Using Historical Social Media Retrieved Trust Attributes to Help Distinguishing Trustworthy Users

Teng-Chieh Huang
The University of Texas at Austin
The Center for Identity
Austin, Texas, USA
e-mail: tengchieh@utexas.edu

K. Suzanne Barber
The University of Texas at Austin
The Center for Identity
Austin, Texas, USA
e-mail: sbarber@identity.utexas.edu

***Abstract*—With the penetration of social media across the world, the information generated by the users has increased exponentially. The wisdom of crowds can now be easily accessed from the Internet. The problem is, how to correctly interpret the true public opinion without distorting it? Considering the spam or malicious users hidden in the social media spreading misinformation and disinformation, the solution might not be trivial. Some previous work uses quantity accumulation trying to mitigate the influence of bad users. More recent works add machine learning techniques to help with the correct judgment. However, the importance of the individual user - the actual person who is behind the screen, does not attract the attention it deserves. In this work, focusing on the history of user behavior, we provide a different angle to understand the connection between the credibility of social media users and the trustworthiness of their virtual representatives. By analyzing Twitter data from November 2017 to November 2021 which contains three types of users (typical, topic-related, and expert) on two target domains (politics and finance), we can gain deeper insights on the social media users and their trustworthiness.**

***Index Terms*—social media, trust, trust attributes, time series forecast, random forests classification.**

I. Introduction

The invention of the Internet and the emergence of social media has exponentially increased the *quantity* of sources a person can contact, but some of these sources may not provide high *quality* information. Social media data is noisy, loaded with contaminated data, advertisements, and scams. The presence of misinformation and disinformation on social media is a rising concern. For those who rely on social media for information, be it for personal use or modern-day analysis of social media data, it is important to know who they can trust. There exist users who abuse the capability of social media by spreading spam, fake news, or biased information. These malicious users not only fool their audience, but may also distort the information retrieved from social media for analytical or prediction purposes.

We aim to find trustworthy information on social media by finding trustworthy users. Specifically, this research addresses how to recommend trustworthy information from specific information sources using an innovative method of combining interpersonal trust on the Internet and classification algorithms for various target domains. We improve the existing and develop new *trust filters*–measurable properties of social media user profile, behavior, connections, posts, etc.–used to identify trustworthy users.

We take a comprehensive list of trust attributes (i.e., properties of a social media user account or posts such as the number of tweets or the number of followers) as potential trust filters from previous work. Further, we consider some established trust filters from previous work.

Using these potential trust filters, we develop a novel, domain-independent method to distinguish three types of social media (specifically Twitter) users from one another: typical users, domain-related users, and experts. We measure trust filters and see how they can rate the three types of users. The value of the trust filter for typical users is the average value for all the users. Domain-related users (e.g., stock-related users) are those that tweet with a set of domain-related handles (e.g., @a_stock_ticker). Experts are real world experts of the field, who we extract from reputable journalistic sources outside social media.

By (1) observing the distribution of the historical trust filter values across user types and domains, and studying the correlation between trust filters, (2) using random forest for time series forecasting to predict new trust filter values, and (3) applying the random forest classifier to compare their performance in distinguishing the types of users, we identify which trust attributes can best distinguish expert, domain-related, and typical users from one another. Most importantly, we keep our work independent of the subject domain by performing the same set of experiments in the finance as well as the politics domain. Moreover, this model can be applied to other target domains, such as health and entertainment.

The rest of the paper is structured as follows. Section II discusses the research that uses social media as a predicting tool in the focus domains of finance and politics. Section III describes how to categorize user groups, data retrieval, and the implementation of the trustworthy users differentiation. Section IV shows the effectiveness of the proposed method with different combinations. Section V concludes the contribution of this work and indicates possible future work.

II. Background

A. Social media predictions in the finance domain

There exists a body of research exploring the power of crowd wisdom as an indicator or a predictor of certain phenomena. In the finance domain, for instance, high social media coverage at the stock level can predict high subsequent return volatility and trading activity [1]. The extracted sentiments of social media users may be used to predict the stock market [2]. Techniques such as text mining [3] and machine learning are frequently used to enhance the predicting power of social media.

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 [4] demonstrated how the Twitter mood, in general, predicts the stock market closing values with high accuracy. Nassirtoussi et al. [5] tried to predict foreign exchange markets based on the text of breaking financial news headlines. Dimpfl et al. [6] 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. [7] 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. Oliveira et al. [8] sought to predict stock markets through Twitter posts. Among all the factors, they found sentiment and posting volume to be particularly important for the forecasting of the Standard and Poor's 500 (S&P 500) index.

B. Social media predictions in the politics domain

The politics domain, particularly elections, is another hotspot for testing the prediction power of social media. Ever since Twitter started becoming an essential online social platform, researchers have been exploring its potential of predicting the election outcome [9]. The 2016 presidential election of the United States, in particular, brought Twitter under the spotlight of public attention. Compared to the Clinton campaign's strategy, the Trump campaign's style in social media points towards de-professionalization and even amateurism as a counter trend in political communication [10]. Moreover, fake news spreading on Twitter [11], social bots distorting public opinion [12], and even Russian interference [13] all played roles on Twitter during the 2016 presidential election. Therefore, it is critical to distinguish real users from fraudulent or malicious sources before utilizing social media as a prediction tool.

The work in [14] showed the distinction among different types of user groups: typical users, target domain-related users and experts of specific domains based on the trust attributes. This work presented the possibility of utilizing social media as a tool to distinguish the more experienced and possibly credible users. This work also suggested some tweet-derived attributes could become promising candidates to differentiate the trustworthy users in specific target domains.

III. Methodology

A. User group categorization

For both the finance and politics domain, the users are categorized based on their level of expertise, involvement, and reputation to the specific target domain. A more detailed definition of the three groups of Twitter users for each target domain is specified below.

1) Finance domain:

- Typical users: The 1% random sampling of all Twitter users (Also known as the Spritzer version of tweets), which stands for the baseline among all groups. In this work, 3000 typical users were randomly selected from the Spritzer version of tweets.
- Stock-related users: The users who posted at least one tweet with a reference to the symbols of the stock market companies, such as $AAPL or $TSLA, during the sampling period. This user group represents the users who might have interest in the stock market, while they may or may not be financial experts. Among all stock-related users collected in 2020-2021 tweet data, 1000 of them were randomly selected for this work.
- Financial experts: This group of users is retrieved from well-known online articles which recommended the top-notch financial experts to follow on Twitter [15]–[17]. These lists are compiled by their authors or by Wall Street analysts and journalists, who are traditionally considered financial authorities. There are 180 recommended financial experts and all of them were included in this work.

2) Politics domain: To keep the consistency between different target domains, the definitions of three groups of Twitter users were similar to the ones defined in the finance domain, which are shown below.

- Typical users: The same 3000 users set as the finance domain.
- Politics-related users: The users who posted tweets with hashtags of the candidates, during the sampling period. 1000 of them were selected in the same manner as stock-related users.
- Political experts: Following the same concept as the financial experts, the political experts were recommended by online articles regarding the specific expertise [18]–[24]. There are 119 recommended political experts and all of them were included in this work.

B. Twitter data retrieval

The list of typical users was randomly selected from the "Spritzer" version of tweets. The lists of two domain-related (finance and politics) users were collected from the same dataset with previous defined requirements. The lists for experts were recommended by the sources mentioned previously. After the list was established, all tweets posted by the listed users were then downloaded using a Python library named "tweepy." The sampling period was from November 1st 2017 to October 31st 2021. The tweets were then divided into 8 segments, each of which contains half-year-long tweets, and trust attributes which will be introduced in the next section.

C. Tweet-retrieved trust attributes

We derive a set of trust attributes retrieved from tweets for each user and use the trust attributes as indicators of the trustworthiness of users. Here, an attribute refers to a user, text, or social connection information of a single social media user. 16 trust attributes (shown in Table I) are selected based on the analysis of previous research [14] which include tweets content, Twitter user information, and social network structure. The trust attributes we chose are neutral, suitable for universal purpose across all domains and do not require any specialized retrieval technique to calculate the attribute for a user. The *expertise_score* is determined by the “keywords” of the corresponding target domain. In the finance domain, the keywords are defined as stock symbols, which is the same as the definition to retrieve stock-related users. In the politics domain, the keywords are sets of frequently used election-related hashtags among political-related users. Those tweets containing keywords are counted as domain-related tweets.

Trust attributes were calculated semiannually from each user’s tweets. Therefore, at most 8 sets (2017$\sim$2021 and twice per year) of trust attributes could be possessed by a single user from the time frame we sampled. The time series of trust attributes represents the change of user behavior in time, which means that by analyzing and understanding how trust attributes change with time, we should be able to forecast the future behavior of each user.

D. Time series analysis

To predict the future trust attributes based on historical data, first we must identify the requirements. The forecast model should be able to forecast multiple trust attributes from multiple historical trust attributes. Here, we assumed trust attributes could be influenced by other trust attributes, which should be self-explanatory. Attributes like Avg_len_tweet and Avg_n_word_tweet are highly related to each other, as are Len_per_word. Many other trust attributes may have implicit influence on others, like Followers_count and Retweet_ratio. However, it is extremely difficult for the classic time series forecast to train a single model that can predict several targets. Not to mention that the short length of data history and huge quantity of users which will require gigantic computation efforts, might provide only mediocre predicting accuracy. Slicing more time frames to the same period of time might be a solution, but since the sampling database was ”Spritzer” version as explained in Section III-B, the tweets generated from a single user in a short period of time could be sparse. If the trust attributes were calculated quarterly or monthly, the majority of users would have merely 1, 2, or even no tweets in most of the time slots.

To mitigate the above-mentioned problems, instead of utilizing traditional time series forecasting methods, this work applied the Random Forests (RF) time series model. There are several advantages for using the RF time series model than other methods of time series forecasting.

- Handles non-linear time series data: many popular classic time series forecasting models such as AutoRegressive Integrated Moving Average (ARIMA) assume linear relationships between variables [28]. However, in the real-world case, many data sets are non-linear and hence require complicated preprocessing of the data to guarantee linearity. This increases the overall complexity of the computation, especially when more variables are considered. On the other hand, RF can handle both linear and non-linear variables well.
- Does not require long historical data: the RF model does not require long continuous time series data, since it only needs the length of predefined lag variables, which is flexible. The lag was set to be 1 in the experiment.
- Decision trees are a great fit to simulate individual user’s behavior: RF is an ensemble learning method. For each individual user, it is not essential for the model to perfectly forecast trust attributes. What we need is the forecast of trust attributes to have high accuracy macroscopically, i.e., the forecast attributes have high accuracy in each user group.
- Generates multiple outputs with only one trained model: Python *sklearn* library supports multiple-outputs fitting in a single model. This is extremely beneficial to this work because it is not just one variable, but 16 variables are being predicted. By training only one model, the computational efforts can be hugely decreased.

To train the RF time series forecasting model, we first converted time series data into supervised learning data. Only users with at least 4 recorded active postings out of 8 time slots were included. There are 902 users in the finance domain (547 typical users, 260 stock-related users, and 95 financial experts) and 967 users in the politics domain (547 typical users, 376 politics-related users, and 44 political experts). The last time slot was used to test the overall prediction accuracy. Therefore, with lag variable setting to be 1, each user could have at least 2 training sets. The number of estimators (decision trees) was 5000. The result will be shown in Section IV.

E. Random forests classification

Different from random forests time series forecast in the previous section, the random forest classification here is used to test the capability of the RF time series forecast to enhance the classification accuracy. To avoid ambiguity, the following section will use the acronym RF-T for random forests time series model and will refer to random forests classification model used to distinguish user groups as RF-C.

Trust attributes were tested having the capability to distinguish between user groups [14]. In this paper, we further tested their capability by training RF-C with different combinations of various historical data and the forecast data made by RF-T. The RF-C models were trained by the following datasets.

- T_{-1}: One time step before the latest set of trust attributes T.

TABLE I
THE DEFINITION OF TRUST ATTRIBUTES.

Trust attributes	Definition	From previous work
Expertise_score	Ratio of tweets which is domain related	[25]
Statuses_count	Total number of posted tweets in user history	[25], [26]
Followers_count	Number of followers of a user	[25]–[27]
Friends_count	Number of friends of a user	[26], [27], [27]
Avg_len_tweet	Average tweet length in characters per tweet	[26]
Avg_n_word_tweet	Average number of words per tweet	[26]
Avg_hashtag	Average number of hashtag symbols per tweet	[25]–[27]
Avg_tweet_URL	Proportion of tweets containing URLs	[25]–[27]
Avg_tweet_question	Proportion of tweets containing "?"	[26]
Avg_tweet_exclamation	Proportion of tweets containing "!"	[25], [26]
Avg_tweet_uppercase	Proportion of tweets composed by upper-case letters	[26]
Retweet_ratio	Proportion of retweets in user's posts	[26]
Len_per_word	Average length of words among all tweets	[25]
Avg_retweet	Average number of retweets a user received per tweet	
Favorite_per_retweet	Average number of favorites received per tweet	
Len_per_word	Average length of characters per word	

- $T_{-1 \to -3}$: 1, 2, and 3 time steps before the latest set of trust attributes T.
- $T_{-1} + T_{Forecast}$: T-1 and the forecast of T .
- $T_{-1 \to -3} + T_{Forecast}$: 1, 2, and 3 time steps before T and the forecast of T.

The latest set of trust attributes T were used as testing sets to verify the accuracy of RF-C in each combination.

IV. Experiment Results

A. Random forests time series forecast

Based on historical trust attributes, RF-T can forecast the future trust attributes of users with great accuracy. For trust attributes having high time-dependency, such as Followers_count (Fig. 1), the forecast values are close to actual values. As for trust attributes with lower time-dependency, such as Avg_len_tweet (Fig. 2), the predicting error tends to be more.

For a baseline comparison, we applied trust attributes one time step before T, which is T_{-1}, as a simplified guess of T. Table II shows Root Mean Square Error (RMSE) of all trust attributes when using T_{-1} and $T_{Forecast}$ as a comparison to T. For both domains, $T_{Forecast}$ has a better RMSE than T_{-1}. This shows the probability of using RF-T to forecast trust attributes.

B. Use time series forecast to enhance social media user classification

Fig. 3 to Fig. 6 highlight the quality of trustworthiness forecast predictions offered by the proposed RF-T method for the finance and the political domains using four training sets. These experimental results report on (1) accuracy measured by the ratio of correct predictions to total predictions and (2) "Area Under the receiver operating characteristic Curve" (AUC), measuring the overall quality of the proposed trust filters (attributes) in predicting classifications of user trustworthiness. A number of decision trees (estimators) were tested from 10 to 1000.

Considering the small difference in RMSEs in the finance domain between T_{-1} and $T_{Forecast}$, it is interesting to see how $T_{Forecast}$ improved the RF-C model, especially regarding AUC. It is also worth mentioning that no matter with or without adding $T_{Forecast}$, $T_{-1 \to -3}$ provides worse accuracy and AUC than T_{-1}. The reason might be that adding older data to train RF-C ends up disrupting the model because outdated data cannot correctly represent the latest trend.

Another observation is that for some cases, adding $T_{Forecast}$ to the training sets did not increase accuracy and AUC. Our explanation is that, when the accuracy or AUC is high enough, adding $T_{Forecast}$ could not enhance the performance since it is already saturated. If we change the classification targets to a different set of users instead of the same set of users, the potential of the enhancement might be better expressed.

Table III and Table IV provide the detailed values of accuracy and AUC when the estimator is 500. We can see that, once accuracy is over 0.95, there is no big difference regarding the addition of $T_{Forecast}$.

TABLE II
RMSE OF T_{-1} AND $T_{Forecast}$.

	T_{-1}	$T_{Forecast}$
Finance Domain	1391	1336
Politics Domain	11872	2848

TABLE III
ACCURACY AND AUC OF FOUR COMBINATIONS OF TRAINING SETS FOR THE FINANCE DOMAIN WHERE THE NUMBER OF ESTIMATORS IS 500.

	Accuracy	AUC
T_{-1}	0.9688	0.9078
$T_{-1} + T_{Forest}$	0.9673	0.9199
$T_{-1 \to -3}$	0.9247	0.8543
$T_{-1 \to -3} + T_{Forecast}$	0.9558	0.9089

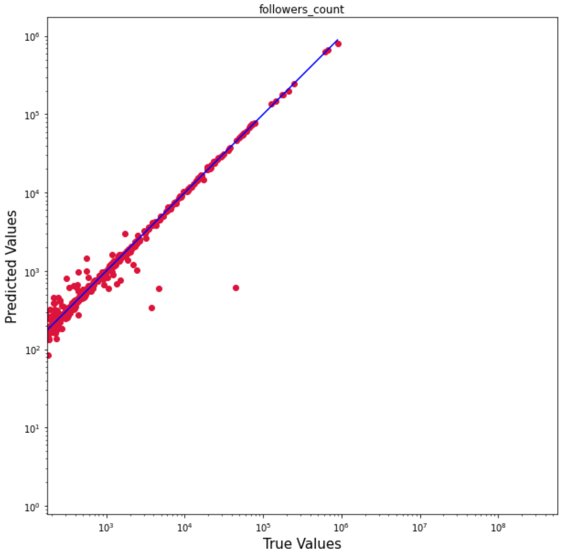

Fig. 1. Comparison of T and $T_{Forecast}$ in Followers_count in the finance domain.

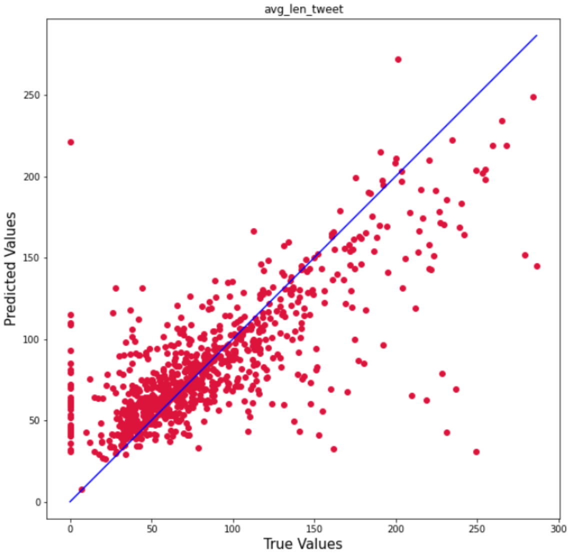

Fig. 2. Comparison of T and $T_{Forecast}$ in Avg_len_tweet in the finance domain.

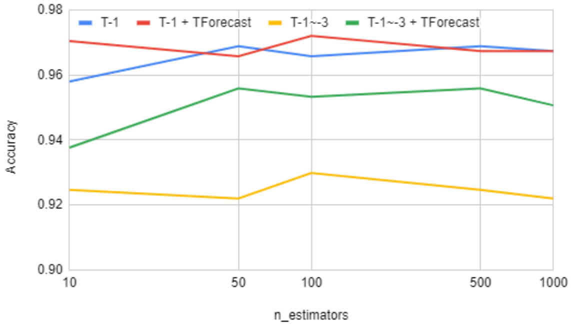

Fig. 3. Accuracy of four combinations of training sets for the finance domain. The number of estimators ranges from 10 to 1000.

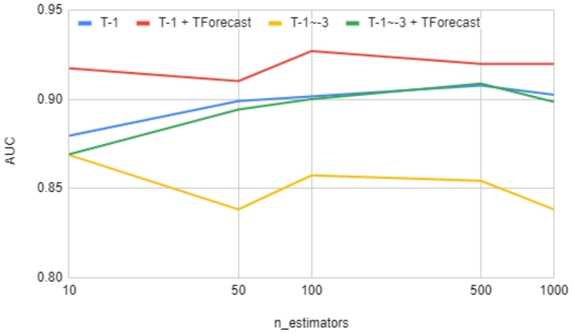

Fig. 4. AUC of four combinations of training sets for the finance domain. The number of estimators ranges from 10 to 1000.

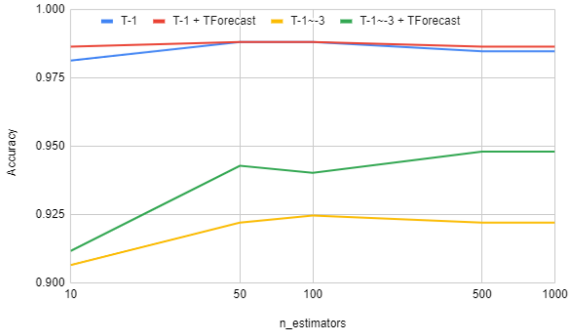

Fig. 5. Accuracy of four combinations of training sets for the politics domain. The number of estimators ranges from 10 to 1000.

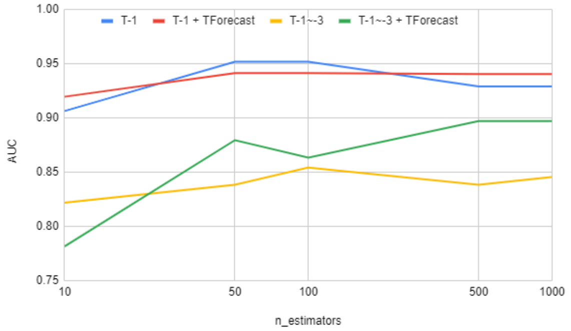

Fig. 6. AUC of four combinations of training sets for the politics domain. The number of estimators ranges from 10 to 1000.

TABLE IV
ACCURACY AND AUC OF FOUR COMBINATIONS OF TRAINING SETS FOR THE POLITICS DOMAIN WHERE THE NUMBER OF ESTIMATORS IS 500.

	Accuracy	AUC
T_{-1}	0.9848	0.9291
$T_{-1} + T_{Forest}$	0.9865	0.9404
$T_{-1 \to -3}$	0.9221	0.8383
$T_{-1 \to -3} + T_{Forecast}$	0.9481	0.8970

V. Conclusion

This work concentrated on developing and evaluating trust filters, measurable trust attributes of social media users that may be able to predict their level of trustworthiness. In this paper, we investigated two target domains, politics and finance, on Twitter. We split social media users into three different groups: typical, domain-related, and expert users. The list of experts was distilled from reputable online sources outside social media. A list of trust attributes was selected, shown to be effective by previous work, as proposed trust filters. We measured the value of these established and proposed trust attributes for Twitter users and generated the distribution of each trust attribute for each of the three user types. The random forest regressor for time series forecast (RF-T) is applied to provide better direction of distinguishing users. We applied the random forests classification algorithm to gauge the effectiveness of trust filters in classifying user types. The results show that, with the addition of RF-T predicted trust attribute values, there is a marked improvement in prediction accuracy and the ability to predict the classification of a user's trustworthiness (Area under the receiver operating characteristic curve, AUC). Our work paves the way for improving the quality of information extracted from social media by focusing on users' trustworthiness and increasing the reliability and utility of this data to explain phenomena, help with decision making, and even predict trends. One possible future work might be using the proposed method to find out more trustworthy users, and extracting users' opinions to make a prediction on certain domains. By comparing the predicting power with other user sets, we can appreciate the effectiveness of this method.

References

[1] P. Jiao, A. Veiga, and A. Walther, "Social media, news media and the stock market," *Journal of Economic Behavior & Organization*, vol. 176, pp. 63–90, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167268120300676

[2] C. Liang, L. Tang, Y. Li, and Y. Wei, "Which sentiment index is more informative to forecast stock market volatility? evidence from china," *International Review of Financial Analysis*, vol. 71, p. 101552, 2020.

[3] A. Sun, M. Lachanski, and F. J. Fabozzi, "Trade the tweet: Social media text mining and sparse matrix factorization for stock market prediction," *International Review of Financial Analysis*, vol. 48, pp. 272–281, 2016.

[4] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *Journal of computational science*, vol. 2, no. 1, pp. 1–8, 2011.

[5] A. K. Nassirtoussi, S. Aghabozorgi, T. Y. Wah, and D. C. L. Ngo, "Text mining for market prediction: A systematic review," *Expert Systems with Applications*, vol. 41, no. 16, pp. 7653–7670, 2014.

[6] T. Dimpfl and S. Jank, "Can internet search queries help to predict stock market volatility?" *European Financial Management*, vol. 22, no. 2, pp. 171–192, 2016.

[7] T. H. Nguyen, K. Shirai, and J. Velcin, "Sentiment analysis on social media for stock movement prediction," *Expert Systems with Applications*, vol. 42, no. 24, pp. 9603–9611, 2015.

[8] N. Oliveira, P. Cortez, and N. Areal, "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*, vol. 73, pp. 125–144, 2017.

[9] A. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welpe, "Predicting elections with twitter: What 140 characters reveal about political sentiment." *ICWSM*, vol. 10, no. 1, pp. 178–185, 2010.

[10] G. Enli, "Twitter as arena for the authentic outsider: exploring the social media campaigns of trump and clinton in the 2016 us presidential election," *European journal of communication*, vol. 32, no. 1, pp. 50–61, 2017.

[11] N. Grinberg, K. Joseph, L. Friedland, B. Swire-Thompson, and D. Lazer, "Fake news on twitter during the 2016 us presidential election," *Science*, vol. 363, no. 6425, pp. 374–378, 2019.

[12] A. Bessi and E. Ferrara, "Social bots distort the 2016 us presidential election online discussion," *First monday*, vol. 21, no. 11-7, 2016.

[13] R. S. Mueller and M. W. A. Cat, "Report on the investigation into russian interference in the 2016 presidential election," 2019.

[14] T.-C. Huang, R. N. Zaeem, and K. S. Barber, "Identifying real-world credible experts in the financial domain," *Digital Threats: Research and Practice*, vol. 2, no. 2, pp. 1–14, 2021.

[15] S. Constable. (2015) "must-follow" twitter feeds on markets and economics. [Online]. Available: https://www.forbes.com/sites/simonconstable/2015/08/14/must-follow-twitter-feeds-on-markets-and-economics/

[16] J. Cummans. (2015) 100 insightful futures traders worth following on twitter. [Online]. Available: https://commodityhq.com/investor-resources/100-insightful-futures-traders-worth-following-on-twitter/

[17] L. Lopez and L. Shen. (2015) The 129 finance people you have to follow on twitter. [Online]. Available: https://www.businessinsider.com/117-finance-people-to-follow-on-twitter-2014-9

[18] S. Emmrich. (2020) The 9 twitter accounts to follow on election day. [Online]. Available: https://www.vogue.com/article/the-nine-best-twitter-accounts-to-follow-on-election-day

[19] J. J. Roberts. (2020) 14 twitter and instagram accounts you should follow for election news day. [Online]. Available: https://fortune.com/2020/11/03/election-night-2020-who-to-follow-twitter-instagram-accounts-follow-updates-trump-biden/

[20] M. Cruz. (2020) 5 journalists to follow on twitter: Politics & news edition. [Online]. Available: https://onepitch.co/blog/5-journalists-to-follow-on-twitter-news-and-politics-edition/

[21] M. Hawkins. (2020) The top 15 conservatives to follow on twitter. [Online]. Available: https://www.thoughtco.com/the-top-conservatives-to-follow-on-twitter-3303615

[22] J. Donatelli. (2017) The 25 must-follow twitter accounts for the anti-trump resistance. [Online]. Available: https://www.lamag.com/culturefiles/social-media-twitter-resistance/

[23] R. Adams. (2020) Top 50 us politics twitter accounts to follow. [Online]. Available: https://www.theguardian.com/world/richard-adams-blog/2010/oct/05/top-50-twitter-accounts-us-politics-election

[24] D. Byers. (2015) Twitter's most influential political journalists. [Online]. Available: https://www.politico.com/blogs/media/2015/04/twitters-most-influential-political-journalists-205510

[25] M. Gupta, P. Zhao, and J. Han, "Evaluating event credibility on twitter," in *Proceedings of the 2012 SIAM International Conference on Data Mining*. SIAM, 2012, pp. 153–164.

[26] C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on twitter," in *Proceedings of the 20th International Conference on World Wide Web*, ser. WWW '11. New York, NY, USA: ACM, 2011, pp. 675–684. [Online]. Available: http://doi.acm.org/10.1145/1963405.1963500

[27] E. J. Ruiz, V. Hristidis, C. Castillo, A. Gionis, and A. Jaimes, "Correlating financial time series with micro-blogging activity," in *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*, ser. WSDM '12. New York, NY, USA: ACM, 2012, pp. 513–522. [Online]. Available: http://doi.acm.org/10.1145/2124295.2124358

[28] M. J. Kane, N. Price, M. Scotch, and P. Rabinowitz, "Comparison of arima and random forest time series models for prediction of avian influenza h5n1 outbreaks," *BMC bioinformatics*, vol. 15, no. 1, pp. 1–9, 2014.