Received April 22, 2019, accepted May 17, 2019, date of publication May 22, 2019, date of current version June 6, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2918196

Spammer Detection and Fake User Identification on Social Networks

FAIZA MASOOD[1], GHANA AMMAD[1], AHMAD ALMOGREN[2], (Senior Member, IEEE), ASSAD ABBAS[1], HASAN ALI KHATTAK[1], (Senior Member, IEEE), IKRAM UD DIN[3], (Senior Member, IEEE), MOHSEN GUIZANI[4], (Fellow, IEEE), AND MANSOUR ZUAIR[5]

[1]Department of Computer Science, COMSATS University Islamabad, Islamabad 44550, Pakistan
[2]Chair of Cyber Security, Department of Computer Science, College of Computer and Information Sciences, King Saud University, Riyadh 11633, Saudi Arabia
[3]Department of Information Technology, The University of Haripur, Haripur 22620, Pakistan
[4]Computer Science and Engineering Department, Qatar University, Doha 2713, Qatar
[5]Department of Computer Engineering, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia

Corresponding authors: Ahmad Almogren (ahalmogren@ksu.edu.sa) and Assad Abbas (assadabbas@comsats.edu.pk)

The authors are grateful to the Deanship of Scientific Research, King Saud University for funding through the Vice Deanship of Scientific Research Chair.

ABSTRACT Social networking sites engage millions of users around the world. The users' interactions with these social sites, such as Twitter and Facebook have a tremendous impact and occasionally undesirable repercussions for daily life. The prominent social networking sites have turned into a target platform for the spammers to disperse a huge amount of irrelevant and deleterious information. Twitter, for example, has become one of the most extravagantly used platforms of all times and therefore allows an unreasonable amount of spam. Fake users send undesired tweets to users to promote services or websites that not only affect legitimate users but also disrupt resource consumption. Moreover, the possibility of expanding invalid information to users through fake identities has increased that results in the unrolling of harmful content. Recently, the detection of spammers and identification of fake users on Twitter has become a common area of research in contemporary online social Networks (OSNs). In this paper, we perform a review of techniques used for detecting spammers on Twitter. Moreover, a taxonomy of the Twitter spam detection approaches is presented that classifies the techniques based on their ability to detect: (i) fake content, (ii) spam based on URL, (iii) spam in trending topics, and (iv) fake users. The presented techniques are also compared based on various features, such as user features, content features, graph features, structure features, and time features. We are hopeful that the presented study will be a useful resource for researchers to find the highlights of recent developments in Twitter spam detection on a single platform.

INDEX TERMS Classification, fake user detection, online social network, spammer's identification.

I. INTRODUCTION

It has become quite unpretentious to obtain any kind of information from any source across the world by using the Internet. The increased demand of social sites permits users to collect abundant amount of information and data about users. Huge volumes of data available on these sites also draw the attention of fake users [1]. Twitter has rapidly become an online source for acquiring real-time information about users. Twitter is an Online Social Network (OSN) where users can share anything and everything, such as news, opinions, and even their moods. Several arguments can be held over different topics, such as politics, current affairs, and important events. When a user tweets something, it is instantly conveyed to his/her followers, allowing them to outspread the received information at a much broader level [2]. With the evolution of OSNs, the need to study and analyze users' behaviors in online social platforms has intensified. Many people who do not have much information regarding the OSNs can easily be tricked by the fraudsters. There is also a demand to combat and place a control on the people who use OSNs only for advertisements and thus spam other people's accounts.

Recently, the detection of spam in social networking sites attracted the attention of researchers. Spam detection is

The associate editor coordinating the review of this manuscript and approving it for publication was Tomohiko Taniguchi.

a difficult task in maintaining the security of social networks. It is essential to recognize spams in the OSN sites to save users from various kinds of malicious attacks and to preserve their security and privacy. These hazardous maneuvers adopted by spammers cause massive destruction of the community in the real world. Twitter spammers have various objectives, such as spreading invalid information, fake news, rumors, and spontaneous messages. Spammers achieve their malicious objectives through advertisements and several other means where they support different mailing lists and subsequently dispatch spam messages randomly to broadcast their interests. These activities cause disturbance to the original users who are known as non-spammers. In addition, it also decreases the repute of the OSN platforms. Therefore, it is essential to design a scheme to spot spammers so that corrective efforts can be taken to counter their malicious activities [3].

Several research works have been carried out in the domain of Twitter spam detection. To encompass the existing state-of-the-art, a few surveys have also been carried out on fake user identification from Twitter. Tingmin *et al.* [4] provide a survey of new methods and techniques to identify Twitter spam detection. The above survey presents a comparative study of the current approaches. On the other hand, the authors in [5] conducted a survey on different behaviors exhibited by spammers on Twitter social network. The study also provides a literature review that recognizes the existence of spammers on Twitter social network. Despite all the existing studies, there is still a gap in the existing literature. Therefore, to bridge the gap, we review state-of-the-art in the spammer detection and fake user identification on Twitter. Moreover, this survey presents a taxonomy of the Twitter spam detection approaches and attempts to offer a detailed description of recent developments in the domain.

The aim of this paper is to identify different approaches of spam detection on Twitter and to present a taxonomy by classifying these approaches into several categories. For classification, we have identified four means of reporting spammers that can be helpful in identifying fake identities of users. Spammers can be identified based on: (i) fake content, (ii) URL based spam detection, (iii) detecting spam in trending topics, and (iv) fake user identification. Table 1 provides a comparison of existing techniques and helps users to recognize the significance and effectiveness of the proposed methodologies in addition to providing a comparison of their goals and results. Table 2 compares different features that are used for identifying spam on Twitter. We anticipate that this survey will help readers find diverse information on spammer detection techniques at a single point.

This article is structured such that Section II presents the taxonomy for the spammer detection techniques on Twitter. The comparison of proposed methods for detecting spammers on Twitter is discussed in Section III. Section IV presents an overall analysis and discussion, whereas Section V concludes the paper and highlights some directions for future work.

II. SPAMMER DETECTION ON TWITTER

In this article, we elaborate a classification of spammer detection techniques. Fig. 1 shows the proposed taxonomy for identification of spammers on Twitter. The proposed taxonomy is categorized into four main classes, namely, (i) fake content, (ii) URL based spam detection, (iii) detecting spam in trending topics, and (iv) fake user identification. Each category of identification methods relies on a specific model, technique, and detection algorithm. The first category (fake content) includes various techniques, such as regression prediction model, malware alerting system, and Lfun scheme approach. In the second category (URL based spam detection), the spammer is identified in URL through different machine learning algorithms. The third category (spam in trending topics) is identified through Naïve Bayes classifier and language model divergence. The last category (fake user identification) is based on detecting fake users through hybrid techniques. Techniques related to each of the spammer identification categories are discussed in the following subsections.

A. FAKE CONTENT BASED SPAMMER DETECTION

Gupta *et al.* [6] performed an in-depth characterization of the components that are affected by the rapidly growing malicious content. It was observed that a large number of people with high social profiles were responsible for circulating fake news. To recognize the fake accounts, the authors selected the accounts that were built immediately after the Boston blast and were later banned by Twitter due to violation of terms and conditions. About 7.9 million distinctive tweets were collected by 3.7 million distinctive users. This dataset is known as the largest dataset of Boston blast. The authors performed the fake content categorization through temporal analysis where temporal distribution of tweets is calculated based on the number of tweets posted per hour.

Fake tweet user accounts were analyzed by the activities performed by user accounts from where the spam tweets were generated. It was observed that most of the fake tweets were shared by people with followers. Subsequently, the sources of tweet analysis were analyzed by the medium from where the tweets were posted. It was found that most of the tweets containing any information were generated through mobile devices and non-informative tweets were generated more through the Web interfaces. The role of user attributes in the identification of fake content was calculated through: (i) the average number of verified accounts that were either spam or non-spam and (ii) the number of followers of the user accounts. The fake content propagation was identified through the metrics that include: (i) social reputation, (ii) global engagement, (iii) topic engagement, (iv) likability, and (v) credibility. After that, the authors utilized regression prediction model to ensure the overall impact of people who spread the fake content at that time and also to predict the fake content growth in future.

Concone *et al.* [7] presented a methodology that provides malignant alerting by using a specified set of tweets in real-time conquered through the Twitter API. Afterwards,

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

Ref.	Proposed Method	Goal	Data Set	Result
[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 behavior in user's interaction	Twitter dataset is used, which contains last 200 tweets of users	Anomalous detection model can be used to analyze effectively the number of URL spammer 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 around 1.8 billion.	Classification of spammer uses a large set of attributes and is significantly more robust to spammers, which familiarize spamming schemes.
[17]	Three types of cascade information, which are created on the basis of spam detection mechanism, have been used, i.e., TSP, SS, and cascade filtering.	Spammers have been classified by using the properties of social networks in the individual social environment.	Real Twitter dataset.	The schemes are scalable because they check users cantered 2-hops social 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 behavior and tweet content have been analyzed 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 sets were utilized for exercising the spam and non-spammer classifiers.	The objective of the study is to 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 existing spammer detection systems.

TABLE 1. *(Continued.)* **Comparison between proposed methods for spam detection in Twitter.**

[21]	Two experiments were conducted, i.e., edge weighting and centrality weighting	To understand the significance of each well-defined edge in order to find the opinion leader and to perceive the weight that could permit more precised opinion based on evaluation algorithms.		Indicates that the low in-degree weight, high betweenness weight, and low or no PageRank weight could provide 100% agreement as compared to other evaluation algorithms in order to find the opinion leader.
[9]	Performance of a comprehensive range of conventional machine learning algorithms for the purpose of identifying the performance of detection and strength based on immense amount of truth data.	The goal of the study is to attain real time Twitter spam detection capabilities.	Around 30 million labelled tweets were randomly selected to form the ground truth data set	The Lfun scheme can enhance the precision of spam detection significantly in the real-world context.
[1]	Entropy minimization discretization (EMD) technique was used on numerical features	To detect fake accounts on Twitter by proposing classification methods and to illustrate the effect of discretization on the basis of Naïve Bayes algorithm in Twitter.	No public dataset is available, thus, created own dataset based on Twitter API	Naive Bayes can perform well with discrete values as compared to continuous vales.
[13]	A hybrid Technique has been used for the identification of spammer on Twitter by utilizing user based, content based, and graph-based features.	Achieve higher accuracy by combining user based, content based, and graph-based features for spam profile detection	Dataset of Twitter with 11k users and approximately 400k tweets were used.	The rate of detection in the study is more accurate and higher as compared to any existing technique.
[6]	Regression prediction model has been used in order to prove the influence of users who spread fake content.	To classify and recommend solutions to counter different forms of spam events on Twitter during activities like Boston Blast.	About 7.8 million Boston marathon blast related tweets extracted using Twitter API.	Approximately 29% content, which are more viral on Twitter during the crisis of Boston blast, were fake. Whereas 51% were general views and comments, and the remaining were correct Information.

the batch of tweets considering the same topic is sum up to generate an alert. The proposed architecture is used to evaluate Twitter posting, recognizing the advancement of admissible event, and reporting of that event itself. The proposed approach utilizes the information contained in the tweets when a spam or malware is recognized by the users or the report of security has been released by the certified authorities. The proposed alerting system comprises of the following components: (i) real time data extraction of both tweets and users, (ii) filtering system based on a pre-processing schedule and on Naïve Bayes algorithm to discard the tweets containing inaccurate information, (iii) data analysis for spammer detection where the detection windows are rigorously abolished according to the Sigmoid function

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

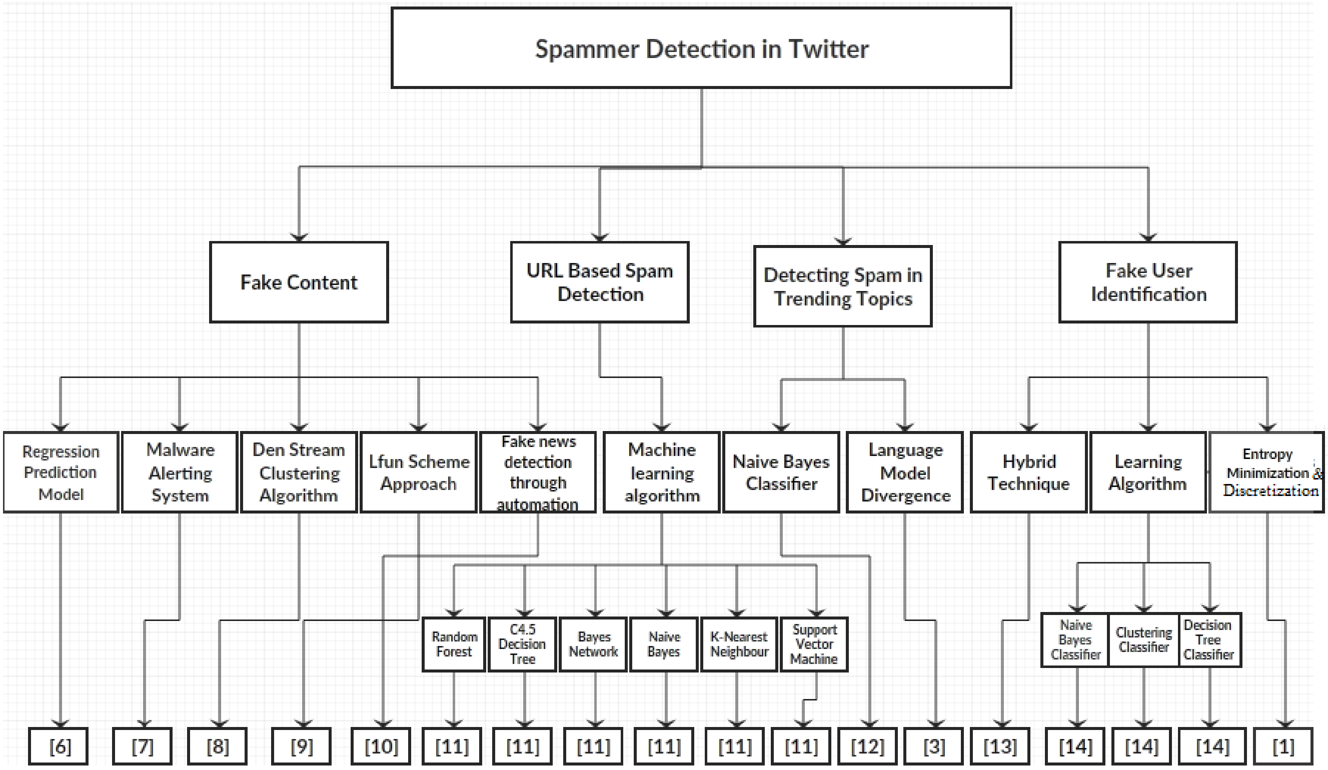

FIGURE 1. Taxonomy of spammer detection/fake user identification on Twitter.

or when the window size reaches the maximum, (iv) alert subsystem that is used when the event is established, the system groups up the tweets that are relevant to the same topic where tweets are distinguished with the cluster barycenter and the one that is nearest to the cluster center is chosen as the representative of the whole system cluster, and (v) feedback analysis. The approach is claimed to be efficient and effective for the detection of some invasive and admirable malignant activities in circulation.

Moreover, Eshraqi *et al.* [8] determined different features to detect the spam and then with the help of a den stream-based clustering algorithm, recognize the spam tweets. Some user accounts were selected from various datasets and afterwards random tweets were selected from these accounts. The tweets are subsequently categorized as spam and non-spam. The authors claimed that the algorithm can divide the data into spam and non-spam with high accuracy and fake tweets maybe recognized with high accuracy and precision.

Various features can be used to determine the spams. For example, feature based on the graph is a state in which Twitter is shaped as a social model of a graph. If the number of followers is low in comparison with the number of followings, the credibility of an account is low and the possibility that the account is spam is relatively high. Likewise, feature based on content includes tweets reputation, HTTP links, mentions and replies, and trending topics. For the time feature, if many tweets are sent by a user account in a certain time interval, then it is a spam account. The dataset of the study comprised 50,000 user accounts. The approach identified the spammers and fake tweets with high accuracy.

A Lfun (learning for unlabeled tweets) scheme, which is used to handle various problems in the detection of Twitter spam, has been presented by Chen *et al.* [9]. Their framework comprises two components, i.e., learn from detected tweets (LDT) and learn from human labelling (LHL). The two components are used to automatically generate spam tweets from the given set of unmarked tweets that are easily collected from the Twitter network side. Once the labelled spam tweets are obtained, random forest algorithm is used to perform classification. The performance of the scheme is evaluated while detecting drifted spam tweets. The experiments were performed on the real-world data of ten continuous days with each day having 100K tweets each for the spam and non-spam. The F-measure and the detection rate were used to evaluate the performance of the presented scheme. The results of the proposed approach showed that the methodology improves the accuracy of spam detection significantly in the real-world situations.

Furthermore, Buntain *et al.* [10] introduced a method for detecting fake news on Twitter automatically by predicting accurate assessment in two credibility-focused datasets. The method was applied on the Twitter fake news dataset and the model was trained against a crowd sourced worker based on the assessment of journalists. The two Twitter datasets were used to study the integrity in OSNs. The first dataset CREDBANK, a crowd-sourced dataset, was used to evaluate the accuracy of events in Twitter whereas the second dataset called PHEME is a journalist-labelled dataset of possible rumors in Twitter and journalistic evaluation of their accuracy. A total of 45 features were described that fall into four categories: structural feature, user feature, content feature, and temporal features. Aligning labels in PHEME and BUZZFEED contain classes that describe whether a story is fake or true. Results of the analysis are helpful in studying information on social media to know whether such stories support similar pattern.

B. URL BASED SPAM DETECTION

Chen *et al.* [11] performed an evaluation of machine learning algorithms to detect spam tweets. The authors analyzed the impact of various features on the performance of spam detection, for example: (i) spam to non-spam ratio, (ii) size of training dataset, (iii) time related data, (iv) factor discretization, and (v) sampling of data. To evaluate the detection, first, around 600 million public tweets were collected and subsequently the authors applied the *Trend micro's web reputation system* to identify spam tweets as much as possible. A total of 12 lightweight features were also separated to distinguish non-spam and spam tweets from this identified dataset. The characteristics of identified features were represented by cdf figures.

These features are grasped to machine learning based spam classification, which are later used in the experiment to evaluate the detection of spam. Four datasets are sampled to reproduce different scenarios. Since no dataset is available publicly for the task, few datasets were used in previous researches. After the identification of spam tweets, 12 features were gathered. These features are divided into two classes, i.e., user-based features and tweet-based features. The user-based features are identified through various objects such as account age and number of user favorites, lists, and tweets. The identified user-based features are parsed from the JSON structure. On the other hand, the tweet-based features include the number of (i) retweets, (ii) hashtags, (iii) user mentions, and (iv) URLs. The result of evaluation shows that the changing feature distribution reduced the performance whereas no differences were observed in the training dataset distribution.

C. DETECTING SPAM IN TRENDING TOPIC

Gharge *et al.* [3] initiate a method, which is classified on the basis of two new aspects. The first one is the recognition of spam tweets without any prior information about the users and the second one is the exploration of language for spam detection on Twitter trending topic at that time. The system framework includes the following five steps.

- The collection of tweets with respect to trending topics on Twitter. After storing the tweets in a particular file format, the tweets are subsequently analyzed.
- Labelling of spam is performed to check through all datasets that are available to detect the malignant URL.
- Feature extraction separates the characteristics construct based on the language model that uses language as a tool and helps in determining whether the tweets are fake or not.
- The classification of data set is performed by shortlisting the set of tweets that is described by the set of features provided to the classifier to instruct the model and to acquire the knowledge for spam detection.
- The spam detection uses the classification technique to accept tweets as the input and classify the spam and non-spam.

The experimental setup was prepared for determining the accuracy of the system. For this purpose, a random sample set of 1,000 tweets was collected from which 60% were legal and the rest were defected.

Stafford *et al.* [12] examined the degree to which the trending affairs in Twitter are exploited by spammers. Although numerous methods to detect the spam have been proposed, the research on determining the effects of spam on Twitter trending topics has attained only limited attention of the researchers. The authors in [12] presented a technique to cooperate with Twitter public API. The aim of the implemented program was to find 10 trending topics from all over the world having a language code within one hour and open the filtered connection related to those topics to acquire a data stream. In the next hour, the authors obtained as much of the tweets and linked metadata as permitted by the Twitter API. Once the data has been collected, the collected tweets were classified into two categories, i.e., spam and non-spam tweets, which can be utilized to instruct classifiers.

To develop such a collection of manual labelling, another program was suggested to sample random tweets, where the idea is based upon URL filtering by Hussain *et al.* [20]. After the completion of labelling tweets, they move toward the next phase of analysis method. Analysis method has two separate phases, where the first phase was to select and evaluate the attribute through information retrieval metrics, while the second phase was to evaluate the effect of spam filtering on the trending topics through statistical test. The result of the evaluation concludes that spammer does not acquire the trending topic in Twitter but alternatively adopts target topics with required qualities. The results signify well for the sustainability of the Twitter and provide a way for improvement.

D. FAKE USER IDENTIFICATION

A categorization method is proposed by Erşahin *et al.* [1] to detect spam accounts on Twitter. The dataset used in the study was collected manually. The classification is performed by analyzing user-name, profile and background image, number of friends and followers, content of tweets, description of account, and number of tweets. The dataset comprised 501 fake and 499 real accounts, where 16 features from the information that were obtained from the Twitter APIs were identified. Two experiments were performed for classifying fake accounts. The first experiment uses the Naïve Bayes learning algorithm on the Twitter dataset including all aspects without discretization, whereas the second experiment uses the Naïve Bayes learning algorithm on the Twitter dataset after the discretization.

Mateen *et al.* [13] proposed a hybrid technique that utilizes user-based, content-based, and graph-based characteristics for spammer profiles detection. A model is proposed to differentiate between the non-spam and spam profiles using three characteristics. The proposed technique was analyzed using Twitter dataset with 11K users and approximately 400K tweets. The goal is to attain higher efficiency and preciseness by integrating all these characteristics. User-based features are established because of relationship and properties of user accounts. It is essential to append user-based features for the spam detection model. As these features are related to user accounts, all attributes, which were linked to user accounts, were identified. These attributes include the number of followers and following, age, FF ratio, and reputation. Alternatively, content features are linked to the tweets that are posted by users as spam bots that post a huge amount of duplicate contents as contrast to non-spammers who do not post duplicate tweets.

These features depend on messages or content that users write. Spammers post contents to spread fake news and these contents contain malicious URL to promote their product. The content-based features include: (i) the total number of tweets, (ii) hashtag ratio, (iii) URLs ratio, (iv) mentions ratio, and (v) frequencyof tweets. The graph-based feature is used to control the evasion strategies that are conducted by spammers. Spammers use different techniques to avoid being detected. They can buy fake followers from different third-party websites and exchange their followers to another user to look like a legal user. Graph-based features include in/out degree and betweenness. The evaluation of the approach is done by using the dataset of previous techniques as, due to the Twitter policy, no data is available publicly. The results are evaluated by integrating three most common approaches, namely Decorate, Naïve Bayes, and J48. The result of the experiment shows that the detection rate of the approach is much accurate and higher than any of the existing techniques.

Gupta *et al.* [14] present a policy for the detection of spammers in Twitter and use the popular techniques, i.e., Naïve Bayes, clustering, and decision trees. The algorithms classify an account as spam or non-spam. The dataset comprises 1064 Twitter users that contain 62 features, which are either user-specific or tweet-specific information. The spammer account contains almost 36% of the used dataset. As the behavior of spammers is different from non-spammers, some attributes or features are recognized in which both categories are different from one another. Feature identification is based on the number of features at user and tweet level such as followers or following, spam keywords, replies, hashtags, and URLs [30], [32].

After the identification of features, pre-processor step transforms all continuous features into discrete. Subsequently, the authors developed a technique using clustering, decision trees, and Naïve Bayes algorithms. With Naïve Bayes, the accounts were identified by estimating the possibility of certain account as the spammer or non-spammer. In clustering-based algorithm, the entire set of accounts is classified into two classes as the spam and non-spam.

In decision tree algorithm, structure of tree was designed, and the decisions were made at every level of the tree. The result of the proposed approach shows that the clustering algorithm's performance to detect the non-spam accounts is better as compared to detection of spam accounts. Results of these integrated algorithm demonstrate the overall accuracy and detection of non-spammer with high effectiveness.

III. COMPARISON OF APPROACHES FOR SPAM DETECTION ON TWITTER

This section provides the comparison of proposed methodology along with their goals, datasets that are used to analyze spams, and results of the experiments of each method, as shown in Table 1.

A. ANOMALY DETECTION BASED ON URL

Chauhan *et al.* [16] proposed a methodology for the detection of anomalous tweets. The type of abnormality that is distributed on Twitter is the type of URL anomaly. Anomalous users use various URL links for creating spams. The proposed methodology, which is used to identify various anomalous activities from social networking sites, for example, Twitter, comprises the following features.

- URL ranking in which the URL rank is identified such that how authentic a URL is.
- Similarity of tweets includes posting of same tweets again and again.
- Time difference between tweets involves posting of five or more tweets during the time period of one minute.
- Malware content consists of malware URL that can damage the system.
- Adult content contains posts that consist of adult content.

For analyzing the anomalous behavior of Twitter based on URL, the dataset is prepared by accumulating 200 tweets of a user.

The dataset is expanded in order to enlarge the size. Five functions are executed on Twitter dataset, which are given below:

- URL rank generation is used to get the URL that a user has used in a tweet. This URL is sent to the website of ALEXA where the source code is obtained and the tree is generated by the help of web scraper from the given source code.
- Tweet similarity in this generation evaluates full tweets instead of analyzing only URL.
- Malware URL rank assignment is used to get the URL from a user that s/he has shared in his/her tweet. The WebOfTrust (WOT) API is used to check the repute of the URL that whether it is a good URL or contains some malware.
- Time difference calculation checks all the tweets with its previous three tweets and the next three tweets, and forms the cluster of seven tweets.
- Adult content identification is used to construct a dataset of all URLs that may contain adult content.

The results ensure that the proposed anomalous detection model can be used to analyze the number of Ueffectively RL spammers.

Moreover, Ghosh *et al.* [22] evaluate the scenarios engaged by new spammers in OSNs by recognizing a spam account in Twitter and controlling their link-creation plans. The analysis of the approach shows that the spammers support intelligent scenarios for the formation of link to evade the detection and to raise the capacity of their spam that are generated. The dataset of eight spam accounts in Twitter was used to detect other doubtful user accounts. It is testified that the spammers on Twitter frequently post tweets that contain URLs of their associated websites, therefore, frequently used URLs are utilized to recognize malignant users. The experiment shows that the spammer not only follows other spammers but also points out legal users who generally follow back. On the other hand, a spammer controls the followers of the spotted legal users and starts to follow them for following these spotted users. Spotted users hope that they can be followed back. This is how spammers identify other spammers and coordinate with them.

The following observations are considered while performing this experimental study:

- A total of 4491 spam accounts, which have around 730,000 links that are directed among them, ensure the presence of huge spam firm with the density of 0.036. It is also reported that spam accounts can easily find other spam accounts within an OSN having the size of Twitter.
- It is estimated that 4.74% of the follow links on average are developed by these spammers and this amount of fraction is as greater as 12% for some of the other accounts.
- It shows that spammers having greater number of following have greater reciprocal on an average. It also shows that more of the spammers' time is spend in the network to create more and more links so that they can filter out more users who can follow them back.
- A huge flap exists on the side of spammers, which implies a large-scale participation among various spammers for recognizing emergent users to follow.

Thus, the result of the analysis recognizes the evidence that is left by the large spam firms within OSNs and provides various insights on the creation of link scenarios of the spammers that needs to be studied while creating anti-spam scenarios.

Furthermore, a study of ambiguous information in Twitter spam has been presented by Chen *et al.* [23]. A complete Twitter feed of two weeks with URLs is collected. A lot of spam tweets, which were analyzed during the research, only a new tweet without URLs is considered as spam. Additionally, spammers primarily use encapsulated URLs for creating it more acceptable for the victims to their independent sides to accomplish their objectives such as scams, downloading malware, and phishing. Two steps were applied to recognize the spam in Twitter. The first one is using Trend Micro's WRT where the false positive rate of WRT is relatively low with a likelihood of missing few spam tweets. In addition, a goal of the research is to achieve high level of understanding on the variety of ambiguous topics that are used in the Twitter spam. The second step involves clustering approach with two folds: a) the clustering approach uncategorizing non-spam and spam tweets into various groups. b) Analyzing spam groups would be more helpful.

The graphical clustering approach is used by bipartite Cliques rather than machine learning algorithm for the grouping of spam tweets. These ambiguous topics are categorized into four groups that include malware, phishing, Twitter follower scam, and advertising. All these groups are organized and developed according to the contrasting deceptive information available in spam groups. The findings of this approach are helpful for the advancement of spam detection policies. Almost 400 million tweets are posted daily in which only 25% include URLs to investigate such a huge number of tweets where removing spams is relatively very expensive to implement in the real world. The result of the analysis shows that the features used in this work face various challenges, i.e., some features are simple to be deceived while others are difficult to be extracted.

B. MACHINE LEARNING ALGORITHMS

Benevenuto *et al.* [2] examined the problem of spammer detection on Twitter. For this, a large dataset of Twitter is collected that contains more than 5400 million users, 1.8 billion tweets, and 1.9 billion links. After that, the number of features, which are associated with tweet content, and the characteristics of users are recognized for the detection of spammers. These features are considered as the characteristics of machine learning process for categorizing users, i.e., to know whether they are spammers or not. In order to recognize the approach for detecting spammers on Twitter, the labelled collection in pre-classification of spammer and non-spammers has been done. Crawling Twitter has been launched to gather the IDs of users, which are about 80 million. Twitter allocates a numeric ID to each user which distinctively identifies the profile of each user. Next, those steps are taken which are needed for the construction of labelled collection and acquired various desired properties. In other words, steps which are essential to be examined to develop the collection of users that can be labelled as spammers or non-spammers. At the end, user attributes are identified based on their behavior, e.g., who they interact with and what is the frequency of their interaction.

In order to confirm this instinct, features of users of the labelled collection has been checked. Two attribute sets are considered, i.e., content attributes and user behavior attributes, to differentiate one user from the other. Content attributes have the property of the wordings of tweets that are posted by the users which gather features that are relevant to the way users write tweets. On the other hand, user behavior attributes gather particular features of the behavior of users in the context of the posting frequency, interaction, and impact on Twitter. The following attributes are considered as user characteristics, which include the total number of followers and following, account age, number of tags, fraction of followers per followings, number of times users replied, number of tweets received, average, maximum, minimum, and median time among user tweets, and daily and weekly tweets. Overall 23 attributes of the user behavior have been considered. The result of the proposed methodology shows that even with the distinguished set of attributes, the framework is able for detecting spammers with high frequency.

Jeong *et al.* [17] analyzed the follow spam on Twitter as an alternative of dispersion of provoking public messages, spammers follow authorized users, and followed by authorized users. Categorization techniques were proposed that are used for the detection of follow spammers. The focus of the social relation is cascaded and formulated into two mechanism, i.e., social status filtering and trade significance profile filtering, where each of which uses two-hop subnetworks that are centered at each other. Assemble techniques and cascading filtering are also proposed for combining the properties of both trade significance profile and social status. To check whether a user is fake or not, a two-hop social network for each user is focused to gather social information from social networks.

The experiment with the real-world data was performed to check the credibility and reliability of Twitter system with positive results. Both TSP and SS filtering were proposed by using partial data for real time and lightweight spammer detection. Both algorithms contain some false positive, but their true positive are not better to collusion rank. A hybrid approach that uses attributes of both filtering are suggested. The experiment was performed on thousand authorized users and thousand spammer accounts with social status and TSP features. The result of the proposed approach shows that the schemes are scalable because they check user centered two-hops social network instead of examining the whole network. This study significantly improves the performance of false and true positives than the previous scheme.

Meda *et al.* [21] presented a technique that utilizes a sampling of non-uniform features inside a machine learning system by the adaptation of random forest algorithm to recognize spammer insiders. The proposed framework focuses on the random forest and non-uniform feature sampling techniques. The random forest is a learning algorithm for the categorization and regression that works by assembling several decision trees at preparation time and selecting the one with the majority votes by individual trees. The scheme integrates bootstrap aggregating technique with the un-planned selection of features.

Non-uniform feature sampling method is used to obtain upper bound of the random forest error generalization. The dataset was prepared by the authors with an aim to gather users with indefinite behaviors for the purpose of testing the performance of random forest algorithm in the reference where the user categorization is undetermined. The choice of features is divided into two sub-categories, i.e., random selection and domain expert selection. Two datasets were used to show the efficiency of the non-feature sampling technique.

The first dataset is constructed having 1,065 users, wherein 355 are labelled as spammer and 710 as non-spammers outlined by 62 features. The second dataset has been constructed by the author. The goal of experiments is to reproduce two opposite situations at the time of feature selection phase. The first group of experiment involves domain experts for the feature choice and the other group utilizes a random selection of features. The results of the experiments reveal the potency of enriched feature sampling technique.

David *et al.* [24] presented an approach for identifying fake user's identity from the Twitter platform. User profiles and timelines were used to produce a feature set of 71 low cost variables. These variables divide timeline-based features into content-based and metadata-based features. Metadata based features refer to all the information that support or define the main content. Feature engineering includes various steps to explore that the data offers short re-collection of some modifications that were recognized while supporting decision trees with the totaled features. The variable importance is used for finding the best feature combination from the feature set effectively and efficiently.

All the feature sets were ranked according to four different measures. The selections by validated classification are used to gain the accuracy of five supervised classifiers, which include decision trees, support vector machine, Naïve Bayes, random forest, and single hidden layer feed forward artificial neutral networks. Results of the proposed approach demonstrate that the highest accuracy on average was acquired with random forest working on 19 feature set. It also confirms that the largely effective detection and the devices that are feasible can be developed by the problem at hand.

Moreover, a study by Keretna *et al.* [26] focuses on verifying actual accounts and fake accounts by the help of using Whiteprint, which is the biometric writing style. The feature sets were separated by using text-mining techniques and the knowledge based is trained using supervised machine learning algorithms. The recognition strategy for separating the characteristics was started and then the similarity of the characteristics vector was measured according to all characterized vectors present in the knowledge base. Subsequently, the most alike vector is recognized as a verified account. The sets of features that are similar to the problem are selected. The Stanford POS is used for extracting features. Using messages of Twitter as a case study, the features are separated according to the nature of Twitter and permitted pictures and videos only through the links that are external. Afterwards, the technique is applied to the number of accounts to investigate the resilience and efficiency of the proposed methodology.

Meda *et al.* [27] designed a technique to identify spammers on Twitter. In the proposed framework the training part is done offline which focuses at the establishment of random forest-based classifier that starts from the set of initial training sets. In the process of feature extraction, the result is parsed to attain a Twitter user profile. As machine learning techniques require to work on numerical features, there is a need to transfer profiles into vectors to well match with the ML-Module. The feature extraction process separates and converts selected features into real numbers to classify between spammers and non-spammers.

The classifier is instructed with the sample acquired from the previous step. Once the classifier is trained, its parameters are fixed and the system is engaged in the run time Twitter messages classification. The run time phase has the following basic steps: (a) Twitter streaming API is used to gather Twitter traffic that reoccurs Twitter reports in the format of JSON, (b) the profiles of Twitter users are constructed based on the features extracted from Twitter reports, (c) the classifier allocates a category as spammer or non-spammer to the trial sample. The results of the study show the effectiveness of the proposed method in comparison to various other models.

C. MISCELLANEOUS METHODS

Chen *et al.* [28] conducted a study on large-scale Twitter dataset and presented an explanation of content polluter. Some novel features are also proposed and combined with other frequently used features to detect the spam. The features were categorized into two classes, namely direct and indirect features. Direct features, which can be obtained from the unprocessed JSON tweets, are further categorized into tweet-based and profile-based features. The indirect features cannot be extricated from the unprocessed JSON tweets such as history of tweets, social relationship, etc.

According to the observation, the indirect features can assist to enhance the rate of detection with the surrender of time performance. The authors identified superior features from the time and accuracy perspective. The location under the ROC curve is employed to illustrate the significance of every individual feature. Moreover, feature selection via recursive feature elimination (RFE) is used to select robust features. The key concept of the RFE is to frequently construct models to abolish the worst or best features. The process is iterated until the entire feature set is visited. The most important features include account age, friends count, retweet count, hashtag count, etc. The results of the study show that random forest classifier achieves high spam detection accuracy in real-time.

Shen *et al.* [29] investigated issues of detecting spammers on Twitter. The proposed method combines characteristics withdrawal from text content and information of social networks. The authors used matrix factorization to determine the underline feature matrix or the tweets and then came up with a social regularization with interaction coefficient to teach the factorization of the underline matrix. Subsequently, the authors combined knowledge with social regularization and factorization matrix processes, and performed experiments on the real-world Twitter dataset, i.e., UDI Twitter dataset.

The dataset that was used in this experiment was basically collected in May 2011 on Twitter which contains 50 million tweets in 140 thousand user profile and 284 million following relationships. The content of the tweets for all users were scanned manually. In the end, 1,629 spammers were separated and 10,450 legal users from 12,079 users in their dataset were extracted. To measure the efficiency of the proposed approach, a conventional assessment measures was used to detect the spammers. The method that is proposed un-seemed to incorporate the features that are obtained from the text, social information network, and supervised information into a single framework. The results of the study demonstrate the effectiveness of the spammer detection.

Washha *et al.* [31] described the Hidden Markov Model for filtering the spam related to recent time. The method supports the accessible and obtainable information in the tweet object to recognize spam tweets and the tweets that are handled previously related to the same topic. The proposed work was based on two various assumptions, which are given below.

- The observation that had been produced by some state St that is hidden from the spectator at given time *t*
- The state where the current state S_t is dependent on the previous state S_{t-1}

The authors explored the consequences of time dependent learning model, which is used for detecting spam tweets of current topics effectively. Moreover, the study also investigated the influence of size training data on the capability of spam detection. The authors claimed that the Hidden Markov Model is capable of detecting spam tweets more effectively as it is better solution to have high quality recent tweets. Table 2 provides the comparison of different techniques for spammer detection.

IV. DISCUSSION

From the survey, we analyzed that malicious activities on social media are being performed in several ways. Moreover, the researchers have attempted to identify spammers or unsolicited bloggers by proposing various solutions. Therefore, to combine all pertinent efforts, we proposed a taxonomy according to the extraction and classification methods. The categorization is based on various classifications such as fake content, URL based, trending topics, and by identifying fake users. The first major categorization in the taxonomy is of techniques proposed for detecting spam, which is injected in the Twitter platform through fake content. Spammers generally combine spam data with a topic or keywords that are malicious or contain the type of words that are likely to be spam. The second categorization considers the techniques for spam detection based on URLs.

Generally, because of the length-limit of tweet description, spammers find it more promising to post URL to spread malicious content than the plain normal text. Therefore, URL based methods are absolutely customized to determine tweets containing excess of URLs specifically on criminal accounts. The third category in the proposed taxonomy contains approaches meant for spam identification from trending topics on Twitter. Hashtag or keywords, which are often seen in tweets at a specific time, appear in the Twitter list of trending topics and are likely to contain spam. Various features for identifying spams in trending topics have been classified with a variety of attributes. The fourth category in the taxonomy is regarding techniques for the identification of fake users to detect spams on Twitter. An assortment of techniques has been introduced for detecting spams of fake users that helps to overcome malicious activities against OSN users.

In addition to reviewing the techniques, the study also provides the comparison of miscellaneous Twitter spam detection features. These features are extracted from user accounts and the tweets that can help to identify spams. These features are categorized into five classes, namely user, content, graph, structure, and time. The user-based features incorporate the number of following and followers, account age, reputation, FF ratio, and number of tweets. The content-based features contain number of retweets, number of URLs, number of replies and propagation of bidirectional, number of characters and digits, and spam words.

The graph-based features include in/out degree and betweenness centrality whereas the structure-based features include average tweet length, thread life time (number of times between first and last tweets), tweet frequency, and depth of conversion tree. On the other hand, time-based features include idle time in days and tweet sent in specific time interval. Therefore, the survey is assembled by the classes that are categorized according to different features that are used for analyzing and detecting Twitter spams in various groups. We further carried out a comparative study on the existing techniques and methods that mainly capture the detection of spams on Twitter social network. This study includes the comparison of various previous methodologies proposed using different datasets and with different characteristics and accomplishments.

Moreover, the analysis also shows that several machine learning-based techniques can be effective for identifying spams on Twitter. However, the selection of the most feasible techniques and methods is highly dependent on the available data. For example, Na ïve Bayes, random forest, bayes betwork, K-nearest neighbor, clustering, and decision tree algorithms are used for predicting and analyzing spams on Twitter with different classes of categorization. This comparative study helps to identify all spam detection techniques under one umbrella, as shown in Figure 1.

V. CONCLUSION AND FUTURE RESEARCH DIRECTIONS

In this paper, we performed a review of techniques used for detecting spammers on Twitter. In addition, we also presented a taxonomy of Twitter spam detection approaches and categorized them as fake content detection, URL based spam detection, spam detection in trending topics, and fake user detection techniques. We also compared the presented techniques based on several features, such as user features, content features, graph features, structure features, and time features. Moreover, the techniques were also compared in terms of their specified goals and datasets used. It is anticipated that the presented review will help researchers find the information on state-of-the-art Twitter spam detection techniques in a consolidated form.

Despite the development of efficient and effective approaches for the spam detection and fake user identification on Twitter [34], there are still certain open areas that require considerable attention by the researchers. The issues are briefly highlighted as under:

False news identification on social media networks is an issue that needs to be explored because of the serious repercussions of such news at individual as well as collective level [25]. Another associated topic that is worth investigating is the identification of rumor sources on social media. Although a few studies based on statistical methods have already been conducted to detect the sources of rumors, more sophisticated approaches, e.g., social network-based approaches, can be applied because of their proven effectiveness.

REFERENCES

[1] B. Erçahin, Ö. Aktaş, 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. Technol. (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 filtering: 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 traffic: 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 filtering," *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 affinity factor analysis for bot detection and stance classification 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.

FAIZA MASOOD received the bachelor's degree in computer science from COMSATS University Islamabad, Islamabad, Pakistan, where she is currently pursuing the master's degree in software engineering. Her research interest focuses on the social networking sites.

GHANA AMMAD received the bachelor's degree in computer science from COMSATS University Islamabad, Islamabad, Pakistan, where she is currently pursuing the master's degree in software engineering. Her research interest focuses on the social networking sites.

AHMAD ALMOGREN received the Ph.D. degree in computer science from Southern Methodist University, Dallas, TX, USA, in 2002. Previously, he was an Assistant Professor of computer science and a member of the scientific council, Riyadh College of Technology. He also served as the Dean of the College of Computer and Information Sciences, and the Head of the Council of Academics, Al Yamamah University. He is currently a Professor and the Vice Dean of the development and quality with the College of Computer and Information Sciences, King Saud University. His research areas of interests include mobile and pervasive computing, cyber security, and computer networks. He has served as a Guest Editor for several computer journals.

ASSAD ABBAS received the Ph.D. degree in electrical and computer engineering from North Dakota State University, Fargo, ND, USA. He is currently an Assistant Professor of computer science with COMSATS University Islamabad, Islamabad, Pakistan. His research interests are mainly, but not limited to, smart health, big data analytics, recommendation systems, patent analysis, software engineering, and social network analysis. Moreover, his research has appeared in several reputable international venues. He is also serving as the referee for numerous prestigious journals and as the technical program committee member for several conferences. Moreover, he is a member of the IEEE-HKN.

HASAN ALI KHATTAK (SM'19) received the B.CS. degree in computer science from the University of Peshawar, Peshawar, Pakistan, in 2006, the master's degree in information engineering from the Politecnico di Torino, Torino, Italy, in 2011, and the Ph.D. degree in electrical and computer engineering degree from the Politecnico di Bari, Bari, Italy, in 2015. He is currently an Assistant Professor of computer science with COMSATS University Islamabad, since 2016. His current research interests focus on web of things, data sciences, social engineering for future smart cities. His perspective research areas are application of machine learning and data sciences for improving and enhancing the quality of life in smart urban spaces through predictive analysis and visualization. He is a Professional Member of the ACM and an active member of the IEEE ComSoc, the IEEE VTS, and the Internet Society.

IKRAM UD DIN (SM'18) received the M.Sc. degree in computer science and the M.S. degree in computer networking from the Department of Computer Science, University of Peshawar, Pakistan, and the Ph.D. degree in computer science from the School of Computing, Universiti Utara Malaysia (UUM). He has also served as the IEEE UUM Student Branch Professional Chair. He has 10 years of teaching and research experience in different universities/organizations. His current research interests include resource management and traffic control in wired and wireless networks, vehicular communications, mobility and cache management in information-centric networking, and the Internet of Things.

MOHSEN GUIZANI (S'85–M'89–SM'99–F'09) received the B.S. degree (Hons.) and M.S. degree in electrical engineering, and the M.S. and Ph.D. degrees in computer engineering from Syracuse University, Syracuse, NY, USA, in 1984, 1986, 1987, and 1990, respectively. He is currently a Professor with the CSE Department, Qatar University, Qatar. Previously, he has served as the Associate Vice President of graduate studies with Qatar University, University of Idaho, Western Michigan University, and University of West Florida. He has also served in academic positions at the University of Missouri-Kansas City, University of Colorado-Boulder, and Syracuse University. His research interests include wireless communications and mobile computing, computer networks, mobile cloud computing, security, and smart grid. He is a Senior Member of the ACM. He is the author of nine books and more than 500 publications in refereed journals and conferences. He has guest-edited a number of special issues in the IEEE journals and magazines. He has also served as a member, Chair, and General Chair for a number of international conferences. He received three teaching awards and four research awards throughout his career. He received the 2017 IEEE Communications Society Recognition Award for his contribution to outstanding research in wireless communications. He was the Chair of the IEEE Communications Society Wireless Technical Committee and the Chair of the TAOS Technical Committee. He is currently the Editor-in-Chief of the *IEEE Network Magazine*, serves on the editorial boards of several international technical journals, and the Founder and Editor-in-Chief of the *Wireless Communications and Mobile Computing Journal* (Wiley). He has served as the IEEE Computer Society Distinguished Speaker, from 2003 to 2005.

MANSOUR ZUAIR received the B.S. degree in computer engineering from King Saud University, Riyadh, Saudi Arabia, and the M.S. and Ph.D. degrees in computer engineering from Syracuse University. He has served as the CEN Chairman (2003–2006), the Vice Dean (2009–2015), and has been the Dean, since 2016. He is currently an Associate Professor with the Department of Computer Engineering, College of Computer and Information Sciences, King Saud University. His research interests are in the areas of computer architecture, computer networks, and signal processing.