The University of Texas at Austin
Center for Identity

# It Is an Equal Failing to Trust Everyone and to Trust Nobody: Stock Price Prediction Using Trust Filters and Enhanced User Sentiment on Twitter

*Teng-Chieh Huang*
*Razieh Nokhbeh Zaeem*
*K. Suzanne Barber*

*UT CID Report # 21-08*

November 2021

# It Is an Equal Failing to Trust Everybody, and to Trust Nobody: Stock Price Prediction Using Trust Filters and Enhanced User Sentiment on Twitter

TENG-CHIEH HUANG, RAZIEH NOKHBEH ZAEEM, and K. SUZANNE BARBER, The University of Texas at Austin, USA

Social media are providing a huge amount of information, in scales never possible before. Sentiment analysis is a powerful tool that uses social media information to predict various target domains (e.g., the stock market). However, social media information may or may not come from trustworthy users. In order to utilize this information, a very first critical problem to solve is to filter credible and trustworthy information from contaminated data, advertisements or scams. We investigate different aspects of a social media user to score his/her trustworthiness and credibility. Furthermore, we provide suggestions on how to improve trustworthiness on social media by analyzing the contribution of each trust score. We apply trust scores to filter the tweets related to the stock market as an example target domain. While social media sentiment analysis has been on the rise over the past decade, our trust filters enhance conventional sentiment analysis methods and provide more accurate prediction of the target domain, here the stock market. We argue that while it is a failing to ignore the information social media provide, effectively trusting nobody, it is an equal failing to trust everybody on social media too: Our filters seek to identify whom to trust.

CCS Concepts: • **Human-centered computing** → **Social media**; • **Information systems** → *Data mining*;

Additional Key Words and Phrases: Trust, credibility, sentiment analysis, stock market, return on investment (ROI), Twitter, social media

**ACM Reference Format:**
Teng-Chieh Huang, Razieh Nokhbeh Zaeem, and K. Suzanne Barber. 2019. It Is an Equal Failing to Trust Everybody, and to Trust Nobody: Stock Price Prediction Using Trust Filters and Enhanced User Sentiment on Twitter. *ACM Trans. Internet Technol.* 1, 1, Article 1 (January 2019), 20 pages. https://doi.org/10.1145/3338855

## 1 INTRODUCTION

Since the emergence of social media, the way people communicate with each other has thoroughly changed. Instead of one-to-one communication such as letters or phone calls, everyone can expand his or her influence to the every inch of the world by simply posting few sentences on the Internet. Twitter is one of the fastest growing social media services in the world. Users are allowed to submit up to 140 characters per post, or so called “tweets”. People can easily catch up with what is currently trending and reinforce it by “re-tweeting” the tweets they find noteworthy.

Researchers have utilized the Twitter and tweets’ sentiment with respect to a certain topic or brand for various purposes, including predicting the brand’s performance in the stock market. A wide variety of research discusses every step of Twitter sentiment analysis (automatic labeling of tweet sentiments for the purpose of training, different sentiment analysis or data mining techniques,

Authors’ address: Teng-Chieh Huang; Razieh Nokhbeh Zaeem; K. Suzanne Barber, The University of Texas at Austin, The Center for Identity, Austin, TX, 78712, USA, tengchieh@utexas.edu,razieh@identity.utexas.edu,sbarber@identity.utexas.edu.


1533-5399/2019/1-ART1 $15.00
https://doi.org/10.1145/3338855

etc.) as well as the applicability of Twitter sentiment analysis to predict various target domains (e.g., stock markets [8], politics [44], and disease outbreaks [28]).

However, Twitter is freely and publicly available and it does not employ any mechanism to separate trustworthy and credible tweets from contaminated data. Just as the presence of fake news is a concern on social media [3], researchers have identified [17, 39, 40] the credibility problem of tweets as a major concern [17] when leveraging Twitter sentiment analysis for prediction. When predicting a target domain, one cannot fully utilize the power of even the best sentiment analysis techniques, if the input data is contaminated and unreliable.

In this paper, we tackle the problem of identifying credible and trustworthy tweets. What we propose in this work is 1) complementary to previous research on social media sentiment analysis and 2) independent from the target domain. First, we introduce an extra layer of filtering to distinguish trustworthy users and their tweets before *any* data mining or sentiment extraction is performed. Second, we predict the performance of the stock market as an example target domain, but our method is readily applicable to other target domains. We focus on the stock market for several reasons: The stock market has a clear boundary on right or wrong when performing the predictions, i.e., the only examination method of the result is by looking at the price. Furthermore, the stock market is a fast changing environment that is hardly predictable, and, therefore, can benefit from Twitter sentiment analysis.

In addition to the stock price forecast, this work also provides suggestions for individuals on how to improve their trustworthiness on social media. By analyzing the distribution of trust scores, we recommend practical ways for those who want to improve their social media trustworthiness in a most effective way.

In this paper, we make the following contributions:

- We present the first work to predict *any* target domain (e.g., the stock market) with credibility filters that are based on PageRank and Hyperlink-Induced Topic Search (HITS) algorithms.
- We provide practical guidelines for social media users to improve their credibility.

The rest of the paper is organized as follows. Section 2 positions our work in the related academic work. Section 3 outlines our trust filters. Section 4 explains where we find our data source, and how we extract useful information from it. Section 5 describes how to implement trust filters for stock price prediction and compares the prediction performance with and without trust filters based on return on investment (ROI). It also provides a possible avenue for mutual feedback on trustworthiness of social network users. Finally, Section 6 concludes the findings in the experiments.

## 2 RELATED WORK

Given the sheer amount of academic research performed on social media and particularly Twitter sentiment analysis, this section seeks to highlight the differences of our work with previous research and explain the gap it fills in the related work.

### 2.1 User Credibility on Social Media

Many researchers [40] have aimed to predict the stock market or other domains with Twitter sentiment analysis. Nonetheless, in the process of data collection and prediction, the credibility of Twitter users has not received the attention it deserves. In fact, researchers have identified [17, 39, 40] the credibility problem as a major concern [17].

An incipient body of work [17, 20, 31] exists that looks at the credibility of Twitter users before basing a forecast on their tweets. For example, Castillo et al. [11] used a cascade of machine learning models (classifiers) to first find newsworthy and then identify credible tweets. In another work [10], they discovered that credible news are propagated through authors who write a large number of posts, originate at a single or few users in the network, and have many re-tweets. Since surely fake news can spread in a similar manner, they also point out that tweets which do not include URLs tend to be related to non-credible news, those with negative sentiment tend to be more credible, and those with question marks or smiling emoticons are more likely to spread non-credible information. Gupta et al. [19] found that only %17 of the dataset they considered contained credible situational awareness information. They used regression analysis to identify two sets of relevant features, namely content-based features (e.g., the number of unique characters or emoticons in a tweet) and user-based features (e.g., the number of followers or length of username). They suggested the use of those features as credibility scores. However, they did not use their credibility scores for further prediction of a domain. Canini et al. [9] performed similar credibility ranking without further application in prediction of target domains.

To the best of our knowledge, this is the first work to predict a target domain with credibility filters that are based on the social network structure (PageRank and HITS) as well as tweet content. Recently, some researchers have experimented with concepts similar to credibility ranking of Twitter users for stock price prediction. For instance, Bartov et al. [6] briefly considered the top posting users when forecasting a firm's earning announcement with Twitter.

### 2.2 Credibility vs. Trust

It is worth mentioning that "user trust"—i.e., the subjective expectation of a social network user of another [2, 42]—is different than the focus of this work which is objective user trust or credibility. Comprehensive surveys exist [4, 42] that cover conceptual differences between subjective trust and objective credibility. On one hand, many researchers have looked at the objective credibility of Twitter information, as Alrubaian et al. classify in their survey [4], at three different levels: single-post, user, and topic level. On the other hand, Sherchan and colleagues [42] dive into various definitions of subjective trust and survey trust in social networks. They consider Psychological trust (with cognitive, emotive, and behavioral aspects), sociological trust (from individual and societal view points), and even trust and its definitions in computer science (with user and system trust notions).

In this work, trust and credibility are used interchangeably and mean the objective quality of being trusted and the quality of offering true statements on social media. For the sake of compatibility with our previously published work [48], we do not abandon the use of the term "trust filter" altogether. Subjective user trust has been utilized to analyze the trust structure of social networks and to predict re-tweets [30] or following relationships [43], but not any target domains.

Furthermore, in offering practical guidelines to users on credibility, we are moving toward subjective trust and its psychological and sociological definitions. This understanding of the mutual effect of subjective and object trust, in fact, can have profound implications. For example, a through survey of the related work reveled that the research communities that study objective credibility and subjective trust are almost completely separate. The inherent relationship between objective credibility and subjective trust can bring these two lines of research together. The coverage of that related work and a more in depth discussion of this relationship is beyond the scope of this paper, but is a very promising future work avenue.

### 2.3 Application of Network-based Algorithms for Credibility on Social Media

PageRank [36] and Hyperlink-Induced Topic Search (HITS) [25] are two famous link-based ranking algorithms that are also utilized to rank social media users. An early work on investigating the network structure of Twitter [47] used PageRank for finding influential users based on the "following" relationship on Twitter. Others have used HITS [23] for finding authorities in question/answer communities. Gupta et al. [21] used a variation of the PageRank algorithm to judge Twitter user credibility, even tough, as mentioned above, they did not go further to predict any particular target domain with their set of credible users. Closest to our work is the work of Ruiz et al. [38] which, among other methods, used PageRank to correlate financial time series with stock market prices. However, several aspects of our work distinguish this research from the work of Ruiz et al., particularly the fact that their work "[did] not pretend to be a prediction model" [38] and merely measured the correlation of the behavior of Twitter with the changes in the stock market. Furthermore, their work never explored sentiment analysis.

### 2.4 Overall Review of Twitter Sentiment Analysis

For the sake of completeness, here we concisely review the history and the most prominent work on Twitter sentiment analysis. Twitter has become a popular platform for public behavior analysis because of its widespread use all over the world. Go et al. [18] were the first to perform sentiment classification on microblogs like Twitter. They were also the first to use emoticons as noisy sentiment labels for training. In their first work on Twitter sentiment analysis, they showed that machine learning algorithms (e.g., Naive Bayes, Maximum Entropy, and SVM) perform well (i.e., have accuracy above 80%) when used with emoticons as labels to identify the sentiment of the training set tweets. Kouloumpis et al. [26] took a similar supervised approach but used Twitter hashtags for the training phase in addition to emoticons. They based their work on linguistic features (such as existing lexical resources) and concluded that microblogging features of Twitter, such as the presence of intensifiers, emoticons and abbreviations, are more useful compared to conventional techniques such as Part of Speech Tagging.

Since the first works and over the past decade, researchers have utilized a wide variety of data mining and text mining techniques for Twitter sentiment analysis. For example, with SemEval workshops[1] [32], tasks of finding tweet sentiments are hugely popular and various different techniques are applied for every single step. "Sentiment Analysis in Twitter" was a notable task in SemEval-17.

### 2.5 Overall Review of Predicting Stock-Markets with Online Data Sources

Even though social media data and similar online sources of data can be used to predict many domains [7, 29] (e.g., stock markets [8], election outcomes [12, 14, 44, 46], and disease outbreaks [1, 5, 28]), predicting markets is an important topic of interest in social media as well as other text mining research [45]. Furthermore, *the method* used to predict a domain varies significantly in the above-mentioned work. We introduce four novel trust filters (Section 3) in which we present the first work to predict *any* target domain, including stock markets, with credibility filters that are based on *PageRank* and *Hyperlink-Induced Topic Search (HITS)* algorithms.

For the sake of completeness, we cover some of the most notable related work that seek to predict stock markets, our example target domain, albeit with very different methods. A widely cited work [8] demonstrated how the Twitter mood, in general, predicts the stock market closing values by high accuracy. Nassirtoussi et al. [33] tried to predict foreign exchange markets based on the text of breaking financial news headlines. Dimpfl et al. [15] studied the dynamics of stock market volatility using Internet search queries and found that high stock market volatility can follow high volumes of Internet searches. Nguyen et al. [34] performed stock market prediction based on social media analysis as well. Instead of taking all sentiments into account, they considered only the sentiments of specific topics of the company to predict stock price movement to increase the forecast accuracy. In a recent work similar to ours, Oliveira et al. [35] sought to predict stock

[1] http://alt.qcri.org/semeval2017 and http://alt.qcri.org/semeval2016/

markets through Twitter posts. They found among all the factors, sentiment and posting volume to be particularly important for the forecasting of S&P 500 index.

## 3 TRUST FILTERS

We propose trust filters to distinguish trustworthy and untrustworthy users. By leveraging trust filters to rate each user's trustworthiness, an individual's opinion reflected via his/her tweet sentiment can be weighed. Therefore, a more reliable forecast can be made. Following previous work [48], we categorize trust by four dimensions, which are Authority, Experience, Expertise and Reputation.

We previously showed [48], through the analysis of Twitter posts of 2,000 users over one month, that the quality of objective trust was–to a good extent–distilled into these four dimensions. We compared [48] our trust filters with other simple features which are widely studied in processing Twitter data and studied the importance of each feature. That study found that the most important feature was the number of related posts, and filters based on it such as Experience and Expertise. We also found how Reputation, Identity, and Proximity filters used in that work outperform many of the other features. As a result, we believe that our trust filters capture a good set of trust dimensions. Due to the information property of the stock market, here we exclude the Proximity and Identity filters from previous work [48]. We have also modified some filters to better fit this specific implementation. For instance, because only a relatively small portion of the users post stock-related tweets, the Authority filter requires more iterations and higher error tolerance to ensure convergence.

Consequently, we introduce four trust scores. Each trust score is a Real number between 0 and 1 (inclusive) that measures the trustworthiness of a user with respect to that particular trust dimension. Higher trust scores indicate higher levels of trustworthiness. The definitions of the four filters are as follows [48]. Expertise measures a user's involvement in the subject of interest. Experience is the difference between a user's Expertise and the average Expertise in the network[2]. Authority[3] is the number and quality of social media links a user receives from Hubs as an Authority. Reputation[4] is the number and quality of social media links to a user.

$$TS_{Expertise}(A) = \frac{\sum_X \#Tweets_X(A)}{\#Tweets(A)} \tag{1.1}$$

$$TS_{Experience}(A) = 1 - |TS_{Expertise}(A) - \sum_{u=1}^{U} \frac{TS_{Expertise}(u)}{U}| \tag{1.2}$$

$$TS_{Reputation}(A) = PR(A) = (1-d) + d(\frac{PR(T_1)}{Q(T_1)} + ...... + \frac{PR(T_n)}{Q(T_n)}) \tag{1.3}$$

$$TS_{Authority}(A)^{(t+1)} = a(A)^{(t+1)} = \sum_{j:j \to A} h(j)^{(t)} \tag{1.4}$$

$$h(i)^{(t+1)} = \sum_{j:i \to j} a(j)^{(t+1)} \tag{1.5}$$

where
$TS_D(A)$=Trust Score of user A generated by dimension (filter) D
$\#Tweets_X(A)$=Number of tweets posted by A related to stock X

[2]The terms Expertise and Experience are borrowed from previous work [48].
[3]The term Authority is named directly after the concept of authority in the HITS algorithm.
[4]The term Reputation—"a collective measure of trustworthiness" [22]— is commonly used in the literature (pointed out by literature surveys [22, 37]) and the use of the PageRank or similar algorithms in its calculation is well-known.

$\#Tweets(A)$=Total number of tweets posted by A
$U$=Number of users
$PR(T_i)$=PageRank of user $T_i$
$Q(T_i)$=The number of tweets that user $T_i$ re-tweets from any user. It counts a re-tweet only once, no matter how many re-tweets for the same (tweet,user) pair
$T_1 ... T_n$=Users that re-tweet the tweets of user A
$d$=The damping factor, set to the default value of 0.85
$h(j)^{(t)}$=$t_{th}$ iteration for the hub value of user j
$a(j)^{(t)}$=$t_{th}$ iteration for the Authority value of user j
$i \rightarrow j$ means user i re-tweets the tweet of user j
For the calculation of $h_i$, both times are $(t+1)$ so it implies that the hub value is formed after the Authority value.

It is important to note that, unless explicitly stated, all interactions between users (e.g., re-tweets) are by default considered in the entire network of users and tweets and not only stock-related tweets. The assumption is that if a user is trustworthy as a person, then he/she is trustworthy when it comes to tweeting about stocks. Furthermore, the graph of the stock-related tweets is sparser and provides less information compared to the entire graph.

The Reputation and Authority filters are based on the PageRank and HITS algorithms respectively [13]. The directed graph is constructed by "quotes" and "re-tweets" in the network. Once a user re-tweets another user's tweet, the former will point to the latter.

When working with a trust score, each user's sentiment is multiplied by his or her trust score as a weighting factor. The weighed sentiments are then summed up for all the users to form the final accumulated sentiment of the entire network using that trust score. The investment strategies (Section 5) use the accumulated sentiment:

$$S^t_{XD} = \sum_{i=1}^{U} TS_D(i) s^t_X(i) \tag{2}$$

Where $S^t_{XD}$ is the combined sentiment for stock X at time t using trust dimension $D$, $TS_D(i)$ is that dimension's trust score for user $i$, and $s^t_X(i)$ is the sentiment of user $i$ toward stock $X$ at time $t$.

In order to store a reasonable amount of data to analyze a user's trustworthiness but still allow for change trough the time, we re-calculate the trust scores monthly. At the beginning of each calendar month, we update all the trust scores for all the users using the previous calendar month data. In fact, we improve over previous work [38] by observing a user's credibility history. A user's credibility should be an accumulated value, even though it might fluctuate over time. One drawback of previous work [38] was that their algorithm would allow malicious users interfere with the prediction by suddenly posting huge amounts of tweets on certain stocks in a single day. Our algorithm prevents this interference because it relies more on trustworthy users based on past history (at least a month). Once the weight (trust score) is assigned, a single user can only influence the combined sentiment by $(+1/-1) \times TS_D$, no matter how many tweets he/she posts.

## 4 DATA EXTRACTION

This section explains how we obtain our experimental set of tweets and how we identify the tweets related to Standard & Poor's 500 (S&P 500) companies by looking for stock symbols in that set. Furthermore, we extract the sentiments of tweets and then combine those sentiments while constructing trust scores. Section 5 introduces investment strategies that seek to predict the movements of the stock market based on the combined sentiments. The investment strategies aim to use the prediction to maximize the Return on Investment (ROI).

Table 1. Statistics about the experimental set of tweets.

| Month | # of Tweets | # of stock-related tweets | # Users that posted about stocks |
|---|---|---|---|
| 11/2015 | 140360 | 7175 | 1792 |
| 12/2015 | 68311 | 3810 | 1651 |
| 01/2016 | 120577 | 6747 | 2722 |
| 02/2016 | 91206 | 5033 | 2002 |
| 03/2016 | 79553 | 3489 | 1591 |
| 04/2016 | 93337 | 5054 | 2533 |
| Total | 593344 | 31308 | 9324 |

### 4.1 The Twitter Dataset

A publicly available collection of tweets for the "spritzer" version is available on Internet Archive[5], which is a non-profit digital library. This dataset has been examined to be consistent with the Tweet2013 collection [41]. Currently, the datasets contain tweets collected from 2011 to 2018, an approximately 1% sample of public posts, which provides us sufficient quantity and length for research purposes. It contains all the tweets of a particular time slot in every second [24].

To better understand the effectiveness of our proposed trust filters during the financial market fluctuation, we investigate a six-month-period of tweets from November 2015 to April 2016 for training and testing purposes. By observing the S&P 500 index from 2011 to 2018, we purposefully selected this time interval, because it contains both increasing and decreasing periods of stock prices. Compared to other time periods where most of the time the stock market was trending upward (from 2011 to 2018), this time slot provides a great illustration on how trust filters function when the prices are rising as well as when they are falling. Table 1 shows statistics of this six-month Twitter dataset. Note that, in the last column, the numbers of users that post about stocks do not add up to the total number shown, since some users post about stocks in multiple months.

### 4.2 Identifying Stock-Related Tweets

We focus on the companies in the Standard & Poor's 500 (S&P 500) index, as it is representative of the whole U.S. stock market and includes most major companies that are frequently mentioned in the tweets. We scan for the $ symbol, a feature that Twitter users utilize to identify stock symbols.

### 4.3 Sentiment Analysis of Tweets

Since the main contribution of our work is the introduction and adoption of trust filters, here we use a basic sentiment analysis method. However, the use of trust filters is complementary to *any* sentiment analysis method as it improves the *input* to the sentiment analysis technique. To further illuminate how our trust filters work with other sentiment analysis methods, we leverage MeaningCloud, a commercial sentiment analysis tool in Section 5.3.

We first use a dictionary-based classification method for tweet sentiment detection similar to what Kumar et al. [27] employed. They applied a combination of corpus-based and dictionary-based methods for sentiment extraction. The dictionary-based part of their approach is very similar to ours, in the sense that they use a list of positive and negative keywords (16 adverbs and 16 verbs from WordNet).

[5]https://archive.org/details/twitterstream

Table 2. List of Frequent Stock Symbols and ETF Names in Stock-Related Tweets.

| Stock Symbol/ETF Name, Frequency | | | |
|---|---|---|---|
| spy, 1346 | dia, 169 | gild, 96 | embr, 74 |
| aapl, 939 | tsla, 166 | nke, 92 | csco, 71 |
| amp, 884 | dis, 164 | yhoo, 91 | baba, 70 |
| amzn, 366 | via, 157 | pfe, 90 | crm, 69 |
| qqq, 351 | googl, 150 | ung, 90 | celg, 67 |
| uso, 316 | fcx, 143 | nyse, 87 | sbux, 66 |
| nflx, 300 | iwm, 123 | jnj, 86 | ups, 63 |
| has, 283 | cmg, 120 | mcd, 86 | wfc, 60 |
| goog, 262 | xom, 118 | dow, 83 | etf, 60 |
| spx, 226 | intc, 117 | cvx, 82 | lnkd, 60 |
| twtr, 213 | ibm, 113 | acn, 80 | nvda, 59 |
| low, 197 | bac, 112 | fxi, 79 | vxx, 59 |
| gld, 188 | vix, 111 | flir, 77 | uvxy, 59 |
| all, 184 | slv, 106 | pcln, 76 | |
| see, 184 | nasdaq, 103 | chk, 76 | |
| msft, 180 | jpm, 102 | wmt, 74 | |

To build the list of keywords for the dictionary-based classification, we gather the most frequently used words in our sample set of tweets in the six-month period. Let $F^{300}_{all}$ be the top 300 most frequently used words in the data set of tweets over the six-month period. Also, let $F^{300}_{stock}$ be the top 300 most frequent words among the stock-related tweets in the same period. We then calculate the set of frequent words in stock-related tweets as $F^{300}_{stock} - F^{300}_{all}$. We manually review this set and identify stock symbols and ETF names (Table 2), and neutral words (Table 3). After excluding stock symbols, ETF names, and neutral words, we manually assign positive or negative sentiment to the rest of the words to be used as positive or negative keywords (Table 4). Note that the set of stock symbols we use in these experiments is beyond the frequently used stock symbols (Table 2). We include all the S&P companies too.

It is through the accumulation of these positive and negative keywords that we categorize tweets as positive, negative or neutral. If a tweet has more positive keywords than negative keywords, we consider its sentiment to be positive (+1). We detect negative sentiment (-1) in a similar manner. If the tweet has an equal number of positive and negative keywords (or no keywords at all for that matter), we consider its sentiment as neutral (0).

Figure 1 shows the decision flow of accumulated sentiment extraction. First, we identify the stock-related tweets by looking for the $ sign followed by the stock symbol, such as $AAPL and $AMZN. Second, we extract the sentiments by finding all the positive and negative keywords as explained above.

To accelerate processing speed and reduce data size, we store only the required data, i.e., stock-related tweets and their sentiments in the six-month period of interest stored by the user. Only the initial stage requires whole tweet dataset scanning. As Figure 2 depicts, the list of users includes those who have posted at least one tweet containing a stock symbol during the six months. Compared to the raw decompressed dataset ( 2.1 TB for six months), the extracted tweets only take 593.3 MB, which is only 0.028% of the original data size.

Table 3. List of Frequent Neutral Words in Stock-Related Tweets.

| Neutral Word, Frequency | | | |
|---|---|---|---|
| inc, 1697 | sales, 154 | services, 94 | incorporated, 69 |
| stock, 1067 | charts, 151 | top, 94 | closed, 69 |
| shares, 938 | chart, 150 | advisors, 93 | dividend, 68 |
| stocks, 519 | reiterated, 146 | ltd, 92 | finance, 68 |
| rating, 477 | markets, 143 | share, 91 | percent, 67 |
| price, 433 | thursday, 143 | sunday, 88 | ispytrading, 67 |
| company, 398 | vetr, 138 | analyst, 87 | daily, 66 |
| corp, 396 | december, 136 | puts, 86 | info, 65 |
| trading, 376 | zacks, 132 | holdings, 85 | weakness, 65 |
| market, 353 | strong, 130 | goldman, 85 | huge, 65 |
| management, 330 | energy, 121 | board, 83 | pick, 65 |
| target, 313 | november, 120 | change, 82 | bstrongstocks, 65 |
| investment, 312 | high, 118 | contracts, 80 | napavalley, 65 |
| earnings, 298 | asset, 118 | extreme, 80 | ihub, 65 |
| stake, 294 | opinions, 114 | wednesday, 80 | tact, 65 |
| philstockworld, 293 | analysts, 113 | executive, 79 | djuvaaoudg, 65 |
| position, 292 | days, 113 | living, 78 | ujif, 65 |
| capital, 250 | trust, 112 | feed, 77 | insider, 64 |
| oil, 247 | may, 108 | sachs, 77 | fund, 64 |
| news, 202 | ceo, 106 | pennystockgang, 77 | systems, 64 |
| february, 201 | friday, 105 | volume, 75 | ahead, 63 |
| research, 196 | investor, 104 | agreements, 74 | going, 63 |
| january, 188 | calls, 103 | japan, 73 | open, 62 |
| hold, 184 | alert, 102 | stockmarket, 72 | deutsche, 61 |
| futures, 183 | plc, 101 | weekly, 71 | quant, 61 |
| bank, 178 | ranked, 100 | eps, 71 | companies, 60 |
| apple, 177 | current, 100 | proshares, 71 | algos, 60 |
| monday, 172 | investors, 98 | jefferies, 71 | given, 59 |
| tuesday, 170 | investing, 97 | financials, 71 | resources, 59 |
| llc, 167 | pennystocks, 95 | growth, 70 | estimates, 59 |
| term, 158 | international, 95 | recent, 70 | gap, 59 |
| trade, 157 | looking, 95 | merger, 70 | credit, 58 |
| financial, 157 | china, 94 | value, 69 | |

## 5 EXPERIMENTAL RESULTS

Section 4 put in place the fundamentals of stock-related tweet sentiment extraction. Utilizing these tweets and their sentiments in our dataset, we now follow different strategies to invest in the stock market and compare Return on Investment (ROI) (Equation 3) results. Particularly, we investigate whether the use of trust filters in sentiment extraction affects the ROI of an investment strategy.

$$\text{ROI} = \frac{\text{Gain from Investment} - \text{Cost of Investment}}{\text{Cost of Investment}} \tag{3}$$

Table 4. List of Frequent Sentiment Words in Stock-Related Tweets.

| Positive Word, Frequency | |
|---|---|
| buy, 1419 | (buy, buys, buying 1089+159+171=1419) |
| long, 360 | |
| bullish, 321 | |
| raise, 264 | (raised, raises 166+98=264) |
| bought, 238 | |
| increase, 234 | (increased, increases, increase 121+71+42=234) |
| boost, 150 | (boosts, boost 92+58=150) |
| up, 63 | |
| bull, 52 | |
| **Negative Word, Frequency** | |
| sell, 828 | (sells, sell 458+370=828) |
| short, 549 | |
| down, 498 | |
| lower, 298 | (lower, lowered 141+157=298) |
| sold, 261 | |
| decrease, 91 | (decreased, decreases, decrease 45+40+6=91) |
| bearish,82 | |
| bear, 36 | |

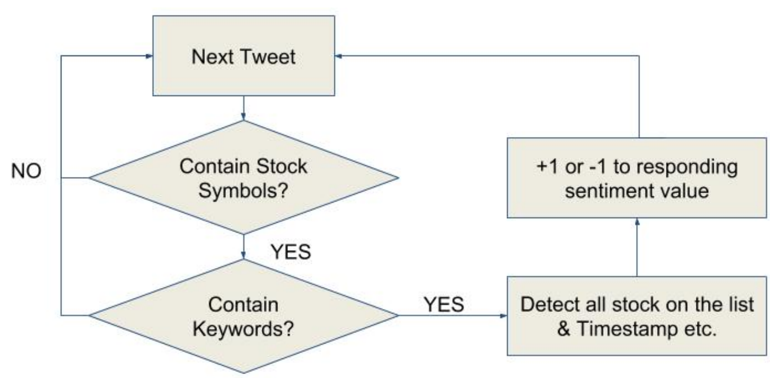


Fig. 1. Decision Flow for Accumulated Sentiment Extraction.

### 5.1 Investment Strategies

We examine various buying and selling strategies based on the general Twitter sentiment toward stocks. On any given day in the six-months period of November 2015 to April 2016, the investor following each of the strategies would buy and sell stock shares. The investment strategies we consider are as follows.

(1) The **Buy & Hold** strategy constitutes buying all listed stocks for a fixed amount each (here $100 per stock), then keeping all the investment until the end of the six-month period. The idea is to provide a strategy which can follow the movement of the S&P 500 index and is easy to implement. Although tracking S&P 500 index [16] is sophisticated and tracking error is

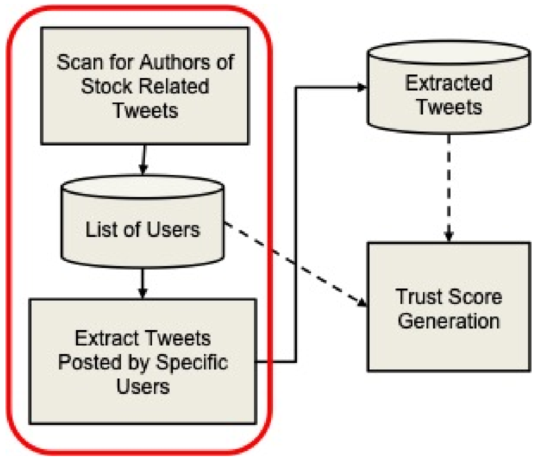


Fig. 2. Tweet Processing Flow. Inside the red frame is the data processing that requires scanning the whole tweet dataset. Dotted arrows distinguish data required by trust scores.

unavoidable, by applying Buy & Hold, we can still establish a primitive method to track the S&P 500 index.

(2) The **Simple Accumulation** strategy takes the Twitter sentiment into account, but sees all users the same. It follows the daily accumulation of all users' sentiments toward a stock to buy, sell or keep the stock. If the sentiment toward a stock is positive and the investor does not already own that stock, the investor buys the stock with an assigned amount (here \$100). If he/she already owns that stock, he/she simply keeps the stock. Consequently, the investor never owns over \$100 of any stock. If the sentiment is negative, the investor sells all the shares for that stock that he/she owns, if any. If the sentiment is not positive nor negative, the investor holds the current shares. Simple Accumulation can be seen as a baseline for comparison with strategies that apply trust filter enhancement.

(3) The strategies based on **Trust Filters**, namely four investment strategies based on **Experience**, **Expertise**, **Authority** and **Reputation** use trust scores as weighting factors for users and combine weighed user sentiments to obtain an overall sentiment as in Equation 2. Each day the weighed sentiment is calculated and then buying and selling takes place similar to the Simple Accumulation strategy, but adding weights to each user's sentiment instead of seeing all users equally. As mentioned before, the trust scores are re-calculated monthly.

In addition, we explore two different initial conditions as follows.

(1) **Buying all available stocks at the beginning** simulates the situation where the investor already owns some shares. At the beginning of the simulation, the investor invests all his/her dedicated dollar amount for each stock. In these experiments, \$100 is dedicated to each stock. As a result, if y different stocks are considered, the investor buys \$100 × y of stocks at the beginning. Here $y = 507$, including the S&P 500 companies as well as the companies frequently mentioned in the stock-related tweets that were not in S&P 500 (Table 2).

(2) **No stocks at the beginning** means that all strategies, except Buy & Hold, hold zero stock at the beginning. This initial condition relies heavily on Twitter sentiment as it only buys stocks suggested by the users. But & Hold still buys \$100 × y of stocks at the beginning.

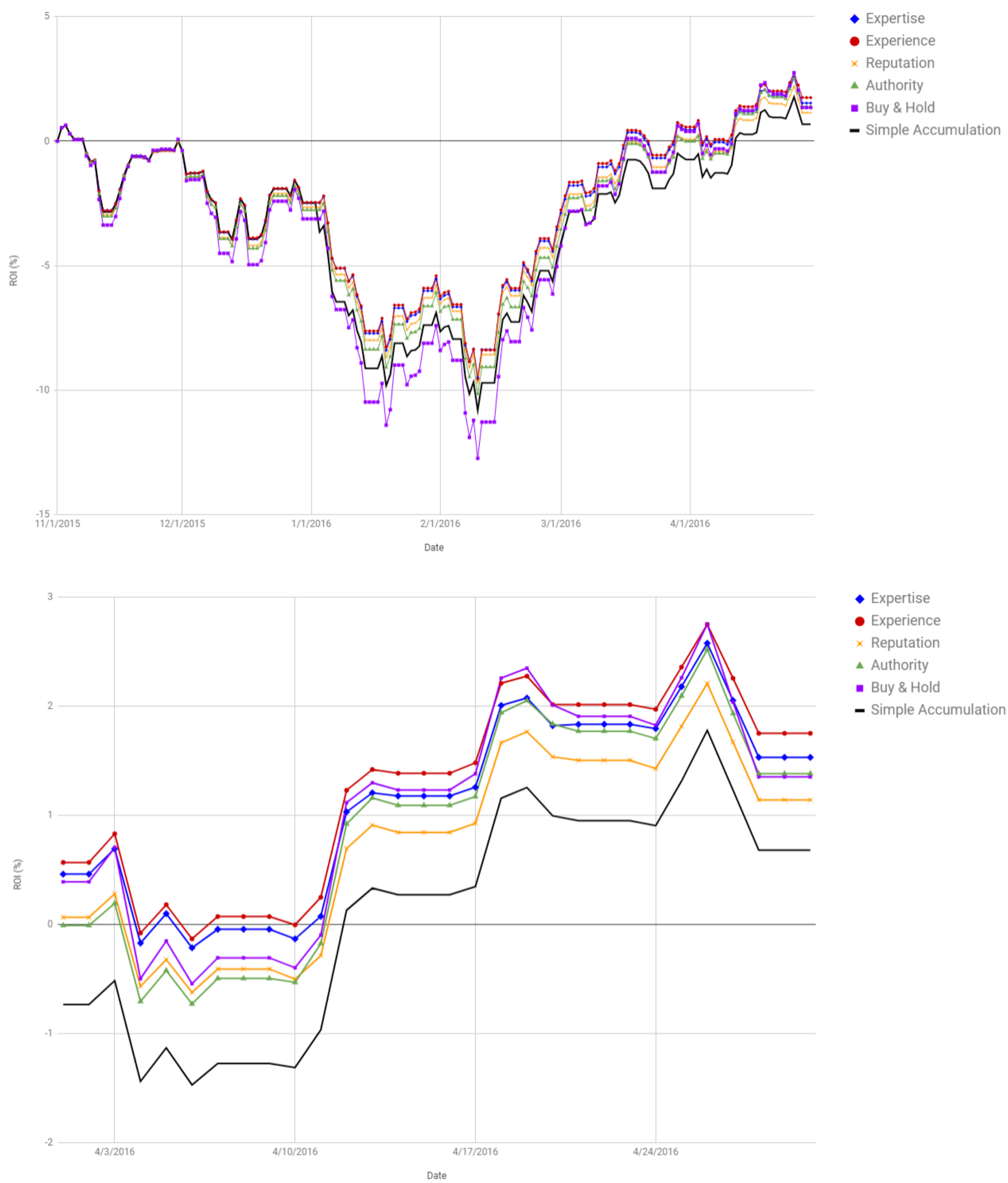


Fig. 3. *Top*: Return on Investment (ROI) with different investment strategies over the six-month period. *Bottom*: Last 30 days of ROI with different investment strategies. Buy all stocks available at the beginning.

Figure 3 shows the Return on Investment (ROI) rate during the six-month period at the top and a magnified view of the last month at the bottom. The initial condition in Figure 3 is to buy all the available shares at first.

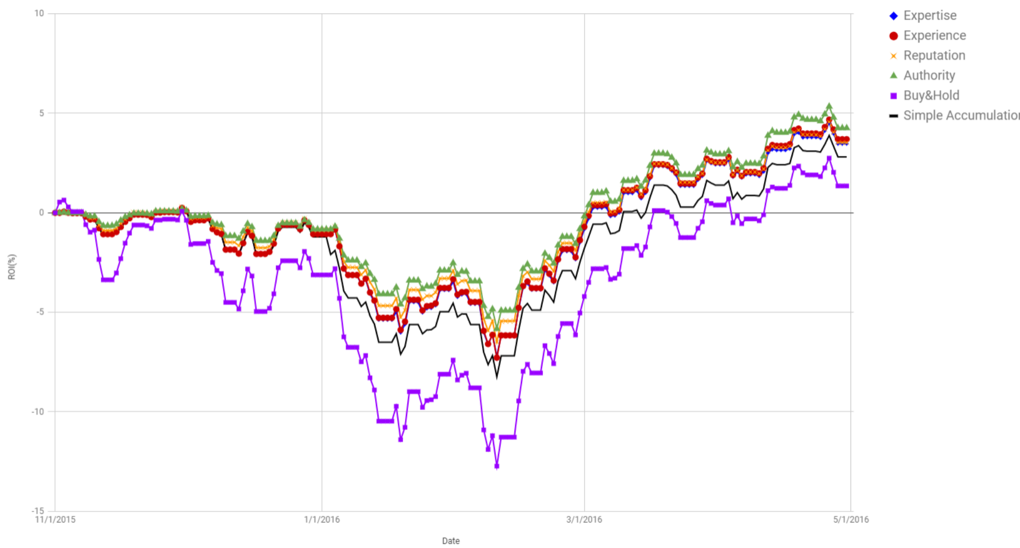


Fig. 4. Return on Investment with different investment strategies over the six-month period. Buy zero shares at the beginning except for the Buy & Hold strategy.

Figure 4 shows the ROI for different investment strategies over the six months with the second initial condition, where, except for Buy & Hold, the investor does not buy any shares at the beginning.

With both of the initial conditions, all the four trust filter-based strategies perform better than Simple Accumulation. Therefore, the trust filters indeed improve the prediction of the stock market through Twitter sentiment analysis. Furthermore, compared to the Buy & Hold strategy, all the four trust filter-based strategies show less loss during the downside of the market for both of the initial conditions, which might imply Twitter users are more sensitive to the stock market during a bear-market period. When utilizing the second initial condition, which is relying more on Twitter sentiment analysis, all the trust filters consistently outperform Simple Accumulation which in turn works better than Buy & Hold (Figure 4).

Table 5 shows a thorough comparison of all strategies and initial conditions. Highest and lowest ROIs are the result of monitoring the ROI over the six months. The final ROI is the value of ROI at the end of the six months. For the initial condition of buying all stocks, Expertise and Experience are the best strategies among all. As for the initial condition of buying zero stocks at first, all Twitter sentiment-based strategies (including the four trust filters and Simple Accumulation) have much better performance than the Buy & Hold strategy. Authority is at the top of all the strategies for this condition. The final ROI shows the potential of outweighing Buy & Hold strategy by adequately applying trust filters.

We also compared the investment strategies according to different time periods of sentiment accumulation (1 day and 7 days)[6]. The ROI result (Figure 5) shows that one-day-accumulation is slightly better than seven-day-accumulation for all the filters. Reputation and Experience show the most significant differences between one-day-accumulation and seven-day-accumulation, which

[6]Note that 1 and 7 days are for sentiment accumulation and not trust score calculation. Trust scores are still updated monthly.

| | Initial condition: buy all stocks | | | Initial condition: buy zero stocks | | |
|---|---|---|---|---|---|---|
| | Highest ROI(%) | Lowest ROI(%) | Final ROI(%) | Highest ROI(%) | Lowest ROI(%) | Final ROI(%) |
| Authority | 2.52 | -10.14 | 1.38 | 5.34 | -5.83 | 4.26 |
| Experience | 2.75 | -9.55 | 1.75 | 4.68 | -7.28 | 3.70 |
| Expertise | 2.57 | -9.48 | 1.53 | 4.51 | -7.21 | 3.48 |
| Reputation | 2.21 | -9.70 | 1.14 | 4.56 | -6.49 | 3.52 |
| Simple Accumulation | 1.78 | -10.81 | 0.68 | 3.89 | -8.24 | 2.81 |
| Buy & Hold | 2.75 | -12.73 | 1.35 | 2.75 | -12.73 | 1.35 |

Table 5. Comparison of Return on Investment (ROI) for all strategies and initial conditions. The green numbers and red numbers are the best and worst ROI performances among all the strategies, respectively.

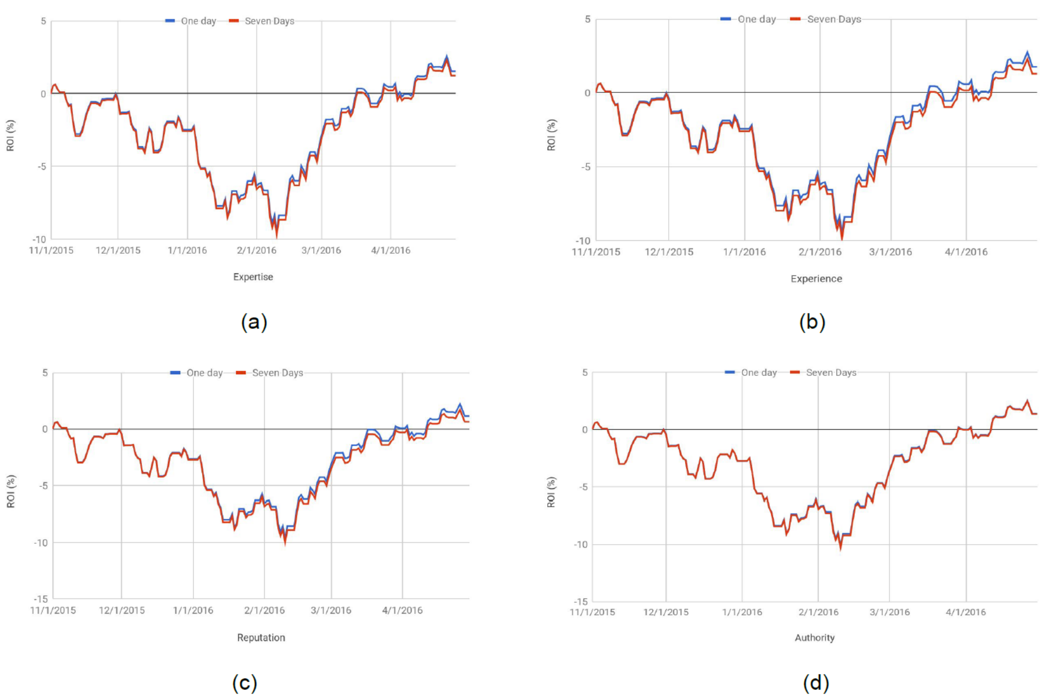


Fig. 5. Comparison of investment strategies based on one-day and seven-day sentiment accumulation. (a) Expertise (b) Experience (c) Reputation and (d) Authority. The initial condition was buying all stocks.

are both around 0.5%. The result implies short time sentiment computation is more suitable for Reputation and Experience. For the Expertise and Authority trust filters, it might be sufficient to change the stock holding weekly instead of daily to decrease the computation effort. The comparison between one-day and seven-day accumulation is consistent with previous work results. Ruiz et al. [38] also found that there is a positive correlation at lag -1, meaning that their tweet features have some predictive power for the stock value of the next day. Similarly, we found one-day accumulation to predict the stock price sufficiently well.


Fig. 6. Histogram chart of trust scores.

## 5.2 Recommendations to Improve Trust Scores

By looking into all the trust scores generated by the trust filters, Figure 6 shows the distribution of trust scores. To make the charts more understandable, $TS_{Authority}$ and $TS_{Reputation}$ are shown in logarithmic scale. Furthermore, we remove the "zero score" points in $TS_{Authority}$ and $TS_{Reputation}$ which account for 89.36% and 28.77%, respectively. Based on the Authority distribution, we can make our first observation about how to improve one's trustworthiness on Twitter: Close to 90% of Twitter users have a 0 Authority score. Therefore, even the slightest success in improving a user's Authority from hubs discerns the user. Of course, it is not very easy to receive a link (here a re-tweet) from hubs as those are the most influential users on Twitter. However, if a user does receive such a link, it is going to be hugely beneficial for his/her trust scores. The same holds true for improving a Reputation score from 0, but with a more modest effect.

To further analyze the distinction between the users with a 0 Authority (or Reputation) score with those with a ">0" Authority (or Reputation) score, let us consider the scenario in which the investor takes the sentiment of one single user to buy and sell stocks. With the initial condition of buying all stocks at first, Table 6 displays the average ROI for this scenario. For example, the first row indicates an average ROI of 0.018, if considering only the sentiment of the users with $TS_{Authority} = 0$. This table shows how ROI increases when making the investment decision based on the sentiment of users with $TS = 0$ versus the users with higher trust scores for Authority and Reputation. We assume that once a user gets approval from other users, in this case is "re-tweeted" by others, the user is certified as a more trustworthy information source. The results show this assumption is reasonable. In fact, basing investment on the sentiment of the users with $TS_{Authority} > 0$ yields an ROI of 22 times more than the users with $TS_{Authority} = 0$. We observe a similar but weaker pattern for Reputation. Because the Authority and Reputation scores peak at $1.0E - 12$ and $1.0E - 4.5$

| | Range of TS | Average ROI |
|---|---|---|
| Authority | "=0" | 0.018 |
| | ">0" | 0.409 |
| | ">1.0E-12" | 0.701 |
| Reputation | "=0" | 0.037 |
| | ">0" | 0.055 |
| | ">1.0E-4.5" | 0.141 |

Table 6. Average ROI by user for different ranges of trust scores. The initial condition was buying all stocks.

respectively, in Table 6 we also aim to find a leading group with respect to Authority and Reputation trust scores. Indeed, the leading groups show a better ROI performance.

From the distributions, we can see $TS_{Authority}$ and $TS_{Reputation}$ are way more diverse than $TS_{Expertise}$ and $TS_{Experience}$. This implies some users have much more influence in terms of stock-related topics in our system. Hence we can make our second recommendation and observation: Compared to the Expertise and Experience filters, one can significantly increase his or her credibility by gaining better trust scores in Authority and Reputation filters.

Looking into users with Expertise of 1 and 0.5 reveals that those are outliers—users with hardly any tweets that are mostly stock-related. Ignoring those two groups, a third observation is that in order to simultaneously improve Expertise and Experience trust scores, it is beneficial for the user to sometimes, but not always, tweet about the topic of interest.

### 5.3 Using Higher Level Sentiment Analysis

As we previously stated, our trust filters are orthogonal to the accompanying sentiment analysis method. To show how trust filters work with higher-than-word-level sentiment analysis methods, in this section, we take advantage of a commercial higher-level sentiment analysis method implemented in the MeaningCloud tool.

MeaningCloud[7] is a commercial text analysis solution. One of the features MeaningCloud provides is sentiment analysis of text. It first analyzes the local polarity (sentiment) of sentences and then the relationship between them. By combining all sentence polarities from the whole text, MeaningCloud generates a global polarity sentiment value of the text. The sentiments are expressed as positive, negative or neutral. It also involves advanced natural language processing techniques to detect the polarity associated with both entities and concepts in the text. Finally, MeaningCloud has the ability to receive as input a dictionary of keywords related to the domain of interest. Considering all these features, we think MeaningCloud's sentiment analysis tool is a great alternative to our sentiment analysis method. MeaningCloud only replaces the dictionary-based sentiment analysis part in our algorithm. We provide our set of keywords as the dictionary to MeaningCloud for its further improvement.

Figure 7 and Table 7 show the ROI of Simple Accumulation of Tweet sentiments when analyzed with MeaningCloud's high level sentiment analysis. They also show how each trust filter can improve the ROI when using MeaningCloud. Finally they display the Buy & Hold strategy's ROI. We observe a similar trend as before: compared to the Simple Accumulation strategy, applying filters can improve the ROI. Similar to when using our dictionary-based sentiment analysis, it is also possible to even outperform Buy & Hold by properly using sentiment analysis and trust filters.

[7]https://www.meaningcloud.com

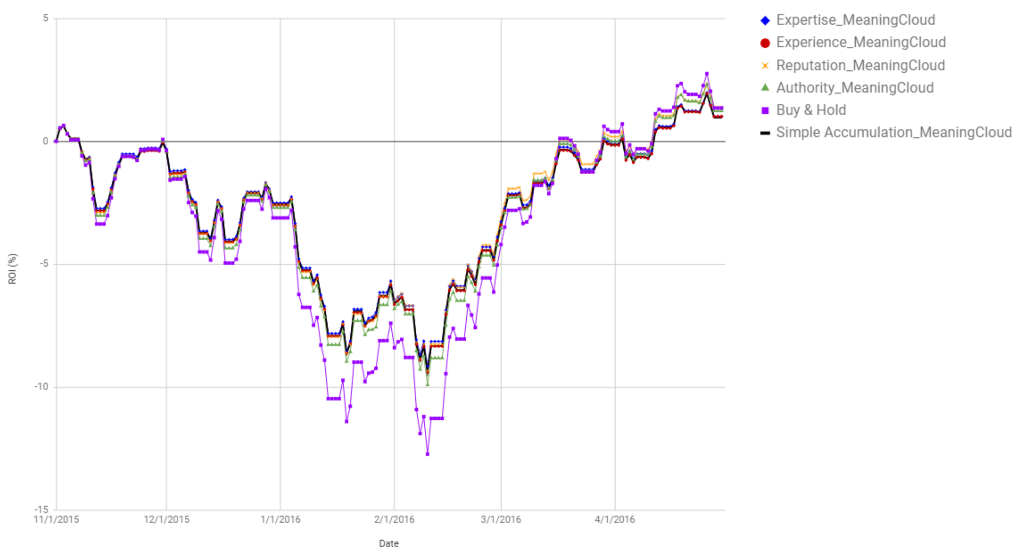


Fig. 7. Return on Investment (ROI) with different investment strategies over the six-month period by applying MeaningCloud's sentiment analysis algorithm including the keyword library model. Buy all stocks available at the beginning.

| | Initial condition: buy all stocks | | |
|---|---|---|---|
| | Highest | Lowest | Final |
| Authority | 2.31 | -9.91 | 1.23 |
| Experience | 1.94 | -9.38 | 1.01 |
| Expertise | 1.97 | -9.15 | 1.01 |
| Reputation | 2.34 | -9.34 | 1.38 |
| Simple Accumulation | 1.94 | -9.38 | 0.95 |
| Buy & Hold | 2.75 | -12.73 | 1.35 |

Table 7. Comparison of Return on Investment (ROI) for all strategies by applying MeaningCloud's sentiment analysis algorithm including the keyword library model. The green numbers and red numbers are the best and worst ROI performances among all the strategies, respectively.

We should note that we are not concerned with the ROI performance comparison between more sophisticated, yet general-purpose sentiment analysis of MeaningCloud and our simpler, but tailor-made dictionary-based sentiment analysis. The contribution of this paper is that no matter what underlying sentiment analysis method is used, trust filters can bring an advantage.

## 6 CONCLUSION

We proposed the application of trust filters for social media sentiment analysis, which is complementary to the underlying text mining techniques used for sentiment analysis and can be applied simultaneously. As an example target domain, we enhanced stock price prediction. We examined four trust filters, namely Expertise, Experience, Reputation, and Authority. Expertise and Experience take into consideration the frequency of discussing a topic by the social media (e.g., Twitter) user.

Reputation and Authority, on the other hand, are based on the network structure of the social media and favor better connected users. We compared all the four trust filters, as well as two base methods named Simple Accumulation (which only accumulates sentiments extracted from tweets and has no trust filter enhancement) and Buy & Hold (which is similar to the concept that is applied by the largest index-based mutual funds and ETFs, representing the conventional investment technique). We also considered two different initial conditions. We measured the return on investment (ROI) of buying and selling stocks, if the investor were to make buying and selling decisions solely based on the sentiment of tweets toward a given stock symbol. Using a sampling of 1% of all the tweets in a six-month period that had both rising and falling stock price periods, we found that all trust filter-based strategies perform better than the Simple Accumulation strategy. This result proves that trust filters can truly filter more trustworthy and useful opinions and constrain the rest. It also decreases loss on bear-market and possibly outperforms the Buy & Hold strategy. The experiments show that better prediction can be achieved through trust filters to enhance users' sentiment accumulation. In addition, trust filters are applicable wherever sentiment analysis is, in several different target domains, e.g, merchandise popularity, movie rating and election prediction.

In this work, we focused on the objective quality of being trusted. However, in offering practical guidelines to users on credibility and trust, we are moving toward subjective trust. The inherent relationship between objective credibility and subjective trust can bring these two lines of research together and is a very promising future work avenue.

## REFERENCES


[1] Harshavardhan Achrekar, Avinash Gandhe, Ross Lazarus, Ssu-Hsin Yu, and Benyuan Liu. 2011. Predicting flu trends using twitter data. In *Computer Communications Workshops (INFOCOM WKSHPS), 2011 IEEE Conference on*. IEEE, 702–707.

[2] Sibel Adali, Robert Escriva, Mark K Goldberg, Mykola Hayvanovych, Malik Magdon-Ismail, Boleslaw K Szymanski, William A Wallace, and Gregory Williams. 2010. Measuring behavioral trust in social networks. In *Intelligence and Security Informatics (ISI), 2010 IEEE International Conference on*. IEEE, 150–152.

[3] Hunt Allcott and Matthew Gentzkow. 2017. Social media and fake news in the 2016 election. *Journal of Economic Perspectives* 31, 2 (2017), 211–36.

[4] Majed Alrubaian, Muhammad Al-Qurishi, Atif Alamri, Mabrook Al-Rakhami, Mohammad Mehedi Hassan, and Giancarlo Fortino. 2018. Credibility in Online Social Networks: A Survey. *IEEE Access* 7 (2018), 2828–2855.

[5] Eiji Aramaki, Sachiko Maskawa, and Mizuki Morita. 2011. Twitter catches the flu: detecting influenza epidemics using Twitter. In *Proceedings of the conference on empirical methods in natural language processing*. Association for Computational Linguistics, 1568–1576.

[6] Eli Bartov, Lucile Faurel, and Partha S Mohanram. 2017. Can Twitter help predict firm-level earnings and stock returns? *The Accounting Review* 93, 3 (2017), 25–57.

[7] Johan Bollen, Huina Mao, and Alberto Pepe. 2011. Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. *ICWSM* 11 (2011), 450–453.

[8] Johan Bollen, Huina Mao, and Xiaojun Zeng. 2011. Twitter mood predicts the stock market. *Journal of computational science* 2, 1 (2011), 1–8.

[9] Kevin R Canini, Bongwon Suh, and Peter L Pirolli. 2011. Finding credible information sources in social networks based on content and social structure. In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on*. IEEE, 1–8.

[10] Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. 2011. Information credibility on Twitter. In *Proceedings of the 20th international conference on World wide web*. ACM, 675–684.

[11] Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. 2013. Predicting information credibility in time-sensitive social media. *Internet Research* 23, 5 (2013), 560–588.

[12] Michael D Conover, Bruno Gonçalves, Jacob Ratkiewicz, Alessandro Flammini, and Filippo Menczer. 2011. Predicting the political alignment of twitter users. In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on*. IEEE, 192–199.

[13] Pooja Devi, Ashlesha Gupta, and Ashutosh Dixit. 2014. Comparative study of hits and pagerank link based ranking algorithms. *International Journal of Advanced Research in Computer and Communication Engineering* 3, 2 (2014), 5749–5754.

[14] Joseph DiGrazia, Karissa McKelvey, Johan Bollen, and Fabio Rojas. 2013. More tweets, more votes: Social media as a quantitative indicator of political behavior. *PloS one* 8, 11 (2013), e79449.
[15] Thomas Dimpfl and Stephan Jank. 2016. Can internet search queries help to predict stock market volatility? *European Financial Management* 22, 2 (2016), 171–192.
[16] Alex Frino and D Gallagher. 2001. Tracking S&P 500 index funds. *Journal of portfolio Management* 28, 1 (2001).
[17] Daniel Gayo-Avello. 2012. " I Wanted to Predict Elections with Twitter and all I got was this Lousy Paper"–A Balanced Survey on Election Prediction using Twitter Data. *arXiv preprint arXiv:1204.6441* (2012).
[18] Alec Go, Richa Bhayani, and Lei Huang. 2009. Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford* 1, 12 (2009).
[19] Aditi Gupta and Ponnurangam Kumaraguru. 2012. Credibility ranking of tweets during high impact events. In *Proceedings of the 1st workshop on privacy and security in online social media*. ACM, 2.
[20] Aditi Gupta, Ponnurangam Kumaraguru, Carlos Castillo, and Patrick Meier. 2014. Tweetcred: Real-time credibility assessment of content on twitter. In *International Conference on Social Informatics*. Springer, 228–243.
[21] Manish Gupta, Peixiang Zhao, and Jiawei Han. 2012. Evaluating event credibility on twitter. In *Proceedings of the 2012 SIAM International Conference on Data Mining*. SIAM, 153–164.
[22] Audun Jøsang, Roslan Ismail, and Colin Boyd. 2007. A survey of trust and reputation systems for online service provision. *Decision support systems* 43, 2 (2007), 618–644.
[23] Pawel Jurczyk and Eugene Agichtein. 2007. Discovering authorities in question answer communities by using link analysis. In *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*. ACM, 919–922.
[24] Dennis Kergl, Robert Roedler, and Sebastian Seeber. 2014. On the endogenesis of Twitter's Spritzer and Gardenhose sample streams. In *Proceedings of the 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. IEEE Press, 357–364.
[25] Jon M Kleinberg. 1999. Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)* 46, 5 (1999), 604–632.
[26] Efthymios Kouloumpis, Theresa Wilson, and Johanna D Moore. 2011. Twitter sentiment analysis: The good the bad and the omg! *ICWSM* 11, 538-541 (2011), 164.
[27] Akshi Kumar and Teeja Mary Sebastian. 2012. Sentiment analysis on twitter. *International Journal of Computer Science Issues (IJCSI)* 9, 4 (2012), 372.
[28] Vasileios Lampos and Nello Cristianini. 2010. Tracking the flu pandemic by monitoring the social web. In *Cognitive Information Processing (CIP), 2010 2nd International Workshop on*. IEEE, 411–416.
[29] Bing Liu. 2012. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies* 5, 1 (2012), 1–167.
[30] Alberto Lumbreras and Ricard Gavaldà. 2012. Applying trust metrics based on user interactions to recommendation in social networks. In *Advances in Social Networks Analysis and Mining (ASONAM), 2012 IEEE/ACM International Conference on*. IEEE, 1159–1164.
[31] Marcelo Mendoza, Barbara Poblete, and Carlos Castillo. 2010. Twitter Under Crisis: Can we trust what we RT?. In *Proceedings of the first workshop on social media analytics*. ACM, 71–79.
[32] Preslav Nakov, Alan Ritter, Sara Rosenthal, Fabrizio Sebastiani, and Veselin Stoyanov. 2016. SemEval-2016 task 4: Sentiment analysis in Twitter. In *Proceedings of the 10th international workshop on semantic evaluation (semeval-2016)*. 1–18.
[33] Arman Khadjeh Nassirtoussi, Saeed Aghabozorgi, Teh Ying Wah, and David Chek Ling Ngo. 2014. Text mining for market prediction: A systematic review. *Expert Systems with Applications* 41, 16 (2014), 7653–7670.
[34] Thien Hai Nguyen, Kiyoaki Shirai, and Julien Velcin. 2015. Sentiment analysis on social media for stock movement prediction. *Expert Systems with Applications* 42, 24 (2015), 9603–9611.
[35] Nuno Oliveira, Paulo Cortez, and Nelson Areal. 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 (2017), 125–144.
[36] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. *The PageRank citation ranking: Bringing order to the web.* Technical Report. Stanford InfoLab.
[37] Josep M Pujol, Ramon Sangüesa, and Jordi Delgado. 2002. Extracting reputation in multi agent systems by means of social network topology. In *Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 1*. ACM, 467–474.
[38] Eduardo J Ruiz, Vagelis Hristidis, Carlos Castillo, Aristides Gionis, and Alejandro Jaimes. 2012. Correlating financial time series with micro-blogging activity. In *Proceedings of the fifth ACM international conference on Web search and data mining*. ACM, 513–522.
[39] Derek Ruths and Jürgen Pfeffer. 2014. Social media for large studies of behavior. *Science* 346, 6213 (2014), 1063–1064.

[40] Harald Schoen, Daniel Gayo-Avello, Panagiotis Takis Metaxas, Eni Mustafaraj, Markus Strohmaier, and Peter Gloor. 2013. The power of prediction with social media. *Internet Research* 23, 5 (2013), 528–543.

[41] Royal Sequiera and Jimmy Lin. 2017. Finally, a downloadable test collection of tweets. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 1225–1228.

[42] Wanita Sherchan, Surya Nepal, and Cecile Paris. 2013. A survey of trust in social networks. *ACM Computing Surveys (CSUR)* 45, 4 (2013), 47.

[43] Mozhgan Tavakolifard, Kevin C Almeroth, and Jon Atle Gulla. 2013. Does social contact matter?: modelling the hidden web of trust underlying twitter. In *Proceedings of the 22nd International Conference on World Wide Web*. ACM, 981–988.

[44] Andranik Tumasjan, Timm Oliver Sprenger, Philipp G Sandner, and Isabell M Welpe. 2010. Predicting elections with twitter: What 140 characters reveal about political sentiment. *ICWSM* 10, 1 (2010), 178–185.

[45] Gang Wang, Tianyi Wang, Bolun Wang, Divya Sambasivan, Zengbin Zhang, Haitao Zheng, and Ben Y Zhao. 2015. Crowds on wall street: Extracting value from collaborative investing platforms. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 17–30.

[46] Hao Wang, Dogan Can, Abe Kazemzadeh, François Bar, and Shrikanth Narayanan. 2012. A system for real-time twitter sentiment analysis of 2012 us presidential election cycle. In *Proceedings of the ACL 2012 System Demonstrations*. Association for Computational Linguistics, 115–120.

[47] Jianshu Weng, Ee-Peng Lim, Jing Jiang, and Qi He. 2010. Twitterrank: finding topic-sensitive influential twitterers. In *Proceedings of the third ACM international conference on Web search and data mining*. ACM, 261–270.

[48] Razieh Nokhbeh Zaeem, David Liau, and K Suzanne Barber. 2018. Predicting Disease Outbreaks Using Social Media: Finding Trustworthy Users. In *Proceedings of the Future Technologies Conference*. Springer, 369–384.

Received February 2018; revised March 2018; accepted June 2018

WWW.IDENTITY.UTEXAS.EDU