

## **SPAMMER DETECTION AND FAKE USER IDENTIFICATION ON SOCIAL NETWORKS**

**SK.Nadeem**

*Department of Computer Science &Engineering, Narayana Engineering College Gudur  
shaiknadeem983@gmail.com*

**N.Kesava Rao**

**Associate Professor**

*Department of Computer Science &Engineering, Narayana Engineering College Gudur  
kesavarn@gmail.com*

**V.Srinath Reddy**

*Department of Computer Science &Engineering, Narayana Engineering College Gudur  
srinathr389@gmail.com*

**K.Nikhil**

*Department of Computer Science &Engineering, Narayana Engineering College Gudur  
nikhil70977@gmail.com*

### **ABSTRACT**

Social networking sites include millions of users worldwide. The interaction of users with these social networking sites, such as Twitter and Facebook, has a profound effect and sometimes has negative effects on daily life. Large social networking sites have become a platform for beneficiaries to disseminate huge amounts of misinformation. For example, Twitter has become one of the most commonly used platforms and therefore unintentionally allows for the wrong amount of spam. False users send unsolicited tweets to users to promote services or websites that not only affect legitimate users and interfere with the use of the services. In addition, Twitter marketing strategies that showcase strategies are based on their ability to find: (i) false content, (ii) spam-based URLs, (iii) spam in the headlines, and (iv) fake users. The strategies you presented were compared based on a variety of factors, such as user characteristics, content features, graph features, layout features, and time elements. We hope that this presented study will be a valuable resource for researchers to find highlights of the recent development of Twitter spam.

**Keywords:** Classification, fake user detection, Online social network, spammer's identification.

### **1.INTRODUCTION**

Twitter is already ineligible to receive any kind of information from any source worldwide using the Internet. The increasing demand for social networking sites allows users to gather more information and data about users. The sheer amount of information available on these sites also attracts the attention of fraudulent users [5]. Twitter is becoming an online source for getting real-time information about users. Twitter is an online social network (OSN) where users can share anything and everything, such as their news, ideas, and experiences. Many debates can be made on a variety of topics such as politics, current affairs and important events. When a user launches something, they are immediately informed by their followers, allowing them to disseminate more detailed information [2]. Most people who do not have much experience with OSNs can be easily deceived by fraudsters. Identification of fake news [8] on social media is an issue that needs to be addressed on an individual and collective level [1] because of the negative effects of such issues.

## **2.LITERATURE SURVEY**

Twitter offers a survey of new and innovative ways to detect spam. The above survey presents a comparative study of the curve method. On the other hand, the author [10] conducted a survey of the various behaviors displayed by the elite players on Twitter. This study also provides a literature review that identifies the presence of spammers on the Twitter social network. Despite all the available studies, there is still some literature available. So, to close the gap, we do a thorough review of spammer detection and false user identification on Twitter. In addition, the survey presents a classification of the Twitter spam detection approach and attempts to provide a detailed description of recent developments in the domain. The purpose of this paper is to identify the different apps of spam detection on Twitter and to classify these approaches into several categories and present a classification. For classification, we have identified four means of reporting spammers that may be helpful in identifying fake identities of users. Spammers can be identified based on: (i) fake content, (ii) URL-based spam detection, (iii) spam detection in trending topics, and (iv) fake user-identification.

## **3. METHODOLOGY**

### **SPAMMER DETECTION ON TWITTER**

In this article, we defined a classification of spammer detection techniques. Fig. 1 shows the taxonomy for identification of spammers on Twitter. The proposed taxonomy is categorized into four main classes, namely,

- Fake content
- URL based spam detection,
- Fetching spam in trending topics
- 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.

### **FAKE CONTENT BASED SPAMMER DETECTION**

[7] The depth of appearance of the elements affected by the newly grown cultivated content. It was seen that a large number of high-profile people were also funded to spread fake news. To obtain fake accounts, the authors opted for accounts created immediately after the Boston explosion and were later blocked by Twitter for violating the terms and conditions. Approximately 7.9 million tweets have been collected by 3.7 million unique users. This data is known as Boston Blast's largest database. calculated based on tweet number per hour. A Lfun (learning for unlabeled tweets) scheme, which is used to solve various problems in the finding of Twitter spam, has been proposed by Chen *et al.* [3]. Their framework comprises two components, i.e., learn from detected tweets (LDT) and learn from human labelling (LHL). The two components will automatically generate spam tweets from the given tweets. The taxonomy is shown in Fig.1

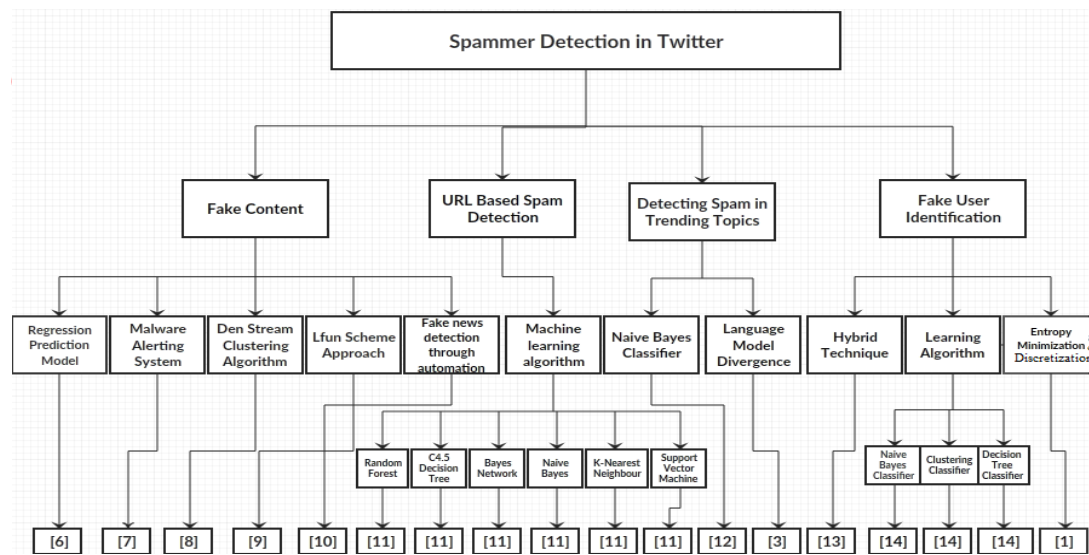


Figure-1: Taxonomy of Spammer detection and fake user identification on Twitter

## URL BASED SPAM DETECTION

We performed machine learning algorithms [4] to detect spam tweets. For example, an analysis of the effect of various characteristics on spam detection: (i) spam on non-spam scales, (ii) training data size, (iii) time-related data, (iv) factors sensitivity, and (v) sample data. To assess ownership, first, nearly 600 million public tweets were collected and later the authors used the Trend Micro web reputation system to identify spam tweets as much as possible. A total of 12 different light features were also classified to distinguish non-spam and spam tweets from the data shown. The features of the identified features were represented by the information in the cdf. These features are understood by being installed in a machine-readable spam machine, later used in the experience to check spam ownership. The four datasets are sampled to reproduce different scenarios. Since no dataset is publicly available for the task, some datasets were used in previous researches. Following the identification of spam tweets, 12 features were collected. . On the other hand, tweet-based features include

- (i) Retweets
- (ii) hash tags,
- (iii) user mentions and
- (iv) number of URLs.

The evaluation results show that changing feature distributions reduced performance, whereas no differences were observed in training dataset distributions.

## DETECTING SPAM IN TRENDING TOPIC

Gharge *et al.*[6] initiated a method. The system framework consists of following five steps:

- Collection of tweets in relation to trending topics on Twitter. After storing tweets in a particular file format, tweets are subsequently analyzed.
- Labeling of spam is done to check through all datasets that are available to detect malicious URLs.
- Feature extraction distinguishes the creation of features based on language models that use the language as a tool and helps determine whether tweets are fake or not.
- Classification of data sets is done by shortlisting the set of tweets that is described by the set of features provided by the classifier to instruct the model and gain knowledge to detect spam.
- The spam detection uses the classification technique to accept tweets as the input and classify the spam and non-spam.

## **FAKE USER IDENTIFICATOIN**

The classification method is proposed by Erasin et al[5]. Finding spam accounts on Twitter. The data used in the study were collected manually. The category is displayed by analyzing the username, profile and background image, number of friends and followers, content of tweets, account information and number of tweets. The database contained 501 fake accounts and 499 original accounts, where 16 references were identified from the Twitter API. Two tests have been performed to separate the fake accounts. The first experiment uses the Naïve Bayes learning algorithm in Twitter data, including all aspects without understanding, while the second experiment uses the Naïve Bayes learning algorithm in Twitter data after re-thinking. These features are based on messages or user-generated content. Spammers post content to spread fake news, and these materials contain malicious URLs to promote their product. Content-based features include

- (i) The total number of tweets,
- (ii) Hashtag ratios,
- (iii) URL ratios,
- (iv) Mentions ratios, and
- (v) Frequency tweets.

The graph-based[9] feature is used to control theft strategies operated by spammers. Spammers use various techniques to avoid detection. They can buy fake followers from different third-party web sites and exchange their followers for another user to look like a legal user.

## **3. RESULTS AND DISCUSSIONS**

From the research, we analyzed how risky activities are performed on social networks in many ways. In addition, researchers have tried to identify unwanted spammers or bloggers by proposing various solutions. Therefore, to combine all relevant efforts, we proposed classification in terms of extraction and classification. Classification is based on identifying various factors such as fake content, based on URLs, headlines and false users. The first major commercial breakdown is the proposed spam detection strategy, which is incorporated into the Twitter segment through counterfeit content. Spammers General - Associate spam data with a topic or keyword that is cruel or contains potentially spam words. The second section considers spam detection strategies based on URLs. In addition to reviewing techniques, this study also provides comparisons of various Twitter spam detection features. These features are extracted from user accounts and tweets that can help identify spam. These features are divided into five categories, namely user, content, graph, structure and time. User-generated features include the following and number of followers, account age, reputation, FF rating and number of tweets. Content-based features include the number of retweets, the number of URLs, the number of replies and the bidirectional spread, characters and numbers, and spam words. Home page is shown in Fig.2



Fig-2.Home Page

## Modules:

This project contains 2 modules

1. User
2. Tweet Server(Admin)

### 1.User:

- If a user has already an account then he can login with those credentials otherwise he has to register a new account .
- User can register new account by providing all the details shown in the fig 3.
- User has to provide all the details shown above to register.
- Once the user can registered he has to wait until the admin (tweet server) should authorize the new user.
- All the fields are mandatory in the registration process.

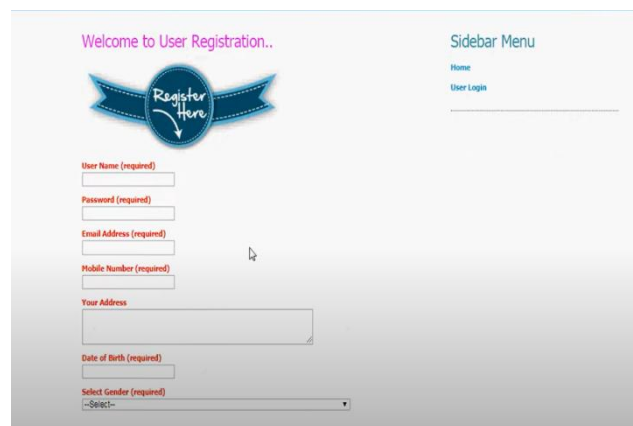


Fig-3.Registration Page

- After admin can authorize the user then he can login with his credentials as shown in Fig 4.



Fig-4.Login Page

- Once the user successfully logged in he can perform different operations like viewing profile, search for friends, create tweets, view friends etc. as shown in Fig.5
- User can create tweets by clicking on create tweet and he can search for friends with their names.
- User search for friend and he can put a request to him. He has to wait until his friend's response.(Accept or decline)

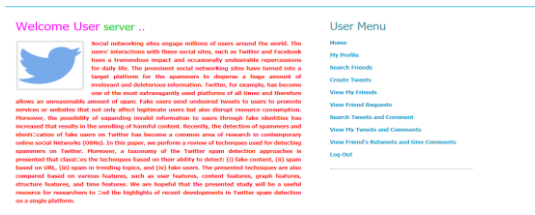


Fig-5.User Home Page

- User can view his profile by clicking “My Profile” as shown in Fig.6

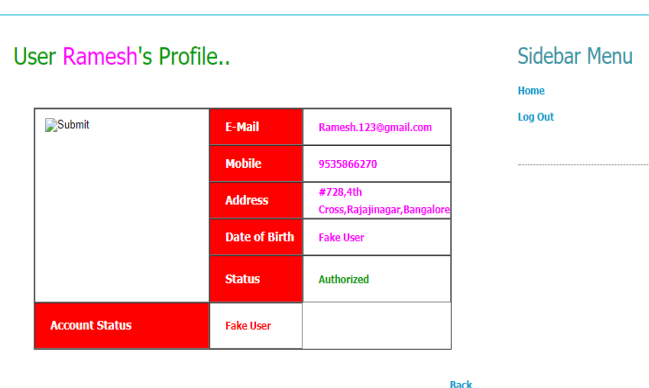


Fig-6.User profile

- User can search for friends by clicking on search friends
- User can search for friends with username or mail id or first name etc..

- The searched friends results will be displayed in “Results found” section at the bottom as shown in Fig. 7

Search Friends..

Enter Friend Name :

Search

Back

Results Found.. ..

Back

Fig-7.Search Friends

## 2.Tweet Server (Admin)

- Tweet server can maintain all the operations performed by the user.
- Admin can authorize the users after they are get registered.
- Admin can filter user the users based in their activities.
- Server has its own username and password.

Admin can perform operations like :

- View and Authorize users
- Add and view Spam Filters
- View All user Posted Tweets
- View All User Tweets Based on URLs
- View Friend Request and Response

Welcome to Tweet Server Login Page..

Login

Server Name (required)

Password (required)





Login

Back

Fig-8.Server Login

- Server can monitor all the tweets posted by the users as shown in Fig. 8
- It will display tweet image, tweet name, tweet description and time and date of tweet posted as shown in Fig. 9

All User Posted Tweets..

ID	Tweet Image	Tweet Name	Tweet Description	Tweet them	Date	Posted By
3		Dell_Laptop	Dell laptop is one of the best laptops which is manufactured by Dell Organisation.	To know about dell laptop.	31/07/2019 12:06:02	Kannan
5		Hp_Printer	Hp Printer is one of the best printer in the world and manufactured by HP.	To print and scan, fax and all in one facilities.	31/07/2019 17:19:47	Manjunath
2		HP_Laptop	The Hewlett Packard Company or Hewlett-Packard was an American multinational information which is in the manufacturing of HP Laptop and Desktops.	To know about HP Laptop	30/07/2019 17:18:14	Ramesh
4		Sandisk_Pendrive	Sandisk Pendrive is a general purpose storage devices which is manufactured by Sandisk.	This device is used to store raw data.	31/07/2019 18:29:49	Uma

Back

Fig-9.Tweets Posted

- Server can authorize the user account and can able to see all fake and authorized user.
  - Admin can filter the normal users and fake users based on the tweet content. If the tweet contains any illegal and offensive words it will automatically falls into offensive section and the user can will be marked as a suspicious. We can also see the fake user identification results and fake tweets results in the form of bar charts. The chart is drawn between the number of re-tweets posted and the user as shown in Fig 10.

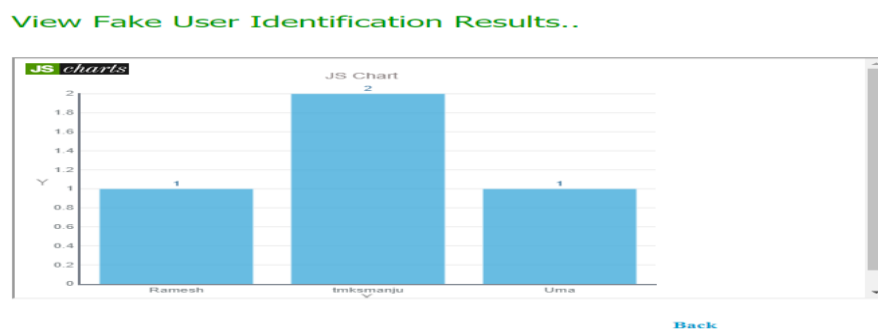


Fig-10.Fake user identification Results

## 4.CONCLUSION

In this paper, we reviewed the strategies used to find spam users on Twitter. In addition, we have also introduced the taxonomy of the Twitter spam Detection method and classified them as fraudulent content discovery, spam-based URL detection, spam detection in headlines, and illegal user access methods. We also compare strategies presented based on all aspects such as user characteristics, content features, graph features, layout features, and time elements. In addition, strategies are also used to target their stated objectives and data. It is expected that the updated version will help researchers find information on ways



to get Twitter spam in a more integrated way. Another related topic to be investigated is the identification of rumors on social media. Although some studies based on mathematical methods have already been done to find the sources of rumors, more informative methods, such as, methods for communicating with forums, can be used because of their proven functionality.

## REFERENCES

- [1] M. Babcock, R. A. V. Cox, 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.
- [2] F. Benevenuto, G. Magno, T. Rodrigues, V. Almeida, "Detecting spammers on Twitter", *Proc. Collaboration Electron. Messaging Anti-Abuse Spam Conf. (CEAS)*, vol. 6, pp. 12, Jul. 2010.
- [3] C. Chen, Y. Wang, J. Zhang, Y. Xiang, W. Zhou, G. Min, "Statistical features-based real-time detection of drifted Twitter spam", *IEEE Trans. Inf. Forensics Security*, vol. 12, pp. 914-925, Apr. .
- [4] C. Chen, J. Zhang, Y. Xie, Y. Xiang, W. Zhou, M. M. Hassan, A. AlElaiwi, 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
- [5] B. Erçahin, Ö. Aktaş, D. Kiliç, C. Akyol, "Twitter fake account detection", *Proc. Int. Conf. Comput. Sci. Eng. (UBMK)*, pp. 388-392, Oct. 2017
- [6] S. Gharge, M. Chavan, "An integrated approach for malicious tweets detection using NLP", *Proc. Int. Conf. Inventive Commun. Comput. Technol. (ICICCT)*, pp. 435-438, Mar. 2017.
- [7] A. Gupta, H. Lamba, P. Kumaraguru, "1.00 per RT #BostonMarathon #prayforboston: Analyzing fake content on Twitter", *Proc. eCrime Researchers Summit (eCRS)*, pp. 1-12, 2013
- [8] M. U. S. Khan, M. Ali, A. Abbas, S. U. Khan, 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
- [9] M. Mateen, M. A. Iqbal, M. Aleem, M. A. Islam, "A hybrid approach for spam detection for Twitter", *Proc. 14th Int. Bhurban Conf. Appl. Sci. Technol. (IBCAST)*, pp. 466-471, Jan. 2017.
- [10] S. J. Soman, "A survey on behaviors exhibited by spammers in popular social media networks", *Proc. Int. Conf. Circuit Power Comput. Technol. (ICCPCT)*, pp. 1-6, Mar. 2016.