Fake Tweets Detection and Its Impacts on the 2020 U.S. Election Prediction

Miftahul Qorib
Department of Computer Science and Information Technology
University of the District of Columbia
Washington, DC, USA
miftahul.qorib@udc.edu

Junwhan Kim
Department of Computer Science and Information Technology
University of the District of Columbia
Washington, DC, USA
junwhan.kim@udc.edu

***Abstract*—The increased social media usage in modern history instigates data collection from various users with different backgrounds. Mass media has been a rich source of information and might be utilized for countless purposes, from business and personal to political determination. Because more people tend to express their opinions through social media platforms, researchers are excited to collect data and use it as a free survey tool on what the public ponders about a particular issue. Because of the detrimental effect of news on social networks, many irresponsible users generate and promote fake news to influence public belief on a specific issue. The U.S. presidential election has been a significant and popular event, so both parties invest and extend their efforts to pursue and win the general election. Undoubtedly, spreading and promoting fake news through social media is one of the ways negligent individuals or groups sway societies toward their goals. This project examined the impact of removing fake tweets to predict the electoral outcomes during the 2020 general election. Eliminating mock tweets has improved the correctness of model prediction from 74.51 percent to 86.27 percent with the electoral outcomes of the election. Finally, we compared classification model performances with the highest model accuracy of 99.74634 percent, precision of 99.99881 percent, recall of 99.49430 percent, and an F1 score of 99.74592 percent. The study concludes that removing fake tweets improves the correctness of the model with the electoral outcomes of the U.S. election.**

Keywords*— *fake tweets detection, machine learning, election prediction, social media

I. Introduction

In this modern era, the internet has been a tremendous source of information [1]. Many individuals convey their opinions through social media [1] [2]. Many suggest that the election results were changed by a substantial increase of fake news stories posted on social media to mislead public opinion for financial or ideological gain [3] [4]. A tweet generator is the ultimate tool for mock-ups and malice with friends designed for tweet campaigns and brainstorming new ideas [5].

Tremendous sources of information from social media have triggered concerns about the impact of poisonous fake news and inaccurate information in societies [6] [7]. During the 2016 U.S. election, over 100 websites were reported to generate and promote major fake news regularly [8]. Moreover, others might use bots and cyborgs to spread false news on social media platforms [8]. In order to fight fake news, it is critical to identify the origins and patterns of the fake stories on social platforms [6]. During the 2016 U.S. election, some tweets were favorited more than 13,000 times and retweeted nearly 1,300 times, but it was written in St. Petersburg, Russia, about 5,000 miles away [9].

Detecting misinformation in mass media is technically challenging because of tedious evidence collection and careful fact-checking [10]. Unverified accounts are believed to be the majority sources of misinformation stories in social media [11]. Identifying distorted news becomes more challenging when displayed by reputed and trusted sources using multiple platforms [12]. About 86 to 91 percent of users spread false information by retweeting or liking the original resources, 5 to 9 percent of users retweeting and asking if the information is correct, while 1 to 9 percent of users doubt or confirm that the original posts are not accurate [13]

As many believe that the news on social media is not always accurate, it might benefit societies to use real news from social media platforms to reflect honest public opinions. Modern technology can produce textual content that is extremely difficult to distinguish from what humans create used for disinformation campaigns and gets amplified detrimental effects once they spread through social media [14]. In this case, there is a possibility that machines can create unlimited quantity of tweets, while humans can only produce limited tweets daily.

Our previous research [15] used public tweets to gain insight into the U.S. presidential election. However, we did not separate fake tweets from the dataset. This project aims to see the impact of removing fake tweets on the result of the U.S. presidential election.

II. Literature Review

Table 1 represents a comparative study of fake tweet detection projects. Some of the projects utilized machine learning methods [16] [17] [18] [19], neural networks [20] [21] [22] [23] [24], while others combined both machine learning and neural networks [25] [26] [27]. Comparing previous studies on fake tweets will aid this project in discovering new benefits in detecting fake tweets. The table is a representation of different fake tweet detection from diverse topics.

According to the previous research on fake tweets, as shown in Table 1, several gaps in the literature need to be addressed. Most of the previous research projects were applied to Covid-19 topics [16] [17] [22] [23] [24] [25]. A research study discussed fake news on the U.S. presidential election [28], but it did not perform fake tweet detection. Most of the projects used less than 50,000 datasets [16] [25] [17] [18] [19] [20] [21] [26] [24]. There is a possibility that fake tweets influence the presidential election, so using an election dataset might enhance the work study.

TABLE I. FAKE TWEETS RESEARCH STUDIES

Title	Authors	Dataset	Method	Conclusion
Using artificial intelligence techniques for detecting Covid-19 epidemic fake news in Moroccan tweets	Youness Madani, Mohammed Erritali, and Belaid Bouikhalene	10,000 tweets	Logistics regression, decision tree, random forest, naïve Bayes, gradient boosting, SVM, multilayer perceptron	Random forest achieves the highest accuracy of 79 percent [16]
CoAID-DEEP: An Optimized Intelligent Framework for Automated Detecting COVID-19 Misleading Information on Twitter	Diaa Salama Abdelminaam, Fatma Helmy Ismail, Mohamed Taha, Ahmed Taha, Essam H. Houssein, and Ayman Nabil	926 Covid-19 tweets, 7613 disaster tweets, 618 politifact tweets, 5328 gossip cop tweets	Decision Tree, Random Forest, KNN, SVM, Logistic Regression, Naïve Bayes, RNN, LSTM	LSTM achieves the best accuracy of 98.57 percent [25]
Fake News Detection of South African COVID-19 Related Tweets using Machine Learning	Yaseen Khan and Surendra Thakur	36,254 tweets	Extra Trees Classifier, Light GBM Classifier, SVC, Random Forest, XGB, Logistic Regression, Bernouli NB, Linear Discriminant Analysis, Bagging, Ridge Classifier CV, Ridge Classifier, Nearest Centroid, Decission Tree, Quadratic Discriminant Analysis, Calibrateed Classifier CV, SGDClassifier, NuSVC, Linear SVC, Ada Boost Classifier, ExtraTree Classifier, Perceptron, Gaussian NB, Passive Aggressive Classifier, K-Neighbors, Label Spreading, Label Propagation, and Dummy Classifier	Achive best model classifier using ExtraTreesClassifier with 0.869 accuracy and 0.823 balance accuracy [17]
Detecting Fake News with Tweets' Properties	Ning Xin Nyow and Hui Na Chua	23,206 tweets	Random Forest and Decision Tree	Random forest achieves the highest accuracy of 98.6% percent [18]
The Detection of Fake News in Arabic Tweets Using Deep Learning	Shatha Alyoubi, Manal Kalkatawi, and Felwa Abukhodair	5,000 tweets	CNN, BiLSTM, and MARBERT-CNN	MARBERT-CNN achieved the best accuracy of 95.48% [20]
CERIST'22: Classifying COVID-19 Related Tweets for Fake News Detection and Sentiment Analysis with BERT-based Models	Rabia Bounaama and Mohammed El Amine Abderrahim	8,661 tweets	BERT	The project achieves an f1-score of 93% on sentiment analysis and 90% on fake news detection [21]
Combining exogenous and endogenous signals with a semi-supervised co-attention network for early detection of COVID-19 fake tweets	Rachit Bansal, William Scott Paka, Nidhi, Shubhashis Sengupta, and Tanmoy Chakraborty	254,402 tweets	HAN, MixText, GCAN, CSI, dEFFEND, CNN-MWSS, RoBERTa-MWSS, FNED, PPC and ENDEMIC	The ENDEMIC (Exogenous and eNDogenous) model achieves the best accuracy of 93.7% [22]
Cross-SEAN: A cross-stitch semi-supervised neural attention model for COVID-19 fake news detection	William Scott Paka, Rachit Bansal, Abhay Kaushik, Shubhashis Sengupta, and Tanmoy Chakraborty	45,261 tweets	MTL, 1HAN, 16HLT-HAN, 3HAN, CSI, dEFEND, MixText, and Cross-SEAN	The cross-SEAN model achieves the best accuracy of 95.4% [23]
Framework for Detecting Fake Retweets Using Deep Neural Network	Sampad Dinesh Hegde, Akhilesh Shetty, Manoj NM, Abhigna Kalasad, and Bharathi R	1,610 users	LDA-KNN, LDA-SVM, LDA-Naïve Bayes, FCNN-LDA, and BiLSTM-BoW	BiLSTM-BoW achieves the best accuracy of 92.4% [26]
Identifying and Classifying Fake COVID-19 Tweets using Transformer Models	Yasmine Eid Mahmoud, Farid Ali Mousa, and Ayat Mahmoud	2,140 tweets	BERT, RoBERTa, and DistilBERT	RoBERTa performs the best with an accuracy of 97% [24]
MVAN: Multi-View Attention Networks for Fake News Detection on Social Media	Shiwen Ni, Jiawen Li, and Hung-Yu Kao	305,904 users	SVM-BOW, BiLSTM, TextCNN, CSI, CRNN, dEFEND, GCAN, G-SEGA, and MVAN	MVAN achieves the highest accuracy of 93.65% [27]
Intelligent Detection of False Information in Arabic Tweets Utilizing Hybrid Harris Hawks-Based Feature Selection and Machine Learning Models	Thaer Thaher, Mahmoud Saheb, Hamza Turabieh, and Hamouda Chantar	3,023 tweets	Random Forest, Support Vector Machines, Linear Regression, Decision Tree, KNN, LDA, Naïve Bayes, and Xgboost	Linear regression with TF-IDF achieves the best accuracy of 81.50 % [19]
Influence of fake news on Twitter during the 2016 U.S. presidential election presidential election	Alexandre Bovet and Hernán A. Makse	171 million tweets	CI algorithm	There are 25% of these tweets are fake tweets or extremely biased information [28]

Title	Authors	Dataset	Method	Conclusion
Behavior-Based Machine Learning Approaches to Identify State-Sponsored Trolls on Twitter	Saleh Alhazbi	2,086,000 tweets	Decision Tree, Random Forest, Adaboost, and Gradient Boosting	Gradient Boost model classifier reaches the highest accuracy of 94.4% [29].

III. Methodology

The experimental design of this project is shown in Figure 1. In this study, we used 2020 presidential election tweets collected by recognizing the names of candidates (Joe Biden and Donald Trump) [30]. Tweets were collected in twenty-five days from October 15, 2020, until November 8, 2020. There are 1.74 million tweets (#biden 776,886 tweets and #trump 970,922 tweets), sufficient for data analysis [31]. For this research project, we used verified accounts with geocode activated only (179,347 tweets for Biden and 212,744 tweets for Trump), about 23 % of the total tweets.

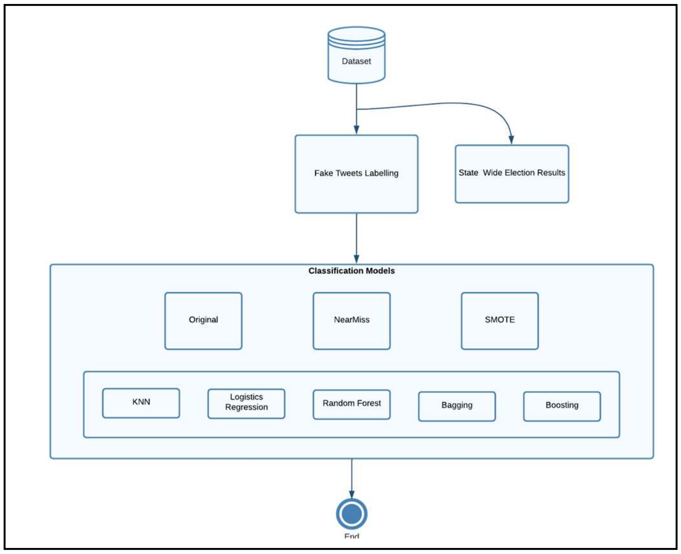

Fig. 1. Experimental Design

The next step is tweet labeling to separate fake tweets from the datasets. We used the number of tweets each user produced during the collection periods (25 days) as a key factor to determine tweet groups. GPT-3, for instance, can easily produce accurate information that humans cannot distinguish it, which can create more compelling disinformation [32]. Consequently, computers can generate more tweets in a certain period than humans. So, the more tweets' users produce in a certain period, the more likely it is a fake tweet. We propose a new probabilistic model indicating the likelihood of a tweet being fake based on the frequency of tweets from a user. Equation 1 is based on logistic regression where the dependent variable is the probability of a tweet being fake, modeled as a function of the tweet rate.

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 * TR \quad (1)$$

Where p is the probability of being fake tweet, β_0 is a constant (intercept), β_1 is the regression coefficient and TR (Tweet Rate) is an independent variable. From the equation, we can derive probability of being fake tweet (p) in Equation 2.

$$p(Fake|TR) = \frac{1}{1+ e^{-(\beta_0+ \beta_1 * TR)}} \quad (2)$$

A user, steveziegenbus2, for example, posted 1259 tweets (about fifty-one tweets per day) during the collection period.

We iterated the accuracy of the tweet sentiment in each state and matched it with the presidential election results based on the number of tweets users posted. We achieved the closest match of the tweet sentiments with the election results when 46 or fewer tweets posted were grouped into real tweets, while 47 or more were labeled fake. Applying this tweets classification, we found that in the 2020 U.S. presidential election there were about 26% (101,908 tweets) of fake tweets and 74% (290,180 tweets) of real one as shown in figure 2. Our result corresponds to the previous research study [28], which stated that in the 2016 U.S. presidential election, there were 25% of the tweets were fake or extremely disinformation tweets.

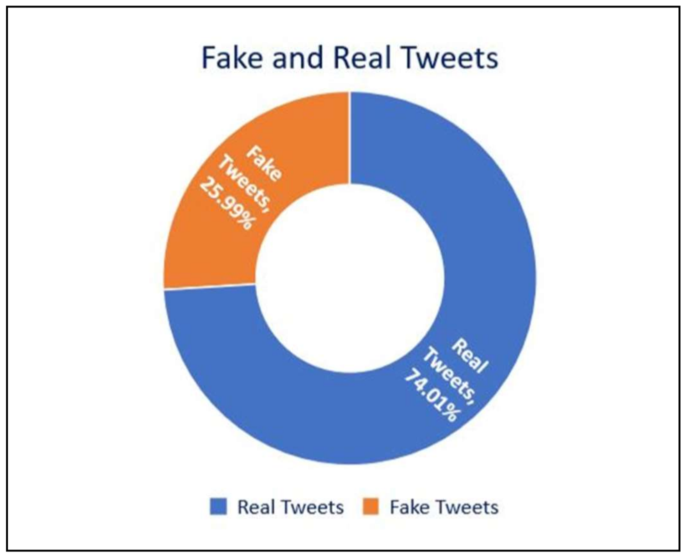

Fig. 2. Fake and Real Tweets

Table 2 represents descriptive statistics of fake tweets users. There is a possibility that 969 users produced fake tweets on the 2020 U.S. election. Within the 25 days, they have generated 101,908 tweets with an average of 105 tweets. Most users generated 50 tweets within the collection period, about 2 tweets per day. The maximum tweets users produced was 1259, about 50.36 tweets per day. The lowest tweets they produced were 47 posts, about 1.88 tweets per day (closed to 2 tweets a day).

$$25\ t \geq 47 \quad (3)$$

$$t \geq 1.88 \quad (4)$$

Where t is the number of tweets posted per day from a certain user.

TABLE II. DESCRIPTIVE STATISTICS FAKE TWEETS USERS

Scores	
Mean	105.1682
Standard Error	3.551643
Median	71
Mode	50
Standard Deviation	110.5583
Sample Variance	12223.13
Kurtosis	36.38257
Skewness	5.248748
Range	1212
Minimum	47
Maximum	1259
Sum	101908
Count	969

Equation 3 and 4 represent the cutoff number of tweets (t) that we labelled as fake in this study. We labeled it as fake tweets when users produced two or more same topic tweets (1.88 tweets) per day for 25 days. To achieve maximum engagement in social media, users suggested to post 1-5 tweets per day, while larger enterprises and major brands recommended 3 tweets per day [33]. From 1.6 million Twitter users, average users posted 4.422 tweets per day [34].

The next step was performing classification models. In this stage, we used *NearMiss, SMOTE* (Synthetic Minority Oversampling Technique), and the original dataset. NearMiss is one of the under-sampling methods that involves randomly selecting examples from the majority class and removing them from the training dataset [35]. On the other hand, SMOTE is an over-sampling technique in which synthetic samples are randomly generated for the minority class [36]. In this study, we applied KNN, logistics regression, random forest, bagging, and boosting classification models.

IV. EXPERIMENT RESULT

Removing the fake tweets from the dataset affects each state's average negative tweet sentiment to get closer to the actual result of the 2020 U.S. presidential election in each state. For example, in Idaho (Trump won), Trump's negative sentiment with fake tweets (27.80%) was more than Biden's (20.79%). After removing the fake tweets, Trump's negative sentiment (14.66%) was less than Biden's (16.09%). With fake tweets in the dataset, we had 28 TRUE and 23 FALSE statements. However, after removing the fake tweets, the TRUE claims improved to 30, and FALSE records lowered to 21. The TRUE assertions increase from 54.90 percent to 58.82 percent, while FALSE ones decrease from 45.10 percent to 41.18 percent. Consequently, removing fake tweets from the dataset improves the model performance.

Positive tweet sentiments in each state also improved after removing the fake tweets from the dataset. Computation of positive sentiments of each candidate in each state gets closer to the 2020 U.S. election results. For example, in Hawaii, where Biden won the state, his positive sentiment (32.44%) was less than Trump's (33.57%). After removing the fake tweets in the dataset, Biden's positive sentiment (35.46%) was more than Trump's (32.21%). The TRUE statements increased from 31 to 37, while the FALSE statements decreased from 20 to 14. Removing the fake tweets improved the TRUE claims from 60.78 percent to 72.55 percent, while FALSE statements decreased from 39.22 percent to 27.45 percent. Consequently, the model performance in the positive sentiment has improved after we removed the fake tweets from the dataset.

Furthermore, we might gain insights into the presidential election results by combining positive and negative sentiments. Removing fake tweets from the dataset improved the election prediction from 74.51 to 86.27 percent. With fake tweets in the dataset, we found 38 TRUE statements and 13 FALSE. On the other hand, removing fake tweets improved the election prediction from 38 to 44 TRUE statements and 13 to 7 FALSE statements. In Massachusetts (Biden won), for instance, with fake tweets in the dataset, Biden was predicted to lose based on negative and positive sentiment computations. After taking away the fake tweets, he was predicted to win the state. Overall, removing the fake tweets improved the possibility of predicting the result of the election by 86.27 percent.

Table 3 represents the classification model performance. The highest accuracy, 99.74634 percent, was achieved when we applied the Random Forest model classifier using SMOTE, followed by the Random Forest with *NearMis*s (99.66207 percent) and the KNN without imbalanced method (99.63478 percent). In addition, the highest model precision was achieved when we were using the Random Forest with SMOTE (99.99881 percent), followed by the Random Forest with *NearMiss* and without imbalanced method (99.99821 percent). Our classification model outperformed previous research study [18] with the highest accuracy of 98.6 percent applying the Random Forest model classifier.

TABLE III. MODEL PERFORMANCE

Model	SMOTE				NearMiss				Original			
	Acc	*Precis*	*Recall*	*F1*	*Acc*	*Precis*	*Recall*	*F1*	*Acc*	*Precis*	*Recall*	*F1*
MLP	96.64653	99.38423	93.88058	96.55404	93.62691	97.94729	89.15842	93.34644	95.53982	99.55241	94.18961	96.79679
Gradient Boosting	97.56198	99.25747	95.8452	97.5215	96.54971	97.4489	95.62224	96.52693	97.50977	97.66908	98.87927	98.27045
Log Reg	91.63563	94.63104	88.29535	91.35348	94.75509	94.59451	94.96662	94.7802	96.7069	97.47662	97.9325	97.70403
KNN	98.84396	99.97328	97.71605	98.83178	99.3485	99.94571	98.75432	99.34644	99.63478	99.98281	99.50665	99.74416

Model	SMOTE				NearMiss				Original			
	Acc	***Precis***	***Recall***	***F1***	***Acc***	***Precis***	***Recall***	***F1***	***Acc***	***Precis***	***Recall***	***F1***
NB	96.36494	95.90546	96.87203	96.38632	94.52497	91.87493	97.72321	94.70887	96.72628	97.45575	97.9824	97.71837
Linear SVC	95.18217	92.64986	98.16002	95.32538	96.34905	97.82316	97.05788	97.43902	93.61068	91.80959	99.99002	95.72535
Decision Tree	96.78475	97.91928	97.57701	97.74784	94.7775	98.53402	90.93754	94.58349	96.80483	97.92838	97.59885	97.76334
Random Forest	99.74634	99.99881	99.4943	99.74592	99.66207	99.99821	99.52955	99.76333	99.49756	99.99821	99.29966	99.64771

V. Contributions

We would like to highlight the following contributions of this study on detection fake tweet on election tweet dataset .

- We were able to generalize a formula to detect a fake tweet ($t \geq 1.88$, where t is the number of tweets produced per day). None of the previous research studies discussed the number of tweets machine produced daily contributed to the possibility of being fake tweets. We believe that the number of tweet human can produce daily is a critical factor in detecting a fake tweet since machines can produce tremendous tweets daily.

- We analyzed the impact of removing fake tweets on the election result. None of the previous research projects were doing it. The closest study [28] was using similar dataset (2016 U.S. election tweets), but it did not analyze the impact of fake tweets on the election result. Their finding (25% of election tweets were extremely biased) is relevant to our finding that 25.99 percent of the 2020 U.S. election tweets were fakes. Our research study suggests that removing fake tweets from the election dataset was able to improve election prediction from 74.51 percent to 86.27 percent.

- Our best model performance was on the Random Forest with an accuracy of 99.74634 percent with the precision of 99.99881 percent, a Recall of 99.4943 percent, and an F1 score of 99.74592 percent. The previous study applied Random Forest classification model [18] achieved the best accuracy of 98.6 percent with a recall of 95.4 percent and an F1 score of 97.2 percent (no precision score).

VI. Future Work

This study successfully gives insight into using sentiment analysis to predict the presidential outcome. Future work studies could also analyze the account's activity before and after the election, as many accounts might be used for election purposes only. In addition, using real-time tweets could give accurate and real-time situations of the election outcome. Furthermore, using neural networks in future research might give a better model performance, become conventional, and adopt prediction models.

VII. Conclusion

Social media is a famous and rich digital state from which many users benefit. It has been widely used in countless sectors from different backgrounds. Tremendous social media usage might promote and generate fake news regularly, especially when dealing with significant campaigns such as elections. Over 100 websites were reported to promote fake news regularly during the 2016 U.S. election. Populating fake news in mass media platforms might distort and bias the information. Detecting fake tweets has become challenging as modern technology can produce undistinguishable textual content from what humans create to disinform and distort news for amplified effect campaigns. Detecting and removing fake tweets from social media datasets could benefit societies remarkably.

In this research study, removing fake tweets has improved the use of sentiment analysis to predict the election outcome. Since not all information presented on social platforms is valid, removing fake tweets means eliminating distorted information from social platforms, so it minimizes biased information in mass media. Detaching fake news has proven beneficial to get a better insight into the electoral outcomes. The Random Forest model classifier applying SMOTE technique achieved the highest accuracy.

The 2024 U.S. presidential election is fast approaching, and social networks have already seen an influx of news and messages, both in support of and against various candidates. These digital communications hold considerable power to shape voter perceptions and decisions. While this study focused on analyzing tweets to predict outcomes in the 2020 U.S. election, the findings offer valuable insights that may apply to the dynamics of the upcoming 2024 election. As one of the most pivotal political events in the country, the 2024 election will likely be influenced by similar patterns of online discourse and engagement.

ACKNOWLEDGMENT

This work has been supported by the National Science Foundation under Grant No. 2131269, through the CISE-MSI Program (RCBP-ED: CNS: Data Science and Engineering for Agriculture Automation)

References

[1] C. McClain, E. A. Vogels, A. Perrin, S. Sechopoulos and L. Rainie, "The Internet and the Pandemic," PEW RESEARCH CENTER, Washington, 2021.

[2] A. Wong, O. Olusanya, M. V. Antonini and D. Lyness, "The use of social media and online communications in times of pandemic COVID-19," *Journal of the Intensive Care Society,* vol. 22, no. 3, pp. 25-260, 22 October 2020.

[3] H. Allcott and M. Gentzkow, "Social Media and Fake News in the 2016 Election," *Journal of Economic Perspectives,* vol. 31, no. 2, pp. 211-236, 2017.

[4] P. Gordon and D. G. Rand, "Prior Exposure Increases Perceived Accuracy of Fake News," *Journal of Experimental Psychology: General,* 2017.

[5] Twiends, "Tweet Generator," Twiends, 2009. [Online]. Available: https://twiends.com/tools/tweet-generator. [Accessed 17 October 2023].

[6] D. Lazer, M. Baum, N. Grinberg, L. Friedland, K. Joseph, W. Hobbs and C. Mattsson, "Combating Fake News: An Agenda for Research and Action," 2017.

[7] P. Mihailidis and S. Viotty, "Spreadable spectacle in digital culture: Civic expression, fake news, and the role of media literacies in "post-fact" society," *American Behavioral Scientist,* vol. 61, no. 4, pp. 441-454, 2017.

[8] C. Shao, G. L. Ciampaglia, O. Varol, K. Yang, A. Flammini and F. Menczer, "The spread of low-credibility content by social bots," *Nature Communication,* 2018.

[9] J. Peebles, "2016 ELECTION INVESTIGATION: Russian trolls targeted Georgia in fake tweets," CMG Corporate Services, Inc., Atlanta, 2018.

[10] S. D. Das, A. Basak and S. Dutta, "A Heuristic-driven Uncertainty based Ensemble Framework for Fake News Detection in Tweets and News Articles," *Neurocomputing,* vol. 491, pp. 607-620, 2022.

[11] S. Muhammed and S. K. Mathew, "The disaster of misinformation: a review of research in social media," *International Journal Data Science Analytics,* vol. 13, no. 4, pp. 271-285, 2022.

[12] S. B. Parikh, S. R. Khedia and P. K. Atrey, "A Framework to Detect Fake Tweet Images on Social Media," in *2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM)*, Singapore, 2019.

[13] Wells Media Group, Inc., "Beware of Fake Tweets During Disasters," Wells Media Group, Inc., San Diego, 2018.

[14] J. Tourille, B. Sow and A. Popescu, "Automatic Detection of Bot-generated Tweets," in *1st ACM International Workshop on Multimedia AI against Disinformation, Jun 2022*, Newark, USA, 2022.

[15] M. Qorib, R. S. Gizaw and J. Kim, "Impact of Sentiment Analysis for the 2020 U.S. Presidential Election on Social Media Data," in *ICMLT 2023, March 10–12, 2023*, Stockholm, Sweden, 2023.

[16] Y. Madani, M. Erritali and B. Bouikhalene, "Using artificial intelligence techniques for detecting Covid-19 epidemic fake news in Moroccan tweets," *Result in Physics,* vol. 25, p. 104266, 2021.

[17] Y. Khan and S. Thakur, "CoAID-DEEP: An Optimized Intelligent Framework for Automated Detecting COVID-19 MisleadingFake News Detection of South African COVID-19 Related Tweets using Machine Learning Information on Twitter," in *2022 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD)*, Durban, KwaZulu Natal, South Africa, 2022.

[18] N. X. Nyow and H. N. Chua, "Detecting Fake News with Tweets' Properties," in *2019 IEEE Conference on Application, Information and Network Security (AINS)*, Penang, Malaysia, 2019.

[19] T. Thaher, M. Saheb, H. Turabieh and H. Chantar, "Intelligent Detection of False Information in Arabic Tweets Utilizing Hybrid Harris Hawks Based Feature Selection and Machine Learning Models," *Symmetry,* vol. 13, no. 556, 2021.

[20] S. Alyoubi, M. Kalkatawi and F. Abukhodair, "The Detection of Fake News in Arabic Tweets Using Deep Learning," *Applied Sciences,* vol. 13, p. 8209, 2023.

[21] R. Bounaama and M. E. A. Abderrahim, "CERIST'22: Classifying COVID-19 Related Tweets for Fake News Detection and Sentiment Analysis with BERT-based Models," in *CERIST'22: CERIST NLP Challenge 2022*, Algiers, Algeria, 2022.

[22] R. Bansal, W. S. Paka, Nidhi, S. Sengupta and T. Chakraborty, "Combining exogenous and endogenous signals with a semi-supervised co-attention network for early detection of COVID-19 fake tweets," in *Lecture Notes in Computer Science*, vol. 12712, Springer, Cham, 2021.

[23] W. S. Paka, R. Bansal, A. Kaushik, S. Sengupta and T. Chakraborty, "Cross-SEAN: A cross-stitch semi-supervised neural attention model for COVID-19 fake news detection," *Applied Soft Computing,* vol. 107, p. 107393, 2021.

[24] Y. E. Mahmoud, F. A. Mousa and A. Mahmoud, "Identifying and Classifying Fake COVID-19 Tweets using Transformer Models," in *2023 Intelligent Methods, Systems, and Applications (IMSA)*, Cairo, Egypt, 2023.

[25] D. S. Abdelminaam, F. H. Ismail, M. Taha, A. Taha, E. H. Houssein and A. Nabil, "CoAID-DEEP: An Optimized Intelligent Framework for Automated Detecting COVID-19 Misleading Information on Twitter," *IEEE,* vol. 9, pp. 27840-27867, 2021.

[26] S. D. Hegde, A. Shetty, M. NM, A. Kalasad and B. R, "Framework for Detecting Fake Retweets Using Deep Neural Network," in *2022 IEEE 7th International conference for Convergence in Technology (I2CT)*, Pune, India, 2022.

[27] S. Ni, J. Li and H.-Y. Kao, "MVAN: Multi-View Attention Networks for Fake News Detection on Social Media," *IEEE,* vol. 9, pp. 106907-106917, 2021.

[28] A. Bovet and H. A. Makse, "Influence of fake news in Twitter during the 2016 US presidential election," *Nature Communications,* vol. 10, no. 7.

[29] S. Alhazbi, "Behavior-Based Machine Learning Approaches to Identify State-Sponsored Trolls on Twitter," *IEEE Access,* vol. 8, pp. 195132 - 195141, 2020.

[30] M. Hui, "US Election 2020 Tweets," Kaggle, 2020.

[31] C. Bayrak and M. Kutlu, "Predicting Election Results via Social Media: A Case Study for 2018 Turkish Presidential Election," *IEEE transactions on computational social systems,* p. 1–12, 2022.

[32] G. Spitale, N. Biller-Andorno, and F. Germani, "AI model GPT-3 (dis)informs us better than humans," *Science Advances,* vol. 9, p. 26, 2023.

[33] N. Ellering, "How Often To Post On Social Media? [Proven Research From 14 Studies]," CoSchedule, 18 October 2017. [Online]. Available: https://coschedule.com/blog/how-often-to-post-on-social-media. [Accessed 28 April 2024].

[34] D. Zarrella, "Is 22 Tweets-Per-Day the Optimum?," HubSpot, 11 October 2022. [Online]. Available: https://blog.hubspot.com/blog/tabid/6307/bid/4594/is-22-tweets-per-day-the-optimum.aspx. [Accessed 28 April 2024].

[35] J. Brownlee, "Undersampling Algorithms for Imbalanced Classification," 27 January 2021. [Online]. Available: https://machinelearningmastery.com/undersampling-algorithms-for-imbalanced-classification/. [Accessed 17 November 2023].

[36] S. Satpathy, "SMOTE for Imbalanced Classification with Python," Analytics Vidhya, 17 November 2023. [Online]. Available: https://www.analyticsvidhya.com/blog/2020/10/overcoming-class-imbalance-using-smote-techniques/. [Accessed 11 December 2023].