CLASSIFICATION OF FINANCIAL MARKETS INFLUENCERS ON TWITTER

Inês Almeida[1] and André Sabino[1,2]
[1]*Autónoma Techlab, Universidade Autónoma de Lisboa, Lisboa, Portugal*
[2]*IADE – Faculdade de Design, Tecnologia e Comunicação, Universidade Europeia, Lisboa, Portugal*

ABSTRACT

Information published by social network users can have an impact on the (virtual and physical) community, which supports the concept of influencer, which is a user generally focused on a set of interests (such as fashion, health, well-being, beauty, finances, etc.), whose opinion correlates with the behavior observed in a community of users who share those interests. The same concept of social network influencer may also be explored in the financial markets, considering user profiles whose expressed opinions align with the performance of a company's shares.
Several users interested in the stock exchange use Twitter as a platform to express their opinions on stocks. Considering that tweets express either a positive, negative, or neutral feeling about a stock, we present a model to describe the profile of an influencer, as well as analyze the correlation between their tweets and the performance of a stock.
Relying on a genetic algorithm to best describe an influencer's profile and selecting the best fit from a set of features, this study provides a comparison between sentiment analysis techniques, applied on social network content, produced by potential influencers. We present a study of several sentiment analyses approaches, namely using Support Vector Machines, Neural Networks, Naïve Bayes, and K-Nearest Neighbors. Results show that the method with the best performance is the Neural Network. This outcome is used as a parameter for analyzing an influencer's tweets.
Validation shows that, by integrating a genetic algorithm for influencer detection with a sentiment analysis technique based on Neuronal Networks, it is possible to identify users with social network content sentiment score correlated with variations in a share price, thus representing an information signal suitable to be used by prediction models.

KEYWORDS

Machine Learning, Sentiment Analysis, Financial Market Analysis

1. INTRODUCTION

Financial markets investment management involves constant assessment and decision making in a fast-pacing environment. Aside from official market data, there are several non-primary sources of information that may prove helpful when designing predictive models. Expert opinion is one of such sources, and it may be expressed through social networks, such as Twitter.

We propose to leverage user generated content on social media as a signal for financial markets behavior, by identifying correlations between the sentiment of a particular user's contributions about a company on the network (i.e., posts about the company) and the performance of the stock of that company.

We focus our study on Twitter, which enables its users to share information and opinions about almost anything (restricted by its terms of service). Posted messages are called tweets, and these may include references to other users (using the @ character, in the form of @UserName), eventually engaging them in conversation. Users may also forward a tweet from another user (i.e., a retweet), and include tags (using the # character, i.e., hashtags) in the tweet content (Drus and Khalid, 2019). Twitter also supports a special type of tag for stock market trading tickers, called cashtag, which prefixes the trading symbol with the $ character, such as $AAPL[1] for the Apple stock[2]. Each tweet has a maximum character limit, which currently is 280 characters, from 140 until 2017 (Gligorić et al., 2018).

[1] https://twitter.com/search?q=%24aapl
[2] https://www.nasdaq.com/market-activity/stocks/aapl

We define influence on a social network, such as Twitter, as the potential of a user's actions to cause actions from other users.

With the evolution of social networks, the concept of influencers is increasingly addressed. Leavitt, et al., focused on the analysis of influencer profiles on Twitter, concluding that the number of followers is not directly related to the number of retweets performed, i.e., there may be accounts with few followers, who's tweets have a high number of retweets. The study also found that an independent user may be more influential than a popularly known entity (Leavitt et al., 2009).

In another study, M. Cha, et al., analyzed the influence of Twitter users based on three measures: the number of followers; the number of retweets; and the number of mentions. The goal was to verify a user's ability to generate dialogues and actions (Cha et al., 2010). Aligning with the results from Leavitt, et al. (Leavitt et al., 2009), the study concludes that a large number of followers do not correlate with a high number of tweets. It was also found that influence is not acquired accidentally and usually requires a strategic effort from the user (e.g., by approaching a single topic per tweet). Furthermore, the authors conclude that users may be influencers over various topics (Cha et al., 2010).

E. Lahuerta-Otero and R. Cordero-Gutiérrez used a data mining tool, that combines graph theory with social influence theory, to investigate influencers on Twitter and discover common features between influencers. The study focused on 3 853 users who posted tweets about the automotive companies Toyota and Nissan. The authors concluded that influencers do not usually embed links in their posts, which usually have a low number of words and a high number of hashtags and mentions. It was also concluded that influencers have a high number of followers who themselves express their opinions. This research, instead of detecting influencers, analyzed the common properties of the tweets of users already identified by the community as influencers, aiming to characterize how influencers communicate with their users (Lahuerta-Otero and Cordero-Gutiérrez, 2016).

X.Zhang, H.Fuehres and P.Gloor forecasted financial market indicators, such as the Dow Jones Industrial Average (DJIA), NASDAQ and S&P 500 indexes using Twitter posts, which were collected for six months and labeled as "Positive" or "Negative" based on a selection of mood words (e.g., "fear" or "hope"). From these annotations, the authors computed a daily collective sentiment based on the number of occurrences of each mood word and report that when users express fear and concern in their tweets, indices such as DJIA, NASDAQ and S&P 500 register loss of value. The study also reports that users post more tweets when they feel positive, and when there is some uncertainty in the stock market, there is a tendency to express feelings such as fear or hope (Zhang et al., 2011).

Bollen, et al., studied the correlation between the collective sentiment measured from Twitter and behavior of the DJIA. The study filtered posts that matched a set of expressions the authors considered to state mood states, and characterized those posts according to 7 mood dimensions from two assessment tools: OpinionFinder (Wilson et al., 2005) and their own Google-Profile of Mood States (GPOMS). The study reported that at least one of those dimensions is able to forecast (i.e., reports a significate Granger causality) the DJIA index, namely, the "Calm" dimension measured by GPOMS. There was a large increase in prediction ability, reaching a success rate of 86.7%. The authors conclude that sentiment analysis of tweets enables automatic, fast, free and large-scale forecasting tools for financial market indices (Bollen et al., 2011).

Xu, et al., also studied potential correlations between sentiment analysis of tweets and the returns of the main global stock markets, based on the Multifractal Detrended Cross-Correlation Analysis (MF-DCCA) method). The study used the Daily Happiness Index[3] as proxy for the collective mood, and the analysis was carried out through the division between developed and underdeveloped countries, and report significant correlations between financial markets returns and tweets from a few developed countries (excluding South Korea and the United Kingdom, where no correlation was found), and if all developing countries considered. This correlation relationship is consistent over time in developed countries and fluctuates over time for developing countries. To a degree, these results confirm a dependence between sentiment and global financial markets (Xu et al., 2021).

These results support the hypothesis that there is a signal available on social networks which may be suitable to forecast financial markets. All works share a processing step where tweets are translated into a sentiment score, which we use to annotate tweets from a set of user profiles that are potential influencers.

The classification of social network profiles whose posts correlate with financial markets behavior has been studied by several authors. D. M. Limpo et al. presents a system for the identification and classification of users

[3] The index is built with data https://hedonometer.org/

whose content may correlate with a particular stock of the S&P 500 Financial Market Index. Most of the influencers found were entities/companies in the same sector of the stock in our study. The methodology yields better results than the recommendation service (Who-To-Follow) available at Twitter (Limpo, David M., 2016).

In a related effort, D. Sousa presents an opinion mining system for the S&P 500 Market Index stocks, which identifies the main characteristics of financial market influencers. The system uses a Naïve Bayes binary sentiment analysis algorithm (which classified the tweets as “positive” or “negative”), reporting accuracies of 72.16% and 75.45% over two distinct datasets, and a Genetic Algorithm to obtain optimal values for each parameter (i.e., characteristics such as account time, number of tweets, or number of followers) that defines an influencer (Sousa, 2017).

The goal of these studies was to propose, for each characteristic, the range of values suitable for an influencer. For example, if the "number of followers" characteristic has a range of values between 500 to 1000, then a user defined as influencer should have their number of followers within that range. Both studies concluded that influencers have a long-time account, are very active on Twitter in terms of published tweets and tend to use tweets to talk to other users.

J. Sacramento implemented a system for collecting, processing, and classifying tweets about a particular company to measure its impact on stock prices. Tweets from Facebook and American Airlines were analyzed. For Facebook, significative correlations of 0.39 in terms of stock volume and 0.42 for stock close price was found. For American Airlines, only the correlation in terms of stock volume (0.51) is significant (Sacramento, 2021).

2. METHODOLOGY

The system is composed of three modules, namely, data collection and processing, sentiment analysis, and influencer detection.

The data collection from Twitter includes tweet content and number of followers, among other useful information, which are preprocessed to filter out content that is not relevant for sentiment analysis. Financial data is collected from Yahoo Finance, namely, the date and adjusted closing price (see section 2.2).

All modules are developed in Python, mainly because it is supported by a rich ecosystem for machine learning.

2.1 Twitter Data Collection

Twitter information is collected with the snscrape[4] module, which enables filtering through Twitter’s public API[5].

Posts about a company are filtered through the cashtag. The module also enables date filters, such as “until:2021-12-06 since:2021-11-24”, which constrains the capture to a date interval, and user filters, such as “from TheTommRobinson”, which constrains the capture to the posts of a given user. The module is only limited by the public Twitter API access policy and can retrieve all tweets from a given user since the creation of their account.

In this study, all collected tweets are in english and must contain a cashtag. Tweets with emojis are excluded to minimize errors in the sentiment analysis module.

Tweets are preprocessed in the extent that the following content is removed: Uniform Resource Locator (URLs); words with numbers; the cashtags; special characters, i.e., non-digits or letters.

Cashtags are removed because there is no value for sentiment analysis. Hashtags, however, are not deleted, since they potentially represent emotions, such as #happy. Note that the # symbol, being a special character, is removed. The same occurs with the $ character on cashtags. Both hashtags and cashtags are marked before the removal of the special characters, to avoid elimination by other criteria. Special characters and words with numbers are removed mainly because they are closely related with automatic spam tweets (Limpo, David M., 2016).

[4] https://pypi.org/project/snscrape/
[5] https://developer.twitter.com/en/docs/twitter-api

Finally, tweets are converted into lower case due to the application of the Naïve Bayes algorithm, which is case sensitive. Although it is not the only algorithm to be tested in this study, it is also evaluated.

2.2 Financial Data Collection and Stock Returns Calculation

For a particular historical record, the value of each stock is represented by the closing price and the adjusted closing price. The first represents the last gross price traded before the market closes, while the second represents the closing price of excluding the impact corporate actions that may affect it, namely, stock splits and dividends.

For example, on a dividend ex-date (i.e., the first day the stock is traded without the right to collect a particular dividend payment), the stock opening price discounts the dividend value which, and the end of the day, is accounted in the closing price. If the day trading session did not move the price of the stock, the closing price still records a loss in the same amount as the dividend. The adjusted closing price will discount the dividend value, more closely reflecting the effects of the actual trading session.

We use the yfinance[6] module to collected financial data from Yahoo Finance and analyze the adjusted closing price within a time interval of 1 month. The daily return of a stock is determined by

$$Return_i(Stock, d_i) = \frac{ClosingPrice_{adjusted}(Stock, d_i) - ClosingPrice_{adjusted}(Stock, d_{i-1})}{ClosingPrice_{adjusted}(Stock, d_i)} \qquad (1)$$

where $Stock$ is the share under study, $\mathrm{ClosingPrice}_{\mathrm{adjusted}}$ is the adjusted closing price of a stock in a particular date, d_i the date of the calculation and d_{i-1} the previous date. The return of a share for a given day is obtained by the difference between the adjusted closing price values that day and the previous day, divided by the adjusted closing price of the previous day.

2.3 Sentiment Analysis

The sentiment analysis module combines combining machine learning with lexicon-based methods.

Most machine learning methods are supervised, which requires a training set, e.g., a set of Twitter posts labeled as “Positive”, “Neutral” or “Negative”. We use a lexicon-based approach to classify this training set.

The VADER (Valence Aware Dictionary for sEntiment Reasoning) Python module works from a set of previously annotated words (hence, lexicon-based) and provides a sentiment prediction for other texts and words.

According C.J. Hutto and E. Gilbert, VADER’s performance stands out from comparable methods for text sentiment analysis on social networks. The VADER project reports correlation coefficient values (r = 0.881) as high as the values generated through human evaluation (r = 0.888). As for classification accuracy, the value produced considering the manual classification (F_1 = 0.84) is lower than the classification accuracy of VADER (F_1 = 0.96) (Hutto and Gilbert, 2014).

In our study, we calculate the sentiment of each tweet (“Positive”, “Neutral” or “Negative”), as well as the number of positive and negative words using VADER with the following encoding: if a word’s polarity, as determined by the VADER module, is equal to or greater than 0.05, the word is annotated as “Positive”; if the polarity is less than or equal to -0.05, it is annotated as “Negative”; if none of these conditions are met, the word is considered as “Neutral”.

With an annotated training set, we consider the following machine learning methods: support vector machines (SVM); neural networks (NN); naïve bayes; k-nearest neighbors (kNN) (see Figure 1).

[6] https://pypi.org/project/yfinance/

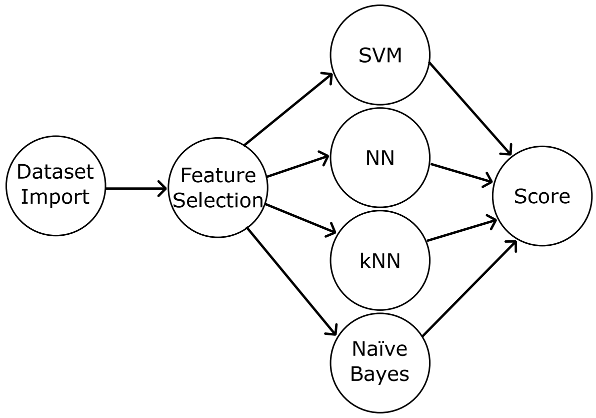

Figure 1. The workflow for sentiment analysis method comparison

SVM and NN rely on training sets of labeled data, which are used to build approximations of functions, optimized to label similar data points, i.e., given a set of features and a target, these methods work as non-linear function approximators.

An SVM builds the approximation function by maximizing separations between the training data points in their representation on expanded geometrical spaces (i.e., hyperplanes). The resulting model is defined by the actual functions.

In a NN, between the input layer (input features) and the output layer, there may be one or more non-linear layers, called hidden layers. Each neuron in the hidden layer transforms the values from the previous layer with a weighted linear sum, followed by a non-linear activation function. As a result, the output layer takes the values from the last hidden layer and transforms them into output values. The resulting model is defined by the set of training data points and their impact on the several neuron layers.

Naïve bayes and kNN are used as simpler base comparison models for SVM and NN.

2.4 Genetic Algorithm

We propose a genetic algorithm to determine which characteristics best define an influencer. This approach tries to select the best solution (i.e., a set of values for the collection of features used to describe an influencer) for a problem from a large set of candidates, through iterations where small changes are introduced into previous solution candidates. Ideally, these candidates evolve into an optimal solution.

The process iterates through generations of solutions, each based on combinations and modifications of the previous one. In each iteration, the most suitable solutions (or, *parents*) are selected according to a *fitness function score*.

Pairs of parents are combined to produce one or more *children* which contain a new set of values for each feature (i.e., *gene*). This combination is actually a crossover operation over the parents' genes, which is then submitted to a chance-based mutation process (Coley, 1999).

The algorithm terminates after a predetermined number of iterations, or when a *satisfactory* solution (see the fitness function, at equation 2) is found.

In the context of the genetic algorithm, an individual represents a set of users. For each feature, users are sorted in descending order according to the values they have for that feature, and each set is divided into several subsets (i.e., ranks). As an example, we consider the "number of followers" feature, where the first subset of users is composed of users with the most followers, and the last subset with the lowest number of followers. This process is repeated for the remaining characteristics under analysis. It is expected that the more users that are in each rank, the greater the range of values. On the other hand, high number of ranks implies that, for each rank, there is a small number of users, thus reducing the range of values.

In this study we focus on a total of 800 users, and consider 20 ranks, each composed of 40 users. Each rank is described by 5 genes (i.e., a *chromosome*), one for each characteristic to be analyzed.

The fitness function represented by equation (2) calculates the impact of tweets made on a given date for each user on a rank. This impact is calculated through the product of the tweet's polarity and the sum of the ACP gain during the 10 days preceding the tweet's publication. The goal is to find positive values, which means that the sentiment of the user's tweets was in line with the variation of the stock. Otherwise, if the value is negative, the user's prediction about the stock was not in line with its variation (Limpo, David M., 2016).

$$f(date) = \sum_{i=1}^{n} (\sum_{d=1}^{10} Return(Stock, tweet_i^{date} + d)) \times tweet_i^{polarity} \quad (2)$$

We set up a tournament selection method to choose between competing ranks is used for this process. n ranks are randomly selected for each tournament, in which the winner yields the best performance in the fitness function. In the study, we consider tournaments of three individuals, amounting to 10 tournaments overall, with 10 winning individuals. Winning ranks move on to the crossover phase.

The Tournament's outcome is a set of winning ranks designated as *parents*. In the crossover phase, *children* (which are also ranks) of each pair of randomly selected parents are formed by the junction of the genes of their parents. This junction if defined by two randomly selected cut-off points (double point crossover), yielding two children for each pair of parents.

In the first child, the genes between the cutoffs are copied from the first parent, with the remaining genes copied from the second parent. In the second child, the inverse occurs, that is, the genes between the cut-off points are copied from the second parent.

After the crossover process, the new population consists of parents and children, over which we apply a mutation process over the smallest number of genes in an individual, yielding a $\frac{1}{numOfGenes}$ rate (Limpo, David M., 2016), e.g., a 20% mutation rate for a set of 5 genes. This rate criterium is arbitrary, and a candidate for further studies.

3. VALIDATION

Our validation process was focused on Apple stock, which is traded with the AAPL ticker (cashtag $AAPL on Twitter). We collected tweets containing the $AAPL cashtag and respective users over 1 month, from January 12, 2022, to February 12, 2022.

We built a data collection script using the Python snscrape module. Along with the text of the tweet, the username and the date of publication of the tweet were also collected. We discarded all tweets with URLs, words with numbers, and/or emojis. This pre-processing helps with the elimination of content that does not contribute to sentiment analysis. Text was converted to lowercase. The final set of tweets has 13443 unique tweets, and 2668 unique usernames.

The study requires information about users, namely, the number of followers (followers), the total number of published tweets, the number of public lists the user is a member of, the ratio between the number of followers and number of friends of the user, and the age of the account since registration. This information will be used as characteristics for the detection of influencers through genetic algorithms and was collected using the Python tweepy module.

The tweepy module has a limitation in terms of the number of queries made to the Twitter API, so for the application of the Genetic Algorithm it will only be applied to a set of collected users. Which means that the total number of users selected for the analysis was 800.

We collected historical market data for the Apple stock between January 12, 2022, and February 12, 2022, which includes the adjusted close price (ACP), enable the calculation of the daily return (or *gain*) for the time interval (represented by equation (1)). Data was collected using the Python yfinance module, which gathers information from the Yahoo Finance platform.

The training set (7943 tweets) includes tweets that were published between January 12, 2022, and January 27, 2022. The test set (5500 tweets) includes tweets published between January 28, 2022, and February 12, 2022.

For each tweet, the polarity of each word was calculated and, consequently, the calculation of number of words with "Positive", "Negative" or "Neutral" sentiment in it, which is then aggregated into the polarity of the entire text. For a polarity value greater than or equal to 0.05, the word/text is classified as "Positive", whereas if a polarity value lower than or equal to -0.05 is obtained, the word/text is classified as "Negative". Otherwise, the word/text is "Neutral". The total number of tweets labeled as "Positive" was 3926, as "Negative" was 1472 and as "Neutral" was 2545.

Precision is the fraction of relevant instances. The best value is 1 and the worst value is 0. This measure is calculated as the quotient of the number of true positives and the sum of the number of true positives and false positives.

$$Precision = \frac{truepositives}{truepositives + falsepositives} \tag{3}$$

Recall is the fraction of relevant instances that were retrieved. Like precision, the range of values is between 0 and 1. This measure is calculated as the quotient of the number of true positives and the sum of the number of true positives and false negatives.

$$recall = \frac{truepositives}{truepositives + falsenegatives} \tag{4}$$

The F1 score can be seen as the harmonic mean of precision and recall measures. Like precision and recall, F1 score takes values between 0 and 1.

$$F_1 = 2 \times \frac{precision \times recall}{precision + recall} \tag{5}$$

We evaluate the performance of the several the sentiment analysis methods using the F1 score (combination between Precision and Recall) and report the following: 0.973 for K-Nearest Neighbors; 0.982 for Support Vector Machines; 0.988 for Naïve Bayes; and 0.999 for Neural Networks. Our results (which are comparable to Hutto and Gilbert (Hutto and Gilbert, 2014)), shows that Neural Networks are the most suitable approach for classifying the tweets.

After applying the genetic algorithm to the set of defined characteristics, it was possible to obtain the most appropriate rank, that is, the range of values, for each characteristic, that best defines an influencer. The evaluation yielded that the number of followers of an influencer user is between 1 842 and 3 073 followers. As for the number of tweets, these are between 62 698 and 1 726 179. In relation to public lists, an influencer user belongs to 9, 10 or 11. For the ratio between followers and friends this assumes a value between 0.26 and 0, 37. Finally, an influencer user has had a Twitter account for 3 or 4 years.

4. CONCLUSIONS

This paper presented a financial market influencer classification system relying on a genetic algorithm and hybrid sentiment analysis techniques, with a focus on Twitter. We concluded that Neural Networks are the best suited approach for sentiment analysis.

Our results show that such influencer is a user with a high number of tweets, with several followers greater than the number of friends, and a high number of followers. The user also belongs to several public lists, and it isn't a recent Twitter user.

Although our focus was on smaller observation windows, for future work, we propose that the high F1 we reported may be due to a lower number of tweets in the analysis and suggest studying the impact of larger observation windows.

We also plan to include the "verified" status of the profile, which may be relevant to refine our model.

ACKNOWLEDGEMENT

We acknowledge the support of Autónoma TechLab for supporting this work and providing a stimulating environment to carry out this research.

REFERENCES

Bollen, J., Mao, H., Zeng, X.-J., (2011). *Twitter mood predicts the stock market*. J. Comput. Sci. 2, 1–8. https://doi.org/10.1016/j.jocs.2010.12.007

Cha, M., Haddadi, H., Benevenuto, F., P. Gummadi, K., (2010). *Measuring User Influence in Twitter: The Million Follower Fallacy*, in: Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media. Washington, D.C., USA, p. 8.

Coley, D.A., (1999). *Introduction to Genetic Algorithms for Scientists and Engineers*, Har/Dskt edition. ed. Wspc, Singapore ; River Edge, NJ.

Drus, Z., Khalid, H., (2019). *Sentiment Analysis in Social Media and Its Application: Systematic Literature Review*. Procedia Comput. Sci. 161, 707–714. https://doi.org/10.1016/j.procs.2019.11.174

Gligorić, K., Anderson, A., West, R., (2018). *How Constraints Affect Content: The Case of Twitter's Switch from 140 to 280 Characters*. Proc. Int. AAAI Conf. Web Soc. Media 12.

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. pp. 216–225.

Lahuerta-Otero, E., Cordero-Gutiérrez, R., (2016). *Looking for the perfect tweet. The use of data mining techniques to find influencers on twitter*. Comput. Hum. Behav. 64, 575–583. https://doi.org/10.1016/j.chb.2016.07.035

Leavitt, A., Burchard, E., Fisher, D., Gilbert, S., (2009). New approaches for analyzing influence on Twitter (No. 4), Web Ecology Project.

Limpo, David M., (2016). *Finding Influential Twitter Users for the Stock Market Optimized by Genetic Algorithm* (MSc Thesis). Universidade Técnica de Lisboa.

Sacramento, J., (2021). *Sentiment Analysis in the Stock Market Based on Twitter Data* (MSc Thesis). University Institute of Lisbon.

Sousa, D.J.M., (2017). *Using Naïve Bayes and Genetic Algorithms to Find Influential Twitter Users to Forecast the S&P 500* (MSc Thesis). Universidade Técnica de Lisboa.

Wilson, T., Hoffmann, P., Somasundaran, S., Kessler, J., Wiebe, J., Choi, Y., Cardie, C., Riloff, E., Patwardhan, S., (2005). *OpinionFinder: a system for subjectivity analysis, in: Proceedings of HLT/EMNLP on Interactive Demonstrations*, HLT-Demo '05. Association for Computational Linguistics, USA, pp. 34–35. https://doi.org/10.3115/1225733.1225751

Xu, K., Pang, Y., Han, J., (2021). *Dynamic Cross-Correlation between Online Sentiment and Stock Market Performance: A Global View*. Discrete Dyn. Nat. Soc. 2021, 1–11. https://doi.org/10.1155/2021/6674379

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 - Soc. Behav. Sci. 26, 55–62. https://doi.org/10.1016/j.sbspro.2011.10.562