages when things go wrong.

Sigmoid Node 4		
Inputs	Threshold	Weights
	Threshold	0.10366072
Attrib	year=y2009	-0.38729648
Attrib	year=y2010	3.44298097
Attrib	year=y2011	-1.09049899
Attrib	year=y2012	-0.33667994
Attrib	year=y2013	-3.24590775
Attrib	year=y2014	1.16582292
Attrib	trim=t01	1.34554736
Attrib	trim=t02	0.55148878
Attrib	trim=t03	-1.83961813
Attrib	trim=t04	-0.16398952
Attrib	vol	3.11105185
Attrib	exp	2.76440289
Attrib	no_tweets	1.47244708
Attrib	premrisk	3.12510297

Table 3 Node 4 weights.

Regarding time variables, years 2010 and 2013 had the highest weights. In 2010, the positive weight means that negative sentiment increased due to the maintenance of the financial crisis that began in 2008. On the contrary, in 2013 the weight was negative, which could relate to either the crisis recovery in the USA, or to the sovereign debt crisis in Europe, which caused international investors to move investments from Europe to the USA, creating a substitution effect.

Conclusions

In this paper, we analyzed the relationship between market variables (i.e., trading volume and risk premium) and social network variables (i.e., sentiment, user experience, and message volume). The results suggest that social media can reflect what happens in the market. For example, if there is an increase in market risk premium, it could lead to a higher negative sentiment. Several papers have proven that social media data can be used to predict market movements (Oh and Sheng, 2011; Makrehchi et al., 2013; Bissattini and Christodoulou, 2013). Therefore, we proved that the market activity also influenced the social media activity, so the relationship can be understood in both directions. Regarding our results, investors and companies must pay attention to social networks, since a message about a company posted in Twitter or StockTwits could cause an increase or a decrease in its stock price. Thus, our analysis of social media activity and investment strategies could be the starting point for further studies in this area.

References

- Aktan, S. (2011). Application of machine learning algorithms for business failure prediction. *Investment Management and Financial Innovations*, 8(2), 52–65.
- Alan, M. A., Yeşilyurt, C., Aydin, S., & Aydin, E. (2014). Comparison of classification performance of selected algorithms using rural development investments support programme data. *Kafkas Üniversitesi Veteriner Fakültesi Dergisi*, 20(3), 351–356.

- 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(3), 1259–1294. [[CrossRef](#)]
- Asur, S., & Huberman, B. A. (2010). Predicting the future with social media. Paper presented at the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), Toronto, AB, Canada, August 31–September 3; Volume 1, pp. 492–499.
- Bissattini, C., & Christodoulou, K. (2013, July 4). Web sentiment analysis for revealing public opinions, trends and making good financial decisions. *Trends and Making Good Financial Decisions*. [[CrossRef](#)]
- Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), 1–8. [[CrossRef](#)]
- Boritz, J. E., & Kennedy, D. B. (1995). Effectiveness of neural network types for prediction of business failure. *Expert Systems with Applications*, 9(4), 503–512. [[CrossRef](#)]
- Coats, P. K., & Fant, L. F. (1993). Recognizing financial distress patterns using neural network tool. *Financial Management*, 22(3), 142–155. [[CrossRef](#)]
- Etherige, H. L., & Sriram, R. S. (1997). A Comparison of the relative costs of financial distress models: Artificial neural networks, logit and multivariate discriminant analysis. *Intelligent Systems in Accounting, Finance and Management*, 6(3), 235–248. [[CrossRef](#)]
- Gilbert, E., & Karahalios, K. (2010). Widespread worry and the stock market. Paper presented at the International Conference on Weblogs and Social (ICWSM), Washington, DC, USA, May 23–26; pp. 59–65.
- Hu, Y.-C., & Ansell, J. (2006). *Developing financial distress prediction models: A study of US, Europe and Japan retail performance*. Scotland: Management School and Economics, University of Edinburgh.
- Larose, D. T. (2006). *Data Mining Methods and Models*. Hoboken: John Wiley & Sons, Inc.
- Makrehchi, M., Shah, S., & Liao, W. (2013). Stock Prediction Using Event-based Sentiment Analysis. Paper presented at the IEEE/ACM International Conference on Web Intelligence, Atlanta, GA, USA, November 17–20.
- Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., & McClosky, D. (2014). The stanford CoreNLP natural language processing toolkit. Paper presented at the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, Baltimore, MD, USA, June 22–27; pp. 55–60.
- Odom, M. D., & Sharda, R. (1990). A Neural Network Model for Bankruptcy Prediction. Paper presented at the International Joint Conference on Neural Network, San Diego, CA, USA, June 17–21; Volume 2, pp. 163–168.
- Oh, C., & Sheng, O. (2011). *Investigating predictive power of stock micro blog sentiment in forecasting future stock price directional movement*. New Orleans: ICIS, Association for Information Systems. ISBN 978-0-615-55907-0.
- Oliveira, N., Cortez, P., & Areal, N. (2013). On the predictability of stock market behavior using stocktwits sentiment and posting volume. In *Progress in Artificial Intelligence* (pp. 355–365). Berlin/Heidelberg: Springer.
- Piñeiro-Chousa, J. R., López-Cabarcos, M. Á., & Pérez-Pico, A. M. (2016). Examining the influence of stock market variables on microblogging sentiment. *Journal of Business Research*, 69(6), 2087–2092.
- Piñeiro-Chousa, J., Vizcaíno-González, M., & Pérez-Pico, A. M. (2017). Influence of social media over the stock market. *Psychology & Marketing*, 34(1), 101–108.
- Rao, T., & Srivastava, S. (2012). Tweetsmart: Hedging in markets through twitter. Paper presented at the 2012 Third International Conference On Emerging Applications of Information Technology (EAIT), Kolkata, India, November 29–December 1; pp. 193–196.

- Ribeiro-Soriano, D., & Urbano, D. (2010). Employee-organization relationship in collective entrepreneurship: An overview. *Journal of Organizational Change Management*, 23(4), 349–359. [[CrossRef](#)]
- Sprenger, T. O., Tumasjan, A., Sandner, P. G., & Welpe, I. M. (2014). Tweets and trades: The information content of stock microblogs. *European Financial Management*, 20(5), 926–957. [[CrossRef](#)]
- Tam, K. Y., & Kiang, M. Y. (1992). Managerial Applications of Neural Networks: The Case of Bank Failure Predictions. *Management Science*, 38(7), 926–947. [[CrossRef](#)]
- Tumarkin, R., & Whitelaw, R. F. (2001). News or noise? Internet postings and stock prices. *Financial Analysts Journal*, 57(3), 41–51. [[CrossRef](#)]
- Wysocki, P. D. (1998). *Cheap talk on the web: The determinants of postings on stock message boards* (Working Paper No. 98025). Ann Arbor, MI, USA: University of Michigan Business School.
- Zhang, X., Fuehres, H., & Gloor, P. A. (2011). Predicting stock market indicators through twitter “I hope it is not as bad as I fear”. *Procedia-Social and Behavioral Sciences*, 26, 55–62. [[CrossRef](#)]