Detecting Malicious Users on Twitter Using Topic Modeling

Myo Myo Swe
Web Mining Lab
University of Computer Studies, Mandalay
Mandalay, Myanmar
myomyoswe@ucsm.edu.mm

Nyein Nyein Myo
Faculty of Information Science
University of Computer Studies, Mandaly
Mandalay, Myanmar
nyeinnyeinmyo@ucsm.edu.mm

***Abstract*— Social networking sites like Twitter, Weibo and Facebook etc. have played a significant role in daily lives of people as they promote innovative ways to connect efficiently and exchange their knowledge. The benefits of these social network services cause them to expand their community rapidly. Most current social network sites face additional issue of coping with unauthorized users and their high levels of violence activities, which distribute fake news, worms and viruses, etc. to the genuine users. Spam distribution degrades user experience and also has a negative effect on server-side functions such as knowledge discovery, user activity analysis and service selection. In this paper, whitelist and blacklist are built which can help to distinguish malicious users and legitimate users. With the aid of blacklist and whitelist, we introduced two new features: malicious probability and legitimate probability. Evaluation has been carried out on the CRESCI-2015 dataset. Three machine learning classifiers like AdaBoost, Bagging and Random Forest. Random Forest obtained the highest 99.7% detection score.**

Keywords*— *whitelist, blacklist, legitimate, spammer

I. INTRODUCTION

Online Social Networks (OSNs) are microblogging sites where anyone can create profiles, communicate and engage with other users. There are hundreds of these microblogging sites accessible around the world. Facebook and Twitter are well known microblogging sites among these OSNs and there are over 500 million active users on those sites. Twitter, one of the OSNs, is being introduced to support communications between people using text-based posts called tweets. Online social networks (OSNs) have evolved to become more and more popular in the last few years. Facebook users love these social networking sites and they absorb much of the attention from internet users. In fact, due to the Pew Research Centre's report, interpersonal media sources were the key new hotspot for around 30% of American. Furthermore, technologies for person-to-person contact in mobile phones affect the entry of consumers.

Billions of internet users are spending enormous energy in OSNs to communicate with strangers or familiar people. Users can not only send messages to each other after links have been established but can also share interesting or latest events with their companions in terms of tweets, mentions or comments. The communication and mutual knowledge of these users forms a gigantic social network with massive distribution of data within it. The social networking sites are of interest to various researchers. Jon et al. analyzes the social network and social search [1], whereas some scientists estimate the informal online organizations. Meanwhile, in OSNs, authors are exploring foresight of connections and induction quality. In informal network, privacy security is a stream. In addition, researchers are still considering sybil attack in OSNs. In OSNs, researchers implemented methods of community detection [2] to analyze user activities within a group.

In fact, a few studies are on the nature of online social networks and the dispersal of information. Some researchers forecast consumer feelings from different perspectives [3] foreseeing influenza slant, or despite distinguishing tremor in OSNs. In spite of the enthusiasm of scientists, the rich data in OSNs has likewise pulled into the consideration of cyber-criminals. For examples, fake users create fake accounts on Twitter to spread worm and malware on online social networks [4]. Such malwares in OSNs are utilizing old plans to use in the new stages. Worms in OSNs like email worms such as LoveLetter utilize friend list of victims to send spam messages to other users. Some fake users forge a same account of victim for stealing the detailed information of victim's friends. They use the stolen information for spreading malware and tricks and customizing victim's daily routine on online social network.

There were some past investigations on the Twitter malicious users recognition for quite a long time. Their methodologies address this issue for distinguishing a Twitter account into a malicious account or a real account. By studying the fake characteristics, the successful features that are extracted from profile-based, content-based, and timing-based of users have been proposed. However, malicious users also utilize new strategies to evade detection and existing methods and features are not enough to effectively detect fake account. The difference characteristics between malicious users and genuine users are initiated to examine in this works. These characteristics are used as features to detect malicious users from legal ones.

To prevent legitimate users from clicking malicious links, implementing a system to combat spam on Twitter. This should be possible by preventing users from posting tweets which contain malicious links at run-time or recognizing users over and again making such posts. This system overcomes both of these issues. We will implement a system which can be used to detect malicious accounts on Twitter.

Our contribution is that we propose an approach to detect malicious users on Twitter using user generated content. Two new corpuses: whitelist and blacklist are built using topic modeling approach.

The main aim of this work is to implement malicious users detection system to prevent genuine users from malicious users which posted malicious content and links. This system can reduce not only false positive rate but also model building time. The four objectives are fulfilled to achieve this aim.

- To build blacklist and whitelist that can effectively classify malicious users from legitimate users
- To extract 26 features from user's profile and timeline

- To evaluate the detection system with three classifiers namely AdaBoost, Bagging and Random Forest

This paper is written according to the following outlines. Section 2 presents the related prior works on malicious users identification on Twitter. Section 3 describes methodology for identification of malicious account. Section 4 deals with assessment of the suggested solution. The paper concludes in section 5.

II. RELATED WORK

The most useful methodology for distinguishing malicious accounts on OSN is applying machine learning strategies. At first, it gathers and recognizes features that can recognize malicious users from real users and then build a binary classifier to isolate these two types of users. This sort of methodology is called feature-based method. Gupta et al. [5] carried out a top to bottom portrayal of the parts that are influenced by the quickly becoming noxious substance. It was seen that an enormous number of individuals with high social profiles were answerable for circling counterfeit news. The authors chose the records that were fabricated following the Boston impact and were later prohibited by Twitter because of infringement of terms and conditions to perceive the phony records. 3.7 million particular clients gathered about 7.9 million unmistakable tweets. This dataset is known as the biggest dataset of Boston impact. The researchers fulfilled the phony substance classification through transient investigation where worldly circulation of tweets is determined dependent on the quantity of tweets posted every hour [6].

Concone et al. [7] introduced a strategy that gives threatening alarming by utilizing a predetermined arrangement of tweets progressively vanquished through the Twitter API. Thereafter, the clump of tweets considering a similar point is summarize to produce a caution. The proposed engineering is utilized to assess Twitter posting, perceiving the progression of permissible function, and announcing of that function itself. The proposed approach uses the data contained in the tweets when a spam or malware is perceived by the clients or the report of security has been delivered by the guaranteed specialists.

Eshraqi et al. [8] decided various highlights to distinguish the spam and afterward with the assistance of a nook stream-based grouping calculation, perceive the spam tweets. Some client accounts were chosen from different datasets and a short time later arbitrary tweets were chosen from these records. The tweets are in this manner sorted as spam and non-spam. The creators asserted that the calculation can partition the information into spam and non-spam with high exactness and phony tweets possibly perceived with high exactness and accuracy.

Researchers presented increasingly powerful features for Twitter fake user discovery in 2011 [9]. They targeted on relations among fake users and their neighbors and then computed important features such as a bidirectional relationship and centrality of betting. Different features dependent on timing and robotization were likewise presented in their papers. Other timing and automation-based features were also included in their papers. Like Yang's work, work in [10] considered the relationships between spam senders and collectors, for example, the briefest ways and least slice to remove features.

To find spammers on Twitter, Naïve Bayesian classifier was used by Wang et al. [11]. Spammers were detected using graph-based features and content-based features. In this approach, reputation is the strongest feature. To evaluate this approach, a ground truth dataset which contain a number of 500 labeled users was used. In this dataset, 3% of users were spammers. This approach classified 392 out of 25817 users as spammers, and by manually checking 342 spammers were real spammers. Therefore, the precision of this approach was 89%. The disadvantage of this approach is that proposed features were not robust for evasion. Therefore, more effective features that can effectively detected spammers were proposed by Yang et al. [12]. The authors carefully analyzed evasion pattern of social spammers and proposed new robust features.

Various researchers proposed various fake accounts detection methods. Some researchers carefully examined the users' profiles and users' tweets and extracted features from these profiles and tweets. Some researchers analyze behavior patterns of fake users and utilizing these stranger behaviors, they can distinguish fake users from normal users. Works utilizing social connection were also developed. Later works focus on early spam discovery to quickly mitigate threats. Content-based detection methods are exceptionally encouraging as they just concentrate data from tweets. In 2013, [13] proposed spam detection system. Reviews datasets were used to classify deceptive or truthful reviews. Their works can only be applied on review datasets. But our approach uses user's profile and timeline information and finds spammers on Twitter.

In our previous works [14][15], only blacklist is created to extract two new features. In this work, only fake users' tweets are utilized for creating blacklist. Characteristics of legitimate users were not considered and tweets generated from legitimate users were not utilized. But, in this paper, we deeply analyst not only malicious users' behaviors but also legitimate too. Whitelist is created from legitimate users' tweets and blacklist is created from malicious users' tweets. In our approach, whitelist and blacklist are built in advance using topic modeling approach. Two new features: (1) malicious probability and (2) legitimate probability are extracted from user's tweets with the help of whitelist and blacklist.

III. METHODOLOGY

Our spammer detection approach is a two-step process. Before spammer detection system have been developed, whitelist and blacklist are created from users' generated content.

A. Building Whitelist and Blacklist

The first step of our approach is to create two corpuses: namely whitelist and blacklist in advance. The following stages are needed to create whitelist and blacklist.

1) Tweets collection: To build whitelist and blacklist, CRESCI-2015 dataset is used. Tweets in this dataset were annotated as two classes – spammers and non spammers. From this dataset, spammers tweets are collected to build blacklist and tweets of legitimate users are gathered for whitelist. 118,327 tweets of 1950 spammers are achieved for blacklist and 1950 legitimate users' 2,631,730 tweets are used for whitelist.

2) Tweet preprocessor: After collecting spammers' tweets and legitimate users' tweets, tweet preprocessor is

applied to preprocess these tweets. Tweet preprocessor does the following works.

a) Removing noise in tweets: In this step, hashtag (#), mentions (@), links (https (or) www), numbers (0-9) and punctuation marks (!,@,#,…) are removed from the tweets. In addition, words that length have less than three and greater than twelve are removed from tweets.

b) Case normalization: Convert all characters into lowercase characters.

c) Remove stopwords: Words like the, have, then, is, was, being and so forth which are essential for sentence arrangement however truly do not pass on any message as individual words are called as stop words. These words are deleted from tweets.

d) Translate slangs and abbreviations: We tarnslate slangs and abbreviations words to their original form. for example, 2nite -> tonight, afaiaa -> as far as i am aware, cul8r -> see you later and ic -> i see.

e) Replace accented characters to unaccented characters: for example, replace accented characters (ã, ê, ñ) to unaccented characters (a, e, n).

f) Lemmatization: Lemmatization is a significant natural linguistic processing (NLP) work. Lemmatization generally relates to doing things correctly using a word vocabulary and morphological analysis, generally aimed only at removing inflectional endings and returning the base or dictionary form of a term known as the lemma. For example, cars -> car, caring -> care.

3) Keyword extraction: In this step, all tweets posted by fake users are regarded as fake tweets. Entire tweets of legitimate users are marked as legitimate tweets. Each user's tweets are collected in each documents. There are 1,950 fake users and 1,950 legitimate users. Therefore, 1,950 fake documents and 1,950 legitimate documents are achieved for keyword extraction. Words in fake documents are candidate for malicious (dangerous) words, but these are not malicious words. The words in legitimate documents are aslo candidate for legitimate words. To extract malicious keywords and legitimate keywords from fake documents and legitimate documents, topic modeling approach is used. For topic modeling, Latent Dirichlet Allocation (LDA) is utilized. LDA is a useful topic modeling techniques in tweets. User can posted 280 characters in one tweet. Tweets are short text message. Traditional keyword extraction method such as term frequency and inverse document frequency (tf_idf) cannot be used in tweets because term frequency of most words are one. LDA overcomes this issue and it can good represent about topics of tweets. Blei et al.[16] suggested Latent Dirichlet Allocation as the basis for the extraction of topics. LDA is a three-steps Bayesian model, in which all of an accumulation is showed as a limited blend over a basic configuration of subjects. Every topic is constructed as an infinite mixture over the fundamental set of topic probability. The thematic probabilities make available an obvious representation of a document in the sense of text mining. Each document is built with set of words $W = \{wi_1, wi_2, \ldots, wi_M\}$. M is the number of words. Each word is distributed to one of the document's topics $Z = \{z_{i1}, z_{i2}, \ldots, z_{iK}\}$. K is the number of topics. Φ_m is a multinomial distribution over words for topic m. θ_i is an another multinomial distribution over topics for document i. α is the parameter of the Dirichlet prior on the per-document topic distributions. β is the parameter of the Dirichlet prior on the per-topic word distribution. We set the valuses of α, β, and M to 0.1, 0.01 and 8. After topic modeling have been applied on spam documents and legitimate documents, whitelist and blacklist are achieved. There are 317 fake keywords in blacklist and 320 legitimate keywords in whitelist. Two new features malicious probability and legitimate probability are calculated using blacklist and whitelist.

B. Twitter Malicious Users Detection

The second step of our approach is implementing malicious users detection system. To identify malicious users on Twitter, features are extracted from user's generated content. User's generated content are users' tweets and user' profile information. Required features are get from users' profile and timeline. Two meta learner classifiers namely AdaBoost and Bagging and one tree-based classifier namely Random Forest are applied to distinguish spammers from legitimate users.

1) Dataset: to assess our way to deal with recognize spammers on twitter, we need a labeled dataset which was annotated into spammers and real users. For evaluating our approach, CRESCI-2015 dataset is used. This data was created by Cresci et al (2012) [17] between Jan 2013 and Feb 2013. This dataset contains the data of 1950 spammers and 2,631,730 their tweets. They marked 1950 clients as legitimate and 118,327 tweets of their tweets. In CRESCI-2015 dataset, the ratio of spammers to legitimate is 1:1.

2) Features Extraction: Features are extracted from users' profile and users' timeline. In our approach, exisiting twenty four features and two new features such as spam probability and legitimate probability are extracted for spammer classification. Twenty six features are as follow.

a) Number of followers: Mailicious users have less number of followers, but legitimate users have a lot of followers.

b) Number of followings: Malicious users follow a lot of another users.

c) Reputation: Reputation of malicious users is smaller than 0.5.

$$reputation = \frac{number\ of\ follower}{number\ of\ follower + numb\quad of\ following} \quad (1)$$

d) Account age: Most of the legitimate users'age are at least three months.

e) Follower rate: Legitimate users have high follower rate.

$$follower\ rate = \frac{number\ of\ follower}{Account\ age} \quad (2)$$

f) Following rate: Malicious users have high following rate.

$$following\ rate = \frac{number\ of\ following}{Account\ age} \quad (3)$$

g) Number of tweets: Legitimate users posted tweets frequently. Number of tweets posted by legitimate users are high.

h) Number of retweets: Malicious users frequently retweet other users tweets or their tweets.

i) Number of urls: Mailicious users posted malicious links in tweets. Number of urls containing in tweets are high in malicious users'tweets.

j) Number of mentions: Malicious users mentions another users in their tweets to get more attention. Therefore, number of mentions are high in mailicious user's tweets.

k) Number of hashtags: Malicious users also use hashtags and posted trending topics to grab social users'attention.

l) Total number of words: Malicious users aim is to persuade legitimate users. Therefore, they post a lot of tweets a day. They use bot to post tweets automatically.

m)Hashtag ratio: Hashtag ratio of malicious user are higher than that of legitimate user.

$$hashtag\ ratio = \frac{number\ of\ hashtag}{number\ of\ tweets} \quad (4)$$

n) Mention ratio: Legitimate users have smaller mention ratio.

$$mention\ ratio = \frac{number\ of\ mention}{number\ of\ tweets} \quad (5)$$

o) url ratio: Malicious users have higher url ratio than legitimate user.

$$url\ ratio = \frac{number\ of url}{number\ of\ tweets} \quad (6)$$

p) Mean time between tweets: Malicious users utilize automated twitting tools or devices to post tweets automatically. They post tweets in a specific time (for example: they posted tweets every 30 minutes). Therefore, mean time between tweets of malicious user are different from legitimate user.

$$\mu = \frac{\sum(timestamp(tweet_i) - timestamp(tweet_j))}{number\ of\ tweets - 1} \quad (7)$$

q) Maximum idle duration time between tweets: Legitimate users post tweets randomly and they have random behavior. Legitimate users have maximum idle duration time.

$$max = max\ (timestamp(tweet_i) - timestamp(tweet_j)) \quad (8)$$

r) Standard deviation time between tweets: Standard deviation time between tweets of legitimate users and malicious user are different.

$$\sigma = \sqrt{\frac{\sum(X-\mu)}{number\ of\ tweets - 1}} \quad (9)$$

s) Number of characters in user's name:

t) Number of characters in user's screen name:

u) Number of lists: Malicious users have a lot of list count.

v) Number of favourites: Malicious users have a lot of favourites users.

w) Number of tweets per day: Malicious users posted specific number of tweets in specific time. Number of tweets per day of malicious users are more than that of legitimate users.

x) Number of malicious keywords in description: Most of the malicious users posted fake keywords in their description.

y) Malicious propability: Aim of malicious user and that of legitimate user are not the same. Malicious users use social networking site to promote their product items, to defame other person, and etc. Behaviours of malicious user and that of mailicious user are also different.Malicious user posted malicious words in their tweets. Malicious words in listed in blacklist. Number of fake words can be counted by using blacklist. If malicious probability is high, malicious suspectiability is also high.

$$malicious\ probability = \frac{number\ of\ fake\ keywords}{number\ of\ tweets} \quad (10)$$

z) Legitimate probability: Leigitimate words can be counted by using whitelist. If legitimate probability is high, user suspectiability is low.

$$Legitimate\ probability = \frac{number\ of\ legitimate\ keywords}{number\ of\ tweets} \quad (11)$$

3) Classification: Three machine learning classifiers are applied to classify spammers and non-spammers. AdaBoost, Bagging and Random Forest are used for spammer detection.

a) Bagging: Bagging is used to reduce the variance of the decision tree classifier. The goal is to create many subsets of data from a randomly selected training sample with a replacement. Each collection of subset data is used to train their decision tree. As a result, we are having a range of different versions. Average of all predictions from various trees is used which is more reliable than a single decision tree classifier.

b) AdaBoost: AdaBoost is used to produce a series of predictors. In this technique, learners are sequentially trained with early learners to fit simple models to the data and then analyze data for errors. Consecutive trees (random sample) are fit and the goal at each stage is to increase the accuracy of the previous tree. If the input is misclassified by the hypothesis, the weight of the input is increased so that the next hypothesis is more likely to interpret it correctly. This method transforms poor learners into a more effective model.

c) Random Forest: Random Forest (RF) is a commonly used technique of machine learning that demonstrates competitive efficiency in different areas, including social science, finance, chemical engineering, biological science, medical analysis, etc. It is an ensemble learning technique for classification and regression. It builts a variety of decision trees while training and outputs the class which is the class mode (classification) or mean prediction (regression) of the individual trees. Random choice forests are right for the practice of decision trees to overfit their training set.

IV. EXPERIMENTAL RESULTS

For experiment, we utilize core i7 processor, 8 GB RAM, 2 TB HDD and 64 bits Window 10 OS. The proposed framework is implemented by means of Java programming language (NetBeans IDE 8.2). Machine learning tool (Weka 3.8) is used for implementing three classifiers. The dataset CRESCI-2015 is used to test the approach. CRESCI-2015 dataset includes 3,900 users (1,950 spam users + 1,950 regular users) and 2,750,097 tweets from these users.

A. Performance Evaluation on Three Classifiers

Evaluation results are based on tenfold cross validation. In tenfold cross validation, dataset is separated into ten subsets. One subset is used for testing and other nine subsets are used for training. In our approach, there are 3,900 instances (users) in dataset. Therefore, 390 instances are utilized for testing and 3510 instances are utilized for training. AdaBoost, Bagging and Random Forest classifiers are used for spam accounts detection. Three accuracy matrics: (1) precision, (2) recall and (3) F-measure are calculated to compare the classifiers' results. In Table 1, these comparative results are shown. Precision, recall and f-measure of AdaBoost is 0.988, 0.986 and 0.9787. Bagging gets 0.988 precision, 0.986 recall and 0.987 f-measure. Random Forest gives 0.997 precision, 0.997 recall and 0.995 f-measure. According to experimental results, Random Forest gives the best result. AdaBoost is better than Bagging. Performance comparison of three classifiers are described in Table 2. Detection rate of Random Forest is 99.5%. This is the highest rate among three classifiers. Model building time of Random Forest is 0.84 seconds and this time is the longest among three classifiers. But this time is acceptable for real time spammers detection and the most important fact of spammers detection is that the system needs to be correctly classify spammers on Twitter. Therefore, Random Forest classifier is chosen for spammer detection on Twitter. Because it handles very well not only a collection of data of a higher dimension but also the missing quantities. It can also ensure consistency for missing data and its accuracy outperforms AdaBoost and Bagging.

TABLE I. COMPARATIVE RESULTS OF THREE CLASSIFIERS

Classifiers	Precision	Recall	F-Measure
AdaBoost	0.988	0.986	0.987
Bagging	0.973	0.973	0.973
Random Forest	0.997	0.997	0.995

TABLE II. PERFORMANCE COMPARISON OF THREE CLASSIFIERS

Classifiers	Detection Rate	Model Building Time
AdaBoost	98.7%	0.25 seconds
Bagging	97.3%	0.39 seconds
Random Forest	99.5%	0.84 seconds

B. Compare with Other Social Spammers Detection Approach

We compare our malicious user detection system with the approach of Meda et. al [18]. In this Meda et. al's approach, the author used fourteen features such as number of followers, number of followings, number of replies, number of hashtags, number of urls, number of characters, number of spam words, number of words, number of mentions, number of numeric characters, number of tweets, number of tweets per day, time between post and age of user account. The author used support vector machine (SVM), extreme learning machine (ELM) and Random Forest (RF) to distinguish spammers on Twitter. Their approach achieved best result on Random Forest classifier. In our proposed approach, we utilize twenty-six features including two new proposed features for detection. Performance comparison of our approach and Meda's approach is shown in Table 3. False positive rate of our approach is 0.012 and that of comparative approach is 0.06. False negative rate of comparative approach is five time greater than that of our approach. Therefore, proposed approach can reduce false positive rate and false negative rate.

TABLE III. PERFORMANCE COMPARISON BETWEEN PROPOSED APPROACH AND MEDA'S APPROACH

Evaluation Metrics	Proposed Approach	Other Approach
Precision	0.997	0.942
Recall	0.997	0.967
False Negative Rate	0.002	0.033
False Positive Rate	0.012	0.06

C. Compare with Proposed Features and without Proposed Features

We also perform comparison with proposed features and without features. Comparative results of with and without proposed features are shown in Table 4. Without proposed features, precision is 0.894, recall is 0.895 and f-measure is 0.894. But proposed features are added to classifier, precision increases to 0.997, recall increases to 0.997 and f-measure also increases to 0.995. The comparative results show that accuracy is increased when two new features are added.

TABLE IV. COMPARATIVE RESULTS OF WITH AND WITHOUT PROPOSED FEATURES SET

Performance Metrics	Without Proposed Features	With Proposed Features
Precision	0.894	0.997
Recall	0.895	0.997
F-Measure	0.894	0.995

V. CONCLUSION AND FUTURE WORK

In this paper, two new features such as malicious probability and legitimate probability are extracted from tweets and user profile. To extract two new features, we build two corpuses such as whitelist and blacklist using topic modeling approach. Experimental results showed that our proposed features can not only improve detection accuracy but also reduces false negative rate and false positive rate. Our proposed two new features can increase accuracies of spammers detection according to the performance comparison of with and without proposed features. Accuracies of three classifiers such as AdaBoost, Bagging and Random Forest are compared and Random Forest achieves the best detection result. In our approach, we created whitelist and blacklist from CRESCI-2015 dataset. This dataset was created in 2015, therefore, we need more up to date data beyond 2015. In future, we will create new dataset with up to date information and we will also extract more robust features that can be applied on other social networking sites such as Facebook and Weibo etc.

REFERENCES

[1] Kossinets, G., Kleinberg, J. and Watts, D., 2008, August. "The structure of information pathways in a social communication network," in *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 435-443), 2008.

[2] Yang, Z., Wilson, C., Wang, X., Gao, T., Zhao, B.Y. and Dai, Y., 2014. "Uncovering social network sybils in the wild," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *8*(1), pp.1-29, 2014.

[3] Bollen, J., Mao, H. and Pepe, A., 2011, July. "Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena," in *Fifth International AAAI Conference on Weblogs and Social Media,* 2011.

[4] Sakaki, T., Okazaki, M. and Matsuo, Y., 2010, April. "Earthquake shakes Twitter users: real-time event detection by social sensors," in *Proceedings of the 19th international conference on World wide web* (pp. 851-860), 2010.

[5] A. Gupta, H. Lamba, and P. Kumaraguru, "1.00 per RT #BostonMarathon # prayforboston: Analyzing fake content on Twitter," in *Proc. eCrime* Researchers Summit (eCRS), 2018, pp. 1-12.

[6] Masood, F., Almogren, A., Abbas, A., Khattak, H.A., Din, I.U., Guizani, M. and Zuair, M., 2019, "Spammer detection and fake user identification on social networks," *IEEE Access*, *7*, pp.68140-68152.

[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] Yang, C., Harkreader, R., Zhang, J., Shin, S. and Gu, G., 2012, April. "Analyzing spammers' social networks for fun and profit: a case study of cyber criminal ecosystem on twitter," in *Proceedings of the 21st international conference on World Wide Web* (pp. 71-80), 2012.

[10] Song, J., Lee, S. and Kim, J., 2011, September. "Spam filtering in twitter using sender-receiver relationship," in *International workshop on recent advances in intrusion detection* (pp. 301-317). Springer, Berlin, Heidelberg, 2011.

[11] Wang, A.H., 2010, July. "Don't follow me: Spam detection in twitter," In *2010 international conference on security and cryptography (SECRYPT)* (pp. 1-10). IEEE, 2010.

[12] Yang, C., Harkreader, R. and Gu, G., 2013. "Empirical evaluation and new design for fighting evolving twitter spammers," *IEEE Transactions on Information Forensics and Security*, *8*(8), pp.1280-1293, 2013.

[13] Li, J., Cardie, C. and Li, S., 2013, August. "Topicspam: a topic-model based approach for spam detection," in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (pp. 217-221), 2013.

[14] Swe, M.M. and Myo, N.N., 2018. "Blacklist Creation for Detecting Fake Accounts on Twitter," *International Journal of Networked and Distributed Computing*, *7*(1), pp.43-50, 2018.

[15] Blei, D.M., Ng, A.Y. and Jordan, M.I., 2003. "Latent dirichlet allocation," *Journal of machine Learning research*, *3*(Jan), pp.993-1022, 2003.

[16] Swe, M.M. and Myo, N.N., 2018, June, "Fake accounts detection on twitter using blacklist," in *2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS),* pp. 562-566, IEEE.

[17] Cresci, S., Di Pietro, R., Petrocchi, M., Spognardi, A. and Tesconi, M., 2015. "Fame for sale: Efficient detection of fake Twitter followers," *Decision Support Systems*, *80*, pp.56-71, 2015.

[18] Meda, C., Bisio, F., Gastaldo, P. and Zunino, R., 2014, October. "A machine learning approach for Twitter spammers detection," In *2014 international carnahan conference on security technology (iccst)* (pp. 1-6). IEEE, 2014.