International Journal of Multidisciplinary Engineering in Current Research - IJMEC
Volume 8, Issue 4, April-2023, http://ijmec.com/, ISSN: 2456-4265

SPAMMER DETECTION AND FAKE USER IDENTIFICATION ON SOCIAL NETWORKS

Asad Saifullah, Mohammed Sajid[2], Mohammed Abdul Muqtadir[3], Saleha Butool[4]
[1,2,3] B.E. student, Department of IT, Lords Institute of Engineering and Technology, Hyderabad
[4] Assistant Professor, Department of IT, Lords Institute of Engineering and Technology, Hyderabad
salehabutool@lords.ac.in

Abstract: Spammers are using social networking platforms like Twitter and Facebook to spread harmful and useless material. For instance, Twitter is now among the most costly networks and allows an overwhelming quantity of spam. Legitimate users are negatively impacted and resource use is disrupted when fake users send unwanted tweets to advertise products or websites. False identities are now more likely to be used to spread false information, increasing the risk of dangerous content spreading. In today's online social networks, spotting spammers and spotting phony users on Twitter has become a popular study topic. This study examines methods for identifying spammers on Twitter and provides a taxonomy of such methods, categorizing them according to how well they can identify bogus material, URL-based spam, hot topics, and fake accounts.

I. INTRODUCTION

Information may now be easily found online, and social media platforms like Twitter have grown to be well-liked sources for up-to-the-minute information. Twitter is an Online Social Network (OSN) where users may discuss current events and politics as well as share news, ideas, and emotions. The necessity to research and examine user behavior on these platforms has grown, though, as it's possible that many individuals may fall for scams. Combating spammers who just utilize OSNs for marketing is also necessary. Therefore, it is essential to keep an eye on and manage these users.

Researchers are focusing on detecting spam on social networking sites (SNS) to protect users from malicious attacks and maintain their privacy. Spammers use various tactics, such as spreading fake news, rumors, and messages, to spread false information. They achieve their goals through advertisements and mailing lists, causing disturbance to non-spammers and lowering the reputation of OSN platforms. Therefore, it is crucial to design a scheme to identify spammers and implement corrective measures to counter their malicious activities.

Surveys on false user identification have been undertaken as part of substantial research on Twitter spam detection. New approaches and strategies were surveyed by Tingmin et al., and Twitter spammer behavior was examined by authors of [5]. The body of literature now in existence nevertheless has a void. This survey examines the state-of-the-art in spammer detection and false user identification on Twitter, offers a taxonomy of Twitter spam detection methods, and gives a thorough overview of recent advancements in an effort to close this gap.

This study examines numerous techniques for Twitter spam detection and provides a taxonomy. false content, URL-based detection, hot themes detection, and false user identification are the four approaches that have been found. In order to give readers with a central location to access a variety of information on spam detection strategies, the research evaluates the characteristics and approaches already in use.

The taxonomy of methods used to identify spammers on Twitter is presented in this article, along with discussion of suggested approaches, an analysis, and recommendations for further research.

II. SPAMMER DETECTION ON TWITTER

The approaches used to identify spammers on Twitter are grouped into four primary categories in this article: false content, URL-based spam detection, spam detection in hot topics, and fake user identification. Specific models, methods, and detection algorithms are used for each category. Regression prediction models, malware warning systems, the Lfun scheme technique, URL-based spam detection with machine learning algorithms, trending topic spam detection with Nave Bayes classifiers, and false user identification are some of the methods for detecting fraudulent material.

A. FAKE CONTENT BASED SPAMMER DETECTION

Gupta et al. examined the expanding harmful material in the Boston bombing and identified prominent people in charge of disseminating false information. They classified the information using temporal analysis using the biggest collection of its kind, 7.9 million tweets from 3.7 million people.

According to the study's analysis of fictitious Twitter user accounts, followers shared the majority of spam messages. While non-informative tweets were mostly produced using online interfaces, mobile devices were the sources of tweet analysis. Fake material was detected using user characteristics including the average number of verified accounts and followers. In order to forecast the effects of propagating bogus material and future growth, the authors employed a regression prediction model.

TABLE 1. Comparison between proposed methods for spam detection in Twitter.

Ref	Proposed Method	Goal	Dataset	Results
[15]	Dirichlet distribution has been used by the statistical framework for identifying spammer in Twitter	Distinguish between spammer and non-spammer	Real data of Twitter and Instagram	Experimentation carried out on Instagram and Twitter data shows that supervised and unsupervised algorithmic methods deliver meaningful outcomes.
[16]	Effective unified weighted for anomalous URL detection	Detection of anomalies behaviour in users interaction.	Twitter dataset is used, which contains last 200 tweets of users.	Anamolous detection model can be used to analyze effectively the number of URLspammer that is done every day
[2]	Using manual inspection, classification of users as spammer and non-spammer	Detection of spammer on Twitter	Twitter dataset that includes more than 1.9 billion links and tweets aroung 1.8 billion.	Classification of spammer uses a large set of atributes
[17]	Three types of cascade information, which are created on the basic of spam detection mechanism, have	Spammers have been classified by using the properties of	Real Twitter dataset	The schemes are scalable because they check users cantered 2-hops social

	been used, I.e., TSP, SS, and cascade filtering	social networks in the individual social environment.		networks instead of examining the whole network.
[18]	Design of 18 robust features by holding the time properties explicitly and implicitly.	Answer the question of how to identify spammer only.	Crawled and manually annotated dataset	The features extracted are able to recognize both authentic users and spammers accurately up to 93%.
[7]	Inductive e-learning technique for the Twitter spammer detection has been used.	User's behaviour and tweet content have been analysed for the purpose of finding the best feature to recognize Twitter spammers.	A set of 62 features has been used for identifying spammers using crawler.	Random-forest system provides adequate results in malicious user spammer detection, having a detection accuracy that exceeds results presented in the existing literature.
[19]	Text pre-processing technique was conducted, and four different feature set	The objective of the study isto detect spam tweets which enhance the quantity of data that needs to be assembled by relying only on tweet-inherent features.	2 large labelled dataset of tweets containing spam.	An inspiring result was achieved by using the limited feature set that is accessible in tweets, which is better as compared to exixting spammer detection systems.

The suggested architecture analyzes tweets, finds occurrences that are allowable, and reports them. When people identify spam or malware or when security reports are published, it leverages tweets to do so. The system consists of real-time data extraction, a preprocessing schedule and Naive Bayes algorithm-based filtering system, data analysis for spammer identification, an alert subsystem for locating pertinent tweets, and feedback analysis. The method is said to be successful and efficient at spotting invasive and cancerous activity in the blood. The technology groups pertinent tweets based on the cluster barycenter and chooses the closest tweet to serve as the lone representation of the cluster as a whole.

TABLE 2. Comparison of different features used for spam detection in Twitter.

Ref.	User feature								Content feature								Graph feature		Structure feature				Time feature	
	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16	F17	F18	F19	F20	F21	F22	F23	F24
[13]	✓	✓	✓	✓	-	-	-	-	-	✓	✓	-	-	-	✓	✓	✓	✓	-	-	-	-	-	-
[11]	✓	✓	✓	-	✓	✓	-	-	✓	✓	✓	✓	✓	✓	✓	-	-	-	-	-	-	-	-	-
[15]	✓	✓	✓	-	-	-	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	✓
[12]	-	-	-	-	-	-	-	-	✓	✓	✓	✓	✓	✓	-	-	-	-	-	-	-	-	-	-
[33]	✓	-	✓	-	-	✓	-	-	✓	-	-	-	-	-	-	-	-	-	✓	-	-	-	-	-
[10]	✓	-	✓	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	✓	✓	✓	✓	-	-
[8]	-	-	-	-	-	-	-	✓	-	-	✓	-	-	-	-	-	-	-	-	✓	-	-	✓	-
[2]	✓	✓	✓	-	-	-	-	✓	-	✓	✓	✓	✓	-	✓	-	-	-	-	✓	-	-	-	-
[14]	✓	✓	-	-	-	-	✓	-	-	✓	-	✓	-	-	-	✓	-	-	-	-	-	-	-	-
[24]	-	-	-	-	✓	-	-	-	-	✓	✓	✓	-	-	-	-	-	-	✓	-	-	✓	✓	-

F1	Number of Followers	**F9**	Number of retweets	**F17**	In/out degree
F2	Number of Following	**F10**	Number of hashtags	**F18**	Betweenness
F3	Age of account	**F11**	Number of user mention	**F19**	Average Tweet Length
F4	Reputation	**F12**	Number of URL	**F20**	Time between first - last Tweet
F5	Number of user favorites	**F13**	Number of Characters	**F21**	Depth of conversion Tree
F6	Number of Lists	**F14**	Number of Digits	**F22**	Tweet frequency
F7	Propagation of Bidirectional	**F15**	Number of Tweets	**F23**	Tweet sent in time interval
F8	Number of replies	**F16**	Spam words	**F24**	Idle time in days

Spam tweets were identified and detected using a novel stream-based clustering technique by Eshraqi et al. [8]. To identify tweets as spam and nonspam, they chose user accounts from diverse datasets and randomly picked tweets. The algorithm was found to partition data with a high degree of accuracy and to identify false tweets with a high degree of precision. Spam was identified using a variety of criteria, including graph-based features, content-based features, and time-based features. The study employed 50,000 user accounts and had great accuracy in identifying spammers and bogus tweets.

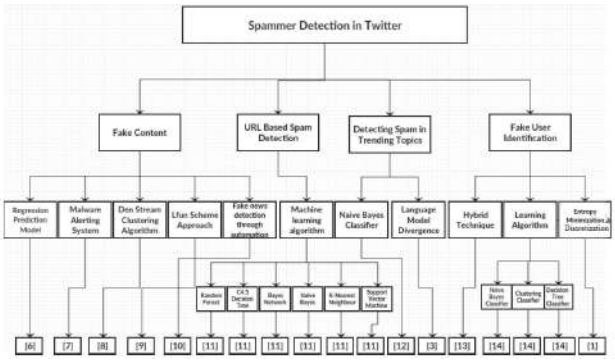

FIGURE 1. Taxonomy of spammer detection/fake user identification on Twitter.

A Lfun technique for Twitter spam detection was presented by Chen et al. It consists of two parts: learn from detected tweets (LDT) and learn from human labeling (LHL). Through the use of an algorithm called random forest, the technique creates spam tweets from unlabeled tweets. The performance was assessed using real-world data over the course of 10 consecutive days, and the accuracy of spam detection significantly improved.

Using two credibility-focused datasets, Buntain et al. devised a system to automatically identify false news on Twitter. Based on journalist evaluations, a crowdsourced employee was used to train the model. The research discovered 45 elements that aid in analyzing material on social media and spotting trends in false news, including structural, user, content, and temporal factors.

B. URL BASED SPAM DETECTION

By examining parameters including the ratio of spam to non-spam tweets, the size of the training dataset, time-related data, factor discretization, and data sampling, Chen et al. assessed machine learning techniques for identifying spam tweets. To find spam tweets, they gathered 600 million public tweets and used Trend Micro's web reputation system.

The study classifies spam tweets using machine learning and identifies 12 user- and tweet-based attributes. To reproduce various scenarios, samples from the databases are used. The findings demonstrate that while there are no variations in the distribution of the training dataset, modifying the feature distribution decreases spam detection effectiveness. The study's goal is to make spam detection techniques better.

C. DETECTING SPAM IN TRENDING TOPIC

A approach is introduced by Gharge et al. [3] and classified based on two novel elements. The first involves identifying spam tweets without any prior knowledge of the users, while the second is studying language to find spam on a current popular subject on Twitter. The system framework consists of the next five phases.

1. A compilation of tweets related to Twitter's hot topics. After being saved in a certain file format, the tweets are then examined.
2. Spam is labeled in order to search through all available datasets and find the malicious URL.
3. Using language as a tool, feature extraction isolates the characteristics construct based on the language model, which aids in identifying whether or not the tweets are phony.
4. The shortlist of tweets that best describe the set of characteristics given to the classifier to teach the model and learn the information necessary for spam identification is used to classify the data set.
5. The classication approach is used by the spam detector to accept tweets as input and classify them as spam and nonspam.

To evaluate the system's accuracy, an experimental setup was created. A random sample set of 1,000 tweets was used for this, of which 60% were lawful and the remaining 40% had flaws.

The effect of spam on Twitter trending topics was studied by Stafford et al. [12]. They suggested a technique to identify spam and assess its effects on these subjects. Within one hour, they identified 10 worldwide hot topics with language codes and gathered all the information permitted by the Twitter API. Following collection, the tweets were divided into groups for spam and non-spam, which may be utilized to train classifiers.

Using URL filtering, a software was created to manually classify random tweets. The analytical approach comprised choosing and analyzing characteristics using information retrieval measures, and then using statistical tests to determine how spam filtering affected trending themes. The findings demonstrated that spammers do not embrace Twitter's hot topics instead choose target themes that meet the necessary criteria, demonstrating Twitter's viability and room for development.

D. FAKE USER IDENTIFICATION

Utilizing a manually gathered dataset of 501 false and 499 legitimate accounts, Erahin et al. suggested a classification approach for identifying spam accounts on Twitter. Username, profile, background picture, friends and followers count, tweet content, account description, and amount of tweets were all examined in the study. The Naive Bayes learning method was used in two studies to categorize false accounts, one before and one after discretization.

A hybrid approach for spammer profile detection was developed by Mateen et al. [13] integrating user-based, content-based, and graph-based features. The model uses three criteria to distinguish between spam and non-spam profiles. A Twitter dataset of 400K tweets and 11K individuals was used to assess the method. For spam identification, user-based attributes including reputation, age, follower-to-follower ratio, and number of followers are crucial. The relevance of these qualities in spam identification is highlighted by the relationship between content features and duplicate tweets from spam bots.

Spammers employ a number of elements to disseminate false information and advertise their goods. The quantity of tweets, hashtag ratio, URL ratio, mention ratio, and tweet frequency are examples of content-based attributes.

Spammers' evasion techniques, such as purchasing and transferring false followers to seem real, are controlled using graph-based characteristics. The method uses historical Twitter datasets for evaluation and combines Decorate, Naive Bayes, and J48. The findings demonstrate an extremely precise detection rate that outperforms current methods. Data access by the public is prohibited per Twitter policy.

A Twitter spam detection strategy was created by Gupta et al. utilizing Naive Bayes, clustering, and decision trees. The dataset categorizes accounts as spam or non-spam using 1064 individuals and 62 attributes. The dataset contains 36% of the spammer account accounts. Based on user and tweet level variables including followers, spam terms, responses, hashtags, and URLs, features are identified.

To recognize spam accounts, the authors created a method utilizing clustering, decision trees, and Naive Bayes algorithms. Naive Bayes calculates the likelihood that a particular account is spammer or not. Accounts are divided into types of spammers and non-spammers by the clustering-based technique. Using a decision tree method, tree topologies are designated and decisions are made at each level. When compared to spam accounts, the suggested technique performs better at identifying non-spammer accounts.

III. COMPARISON OF APPROACHES FOR SPAM DETECTION ON TWITTER

This section compares the suggested methodologies, along with their aims, datasets used to assess spam, and the outcomes of each method's tests, as given in Table 1.

A. ANOMALY DETECTION BASED ON URL

A approach for identifying URL irregularities in tweets on social networking sites like Twitter was put out by Chauhan et al. The technique takes into account adult material, virus content, tweet similarity, timing difference, and URL ranking. 200 tweets have been collected as a dataset to examine unusual behavior depending on URL. Five new functions are added to the dataset: time difference calculation, adult content recognition, malware URL rank assignment, URL rank generation, and tweet similarity.

ALEXA's source code is used to determine the URL rank, whereas tweet similarity analyzes whole tweets. Malware URL rank assignment checks the reputation of the URL using the WebOfTrust (WOT) API. Time difference calculations create clusters of seven tweets by comparing each tweet with its predecessor and succeeding tweet. A collection of all URLs that include adult material is created using adult content identification. The findings demonstrate that the suggested anomaly detection approach is capable of accurately estimating the quantity of Ueffectively RL spammers.

Ghosh et al.'s [22] analysis of Twitter spam accounts and their link-building activities allowed them to study the tactics new spammers in Online Social Networks (OSNs) utilize. They discovered that spammers employ clever situations to avoid detection and boost spam production. A dataset of eight Twitter spam accounts was employed in the study, and it was discovered that spammers regularly publish tweets with related website URLs, which are used to spot malicious individuals. They also point out legitimate users who return the favor by following other scammers.

Ghosh et al.'s [22] analysis of Twitter spam accounts and their link-building activities allowed them to study the tactics new spammers in Online Social Networks (OSNs) utilize. They discovered that spammers employ clever situations to avoid detection and boost spam production. A dataset of eight Twitter spam accounts was employed in the study, and it was discovered that spammers regularly publish tweets with related website URLs, which are

used to spot malicious individuals. They also point out legitimate users who return the favor by following other scammers.

In a research on confusing information in Twitter spam, Chen et al. examined a two-week Twitter feed that contained URLs. They discovered that spammers employ enclosed URLs to make it more likely for victims to fall for scams, download malicious software, and fall for phishing. They utilized Trend Micro's WRT, which has a low false positive rate, to recognize spam. The research employed a clustering technique to divide non-spam and spam tweets into categories in order to better understand the diversity of unclear subjects used in Twitter spam.

Malware, phishing, Twitter follower scams, and advertising are the four categories into which spam tweets are divided using the graphical clustering method using bipartite cliques. Given that eradicating spam is expensive to implement in the real world, this method aids in the advancement of spam detection strategies. The investigation reveals that the traits utilized have drawbacks including being simple to trick or hard to extract. This strategy is difficult since about 400 million tweets every day contain URLs.

B. MACHINE LEARNING ALGORITHMS

Benevenuto et al. used a massive dataset of more than 5400 million users, 1.8 billion tweets, and 1.9 billion connections to do research on spammer identification on Twitter. They found user attributes and tweet content aspects as machine learning criteria for user classification. They employed a tagged collecting technique to obtain 80 million user IDs from Twitter in order to identify spammers. An individual numeric ID was given to each user, and efforts were made to create a tagged collection with the appropriate attributes. Based on their actions, such as the frequency of their interactions, user traits were discovered.

Two sets of attributes—content attributes and user behavior attributes—are used in the study to identify users. While user behavior characteristics collect particulars like posting frequency, interaction, and influence on Twitter, content attributes concentrate on the wordings of tweets. The 23 qualities that are taken into account by the approach include followers, age, tags, responses, tweets that were received, time, and daily and weekly tweets. Despite these characteristics, the framework can frequently identify spammers.

In their investigation of follow spam on Twitter, Jeong et al. [17] proposed categorization methods for identifying spammers. Using two-hop subnetworks, they presented two mechanisms: social status filtering and trade importance profile filtering. They also suggested using assembly approaches and cascade filtering to merge social status and trade importance profile attributes.

Using actual data, the study evaluated Twitter's dependability and believability. It suggested using incomplete data in TSP and SS filters to detect spammers in real-time. In comparison to earlier methods, the results demonstrated a scalable methodology that focused on user-centered two-hop social networks and dramatically improved false and true positive performance.

In order to find spammer insiders in machine learning systems, Meda et al. [21] presented a method employing the random forest algorithm. The framework integrates bootstrap aggregation, non-uniform feature sampling, random forest, and unplanned feature selection.

The authors used a non-uniform feature sampling method to test the random forest algorithm's performance on users with indefinite behaviors. They divided feature selection into random selection and domain expert selection. Two datasets were used to demonstrate the efficiency of the non-feature sampling technique. The experiments showed the potency of the enriched feature sampling technique.

A technique for spotting fraudulent Twitter user identities was developed by David et al. utilizing user profiles and timelines. They divided timeline-based characteristics into content-based and metadata-based ones, resulting in 71 low-cost variables. To choose the optimum feature combination, the approach employed variable significance. There were utilized five supervised classifiers, with random forest averaging the greatest accuracy. The method validated the viability of practical devices and efficient detection.

The research used five supervised classifiers to rank feature sets, including decision trees, support vector machines, Naive Bayes, random forests, and single hidden layer feed-forward artificial neutral networks. On average, across 19 feature sets, random forest produced the best accuracy. Text-mining methods and supervised machine learning algorithms were also utilized in the study to validate accounts using Whiteprint, a biometric writing style. We examined the robustness and effectiveness of Twitter tweets by extracting features using the Stanford POS.

A method to locate spammers on Twitter was created by Meda et al. using a random forest-based classifier. To construct Twitter user profiles, feature extraction and parsing are used. After that, the classifier divides the sample into spammers and non-spammers. The run-time phase entails gathering JSON traffic using the Twitter streaming API, building user profiles using features that were retrieved, and categorizing the trial sample as spammer or non-spammer.

C. MISCELLANEOUS METHODS

Using both innovative and regularly used variables, Chen et al.'s investigation on the Twitter dataset finds content pollutants. These features, which include tweet-based and profile-based elements, are divided into direct and indirect categories.

The authors contend that while compromising time performance, indirect characteristics can increase detection rates. They use ROC curves to emphasize each feature's significance and recursive feature elimination (RFE) to choose the most reliable ones. The study demonstrates that the real-time spam detection accuracy of the random forest classifier is high. They also suggest a technique that combines text content removal with social network data to find spammers on Twitter.

The study examined 140 thousand user profiles and 284 million follower relationships using a dataset of 50 million tweets from Twitter in May 2011. From a total of 12,079 accounts, the researchers removed 10,450 legitimate users while separating 1,629 scammers. The technique created a single framework from text, social information networks, and supervised data. Based on two assumptions—the observation created by a hidden state and the state depending on the previous state—the study discovered that the Hidden Markov Model is good for detecting spam connected to recent time. The study looked at how size training data affected spam recognition.

The study examined 140 thousand user profiles and 284 million follower relationships using a dataset of 50 million tweets from Twitter in May 2011. From a total of 12,079 accounts, the researchers removed 10,450 legitimate users while separating 1,629 scammers. The technique created a single framework from text, social information networks, and supervised data. Based on two assumptions—the observation created by a hidden state and the state depending on the previous state—the study discovered that the Hidden Markov Model is good for detecting spam connected to recent time. The study looked at how size training data affected spam recognition.

IV. DISCUSSION

A taxonomy based on extraction and classification techniques was proposed after the study examined malevolent actions on social media. The taxonomy groups spam detection systems, URL-based approaches, popular themes, and methods for identifying bogus users. While URL-based algorithms can identify tweets with too many URLs, spammers frequently combine spam data with dangerous keywords. In order to detect spam on Twitter, strategies for identifying bogus users are introduced, along with hashtags or phrases that are likely to include spam. This taxonomy aids in preventing harmful actions taken against OSN users.

In this study, user, content, graph, structure, and time categories are used to compare Twitter spam detection features that are taken from user accounts and tweets. The amount of followers, account age, reputation, and number of tweets are all user-based characteristics. The features that are content-based include spam terms, URLs, replies, and retweets. In/out degree and betweenness centrality are properties of graphs. Average tweet length, thread life, tweet frequency, and conversion tree depth are examples of structure-based features. Idle time and tweets sent at predetermined intervals are two examples of time-based functionalities. The study also identifies machine learning-based methods for locating Twitter spam.

V. CONCLUSION AND FUTURE RESEARCH DIRECTIONS

The methods used to detect spam on Twitter are reviewed in this study and are divided into four categories: fake content identification, URL-based spam detection, spam detection in trending topics, and fake user detection methods. Based on user, content, graph, structure, time, and dataset-specific aspects, techniques are contrasted. The review intends to assist academics in discovering cutting-edge Twitter spam detection methods. Research is still needed, though, particularly in the areas of identifying false news and identifying the sources of rumors.

REFERENCES

[1] B. Erçahin, Ö. Akta³, D. Kilinç, and C. Akyol, ``Twitter fake account detection," in *Proc. Int. Conf. Comput. Sci. Eng. (UBMK)*, Oct. 2017, pp. 388_392.

[2] F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida, ``Detecting spammers on Twitter," in *Proc. Collaboration, Electron. Messaging, Anti- Abuse Spam Conf. (CEAS)*, vol. 6, Jul. 2010, p. 12.

[3] S. Gharge, and M. Chavan, ``An integrated approach for malicious tweets detection using NLP," in *Proc. Int. Conf. Inventive Commun. Comput. Technol. (ICICCT)*, Mar. 2017, pp. 435_438.

[4] T. Wu, S. Wen, Y. Xiang, and W. Zhou, ``Twitter spam detection: Survey of new approaches and comparative study," *Comput. Secur.*, vol. 76, pp. 265_284, Jul. 2018.

[5] S. J. Soman, ``A survey on behaviors exhibited by spammers in popular social media networks," in *Proc. Int. Conf. Circuit, Power Comput. Tech- nol. (ICCPCT)*, Mar. 2016, pp. 1_6.

[6] A. Gupta, H. Lamba, and P. Kumaraguru, ``1.00 per RT #BostonMarathon # prayforboston: Analyzing fake content on Twitter," in *Proc. eCrime Researchers Summit (eCRS)*, 2013, pp. 1_12.

[7] F. Concone, A. De Paola, G. Lo Re, and M. Morana, ``Twitter analysis for real-time malware discovery," in *Proc. AEIT Int. Annu. Conf.*, Sep. 2017, pp. 1_6.

[8] N. Eshraqi, M. Jalali, and M. H. Moattar, ``Detecting spam tweets in Twitter using a data stream clustering algorithm," in *Proc. Int. Congr. Technol., Commun. Knowl. (ICTCK)*, Nov. 2015, pp. 347_351.

[9] C. Chen, Y. Wang, J. Zhang, Y. Xiang, W. Zhou, and G. Min, ``Statistical features-based real-time detection of drifted Twitter spam," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 4, pp. 914_925, Apr. 2017.

[10] C. Buntain and J. Golbeck, ``Automatically identifying fake news in popular Twitter threads," in *Proc. IEEE Int. Conf. Smart Cloud (SmartCloud)*, Nov. 2017, pp. 208_215.

[11] C. Chen, J. Zhang, Y. Xie, Y. Xiang,W. Zhou, M. M. Hassan, A. AlElaiwi, and M. Alrubaian, ``A performance evaluation of machine learning-based streaming spam tweets detection," *IEEE Trans. Comput. Social Syst.*, vol. 2, no. 3, pp. 65_76, Sep. 2015.

[12] G. Stafford and L. L. Yu, ``An evaluation of the effect of spam on Twitter trending topics," in *Proc. Int. Conf. Social Comput.*, Sep. 2013, pp. 373_378.

[13] M. Mateen, M. A. Iqbal, M. Aleem, and M. A. Islam, ``A hybrid approach for spam detection for Twitter," in *Proc. 14th Int. Bhurban Conf. Appl. Sci. Technol. (IBCAST)*, Jan. 2017, pp. 466_471.

[14] A. Gupta and R. Kaushal, ``Improving spam detection in online social networks," in *Proc. Int. Conf. Cogn. Comput. Inf. Process. (CCIP)*, Mar. 2015, pp. 1_6.

[15] F. Fathaliani and M. Bouguessa, ``A model-based approach for identifying spammers in social networks," in *Proc. IEEE Int. Conf. Data Sci. Adv. Anal. (DSAA)*, Oct. 2015, pp. 1_9.

[16] V. Chauhan, A. Pilaniya, V. Middha, A. Gupta, U. Bana, B. R. Prasad, and S. Agarwal, ``Anomalous behavior detection in social networking," in *Proc. 8th Int. Conf. Comput., Commun. Netw. Technol. (ICCCNT)*, Jul. 2017, pp. 1_5.

[17] S. Jeong, G. Noh, H. Oh, and C.-K. Kim, ``Follow spam detection based on cascaded social information," *Inf. Sci.*, vol. 369, pp. 481_499, Nov. 2016. [18] M. Washha, A. Qaroush, and F. Sedes, ``Leveraging time for spammers detection on Twitter," in *Proc. 8th Int. Conf. Manage. Digit. EcoSyst.*, Nov. 2016, pp. 109_116.

[19] B. Wang, A. Zubiaga, M. Liakata, and R. Procter, ``Making the most of tweet-inherent features for social spam detection on Twitter," 2015, *arXiv:1503.07405*. [Online]. Available: https://arxiv.org/abs/1503.07405

[20] M. Hussain, M. Ahmed, H. A. Khattak, M. Imran, A. Khan, S. Din, A. Ahmad, G. Jeon, and A. G. Reddy, ``Towards ontology-based multilingual URL _ltering: A big data problem," *J. Supercomput.*, vol. 74, no. 10, pp. 5003_5021, Oct. 2018.

[21] C. Meda, E. Ragusa, C. Gianoglio, R. Zunino, A. Ottaviano, E. Scillia, and R. Surlinelli, ``Spam detection of Twitter traf_c: A framework based on random forests and non-uniform feature sampling," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, Aug. 2016, pp. 811_817.

[22] S. Ghosh, G. Korlam, and N. Ganguly, ``Spammers' networks within online social networks: A case-study on Twitter," in *Proc. 20th Int. Conf. Companion World Wide Web*, Mar. 2011, pp. 41_42.

[23] C. Chen, S. Wen, J. Zhang, Y. Xiang, J. Oliver, A. Alelaiwi, and M. M. Hassan, ``Investigating the deceptive information in Twitter spam," *Future Gener. Comput. Syst.*, vol. 72, pp. 319_326, Jul. 2017.

[24] I. David, O. S. Siordia, and D. Moctezuma, ``Features combination for the detection of malicious Twitter accounts," in *Proc. IEEE Int. Autumn Meeting Power, Electron. Comput. (ROPEC)*, Nov. 2016, pp. 1_6.

[25] M. Babcock, R. A. V. Cox, and S. Kumar, ``Diffusion of pro- and anti-false information tweets: The black panther movie case," *Comput. Math. Org. Theory*, vol. 25, no. 1, pp. 72_84, Mar. 2019.

[26] S. Keretna, A. Hossny, and D. Creighton, ``Recognising user identity in Twitter social networks via text mining," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Oct. 2013, pp. 3079_3082.

[27] C. Meda, F. Bisio, P. Gastaldo, and R. Zunino, ``A machine learning approach for Twitter spammers detection," in *Proc. Int. Carnahan Conf. Secur. Technol. (ICCST)*, Oct. 2014, pp. 1_6.

[28] W. Chen, C. K. Yeo, C. T. Lau, and B. S. Lee, ``Real-time Twitter content polluter detection based on direct features," in *Proc. 2nd Int. Conf. Inf. Sci. Secur. (ICISS)*, Dec. 2015, pp. 1_4.

[29] H. Shen and X. Liu, ``Detecting spammers on Twitter based on content and social interaction," in *Proc. Int. Conf. Netw. Inf. Syst. Comput.*, pp. 413_417, Jan. 2015.

[30] G. Jain, M. Sharma, and B. Agarwal, ``Spam detection in social media using convolutional and long short term memory neural network," *Ann. Math. Artif. Intell.*, vol. 85, no. 1, pp. 21_44, Jan. 2019.

[31] M. Washha, A. Qaroush, M. Mezghani, and F. Sedes, ``A topic-based hidden Markov model for real-time spam tweets _ltering," *Procedia Comput. Sci.*, vol. 112, pp. 833_843, Jan. 2017.

[32] F. Pierri and S. Ceri, ``False news on social media: A data-driven survey," 2019, *arXiv:1902.07539*. [Online]. Available: https://arxiv. org/abs/1902.07539

[33] S. Sadiq, Y. Yan, A. Taylor, M.-L. Shyu, S.-C. Chen, and D. Feaster, ``AAFA: Associative af_nity factor analysis for bot detection and stance classi_cation in Twitter," in *Proc. IEEE Int. Conf. Inf. Reuse Integr. (IRI)*, Aug. 2017, pp. 356_365.

[34] M. U. S. Khan, M. Ali, A. Abbas, S. U. Khan, and A. Y. Zomaya, ``Segregating spammers and unsolicited bloggers from genuine experts on Twitter," *IEEE Trans. Dependable Secure Comput.*, vol. 15, no. 4, pp. 551_560, Jul./Aug. 2018.