SPAMMER DETECTION AND FAKE USER IDENTIFICATION ON SOCIAL NETWORKS

Ayesha Fatima, H.No: 21S41D5821,Mtech (Cse), Department Of Computer Science, Vaageswari College Of Engineering Thimmapur,Karimnagar, Telangana, India, Email-Id: ayeshafatimah567@gmail.com.

E Srikanth Reddy,Associate Professor, Department Of Computer Science, Vaageswari College Of Engineering Thimmapur,Karimnagar, Telangana, India,Email-Id: bapuji.vala@gmail.com.

ABSTRACT

Millions of people worldwide utilise social networking services. The way people connect on social media platforms like Facebook and Twitter has a significant influence on everyday life, often in an unfavourable way. Prominent social media platforms have become a target for spammers who want to spread a great deal of harmful and useless content. For example, Twitter has grown to be one of the most widely utilised platforms ever, which means that it accepts an excessive quantity of spam. False users spam people with unwanted tweets promoting websites or services that negatively impact real users and waste resources. Furthermore, there is now a greater chance of providing users with false identities and inaccurate information, which may lead to the unrolling of hazardous material. The identification of phoney users and the detection of spammers on Twitter have recently gained popularity as study topics in modern online social networks (OSNs). We conduct an overview of methods for identifying spammers on Twitter in this research. Additionally, a taxonomy of Twitter spam detection methods is provided, which groups the methods according to how well they can identify false users, fake content, URL-based spam, trending topic spam, and spam in general. Additionally, a comparison of the methods is made according to a number of factors, including user, content, graph, structure, and temporal aspects. With the help of this study, researchers should be able to locate the most noteworthy advancements in Twitter spam detection on one convenient platform.

Index Terms:— spammers, online social networks (OSNs), , fake content, URL-based spam, trending topic spam.

I.INTRODUCTION

1.1 MOTIVATION:

In the real world, these dangerous tactics used by spammers severely damage communities. The goals of Twitter spammers include disseminating false information, rumours, fake news, and impromptu messages. Spammers use a variety of techniques, including adverts, to further their malevolent goals. They maintain several mailing lists and then send out spam messages at random to further their agendas. The original users—also referred to as non-spammers—are disturbed by these actions. Furthermore, it also damages the OSN platforms' reputation. Therefore, in order to counteract spammers' destructive behaviours, it is essential to build a technique to identify spammers [3]. Numerous studies have been conducted in the field of Twitter spam detection. A few polls on phoney user identities from Twitter have also been conducted in order to cover the state-of-the-art at this time. A review of novel approaches and strategies for Twitter spam detection is given by Tingmin et al. [4]. A comparison of the existing methods is shown in the survey mentioned above. Conversely, the authors of [5] carried out a study about the various actions shown by spammers on the social media platform Twitter. A review of the literature that acknowledges the presence of spammers on the social network Twitter is also provided by the research. There is a void in the body of literature despite the existence of all the investigations. Thus, in order to close the gap, we examine the most recent developments in Twitter's spammer and false user identification systems. Additionally, this survey aims to provide a thorough explanation of current advancements in the field and offers a taxonomy of Twitter spam detection techniques.

1.2 PROBLEM DEFINITION:

Issue of spam on online social networks (OSNs), with a particular emphasis on Twitter. Spamming activities, including the dissemination of false information, rumors, fake news, and unwanted messages, pose significant challenges to genuine users and the integrity of OSN platforms. These activities not only disrupt user experiences but also damage the reputation of OSNs as reliable sources of information.

Given the prominence of Twitter as a platform for real-time information sharing, the need to identify and mitigate spamming behaviors on this platform is paramount. However, current spam detection techniques face several challenges, including the evolving tactics employed by spammers and the difficulty in distinguishing between genuine and spam accounts.

To address this problem effectively, there is a critical need for research and evaluation of user behavior on OSNs, specifically focusing on Twitter. Existing studies and surveys provide valuable insights into spam detection methods, but there remains a gap in the literature regarding recent developments and emerging trends in combating spam on Twitter.

Therefore, the overarching goal is to develop and evaluate robust spam detection techniques tailored to the unique characteristics of Twitter and other OSNs. This entails exploring innovative approaches to identify spammers and false users, leveraging advancements in machine learning, natural language processing, and network analysis. By doing so, the aim is to enhance the trustworthiness and reliability of OSNs, ultimately improving the user experience and preserving the integrity of online communication channels.

II.LITERATURE SURVEY

- C. Chen, Y. Wang, J. Zhang, Y. Xiang, W. Zhou, and G. Min proposed a twitter spam has become a critical problem nowadays. Recent works focus on applying machine learning techniques for Twitter spam detection, which make use of the statistical features of tweets. In our labeled tweets data set, however, we observe that the statistical properties of spam tweets vary over time, and thus, the performance of existing machine learning-based classifiers decreases. This issue is referred to as “Twitter Spam Drift”. In order to tackle this problem, we first carry out a deep analysis on the statistical features of one million spam tweets and one million non-spam tweets, and then propose a novel Lfun scheme. The proposed scheme can discover “changed” spam tweets from unlabeled tweets and incorporate them into classifier's training process. A number of experiments are performed to evaluate the proposed scheme. The results show that our proposed Lfun scheme can significantly improve the spam detection accuracy in real-world scenarios.

- C. Buntain and J. Golbeck describe about information quality in social media is an increasingly important issue, but web-scale data hinders experts' ability to assess and correct much of the inaccurate content, or "fake news," present in these platforms. This paper develops a method for automating fake news detection on Twitter by learning to predict accuracy assessments in two credibility-focused Twitter datasets: CREDBANK, a crowdsourced dataset of accuracy assessments for events in Twitter, and PHEME, a dataset of potential rumors in Twitter and journalistic assessments of their accuracies. We apply this method to Twitter content sourced from BuzzFeed's fake news dataset and show models trained against crowdsourced workers outperform models based on journalists' assessment and models trained on a pooled dataset of both crowdsourced workers and journalists. All three datasets, aligned into a uniform format, are also publicly available. A feature analysis then identifies features that are most predictive for crowdsourced and journalistic accuracy assessments, results of which are consistent with prior work. We close with a discussion contrasting accuracy and credibility and why models of non-experts outperform models of journalists for fake news detection in Twitter.

- C. Chen, J. Zhang, Y. Xie, Y. Xiang,W. Zhou, M. M. Hassan, A. AlElaiwi, and M. Alrubaian proposed the popularity of Twitter attracts more and more spammers. Spammers send unwanted tweets to Twitter users to promote websites or services, which are harmful to normal users. In order to stop spammers, researchers have proposed a number of mechanisms. The focus of recent works is on the application of machine learning techniques into Twitter spam detection. However, tweets are retrieved in a streaming way, and Twitter provides the Streaming API for developers and researchers to access public tweets in real time. There lacks a performance evaluation of existing machine learning-based streaming spam detection methods. In this paper, we bridged the gap by carrying out a performance evaluation, which was from three different aspects of data, feature, and model. A big ground-truth of over 600 million public tweets was created by using a commercial URL-based security tool. For real-time spam detection, we further extracted 12 lightweight features for tweet representation. Spam detection was then transformed to a binary classification problem in the feature space and can be solved by conventional machine learning algorithms. We evaluated the impact of different factors to the spam detection performance, which included spam to nonspam ratio, feature discretization, training data size, data sampling, time-related data, and machine learning algorithms. The results show the streaming spam tweet detection is still a big challenge and a robust detection technique should take into account the three aspects of data, feature, and model.

- F. Fathaliani and M. Bouguessa presented a survey on identifying spammers in social networks from a mixture modeling perspective, based on which we devise a principled unsupervised approach to detect spammers. In our approach, we first represent each user of the social network with a feature vector that reflects its behaviour and interactions with other participants. Next, based on the estimated users feature vectors, we propose a statistical framework that uses the Dirichlet distribution in order to identify spammers. The proposed approach is able to automatically discriminate between spammers and legitimate users, while existing unsupervised approaches require human intervention in order to set informal threshold parameters to detect spammers. Furthermore, our approach is general in the sense that it can be applied to different online social sites. To demonstrate the suitability of the proposed method, we conducted experiments on real data extracted from Instagram and Twitter.

- C. Meda, E. Ragusa, C. Gianoglio, R. Zunino, A. Ottaviano, E. Scillia, and R. Surlinelli proposed a law

Enforcement Agencies cover a crucial role in the analysis of open data and need effective techniques to filter troublesome information. In a real scenario, Law Enforcement Agencies analyze Social Networks, i.e. Twitter, monitoring events and profiling accounts. Unfortunately, between the huge amount of internet users, there are people that use microblogs for harassing other people or spreading malicious contents. Users' classification and spammers' identification is a useful technique for relieve Twitter traffic from uninformative content. This work proposes a framework that exploits a non-uniform feature sampling inside a gray box Machine Learning System, using a variant of the Random Forests Algorithm to identify spammers inside Twitter traffic. Experiments are made on a popular Twitter dataset and on a new dataset of Twitter users. The new provided Twitter dataset is made up of users labeled as spammers or legitimate users, described by 54 features. Experimental results demonstrate the effectiveness of enriched feature sampling method.

III. EXISTING SYSTEM

- Several studies have contributed to the understanding of Twitter spam detection and the behaviors exhibited by spammers on the platform. Tingminet al. conducted a survey that presents a comparative study of current approaches to identifying Twitter spam, offering insights into new methods and techniques in this field. Meanwhile, S. J. Somanet al. conducted a survey focusing on the various behaviors displayed by spammers on Twitter and provided a literature review that acknowledges the presence of spammers within the platform.
- Despite the wealth of existing studies, there remains a gap in the literature surrounding Twitter spam detection and the identification of fake users. To address this gap, there is a need to review the state-of-the-art in spammer detection and fake user identification on Twitter comprehensively. By synthesizing and analyzing the latest advancements and methodologies in this area, researchers can contribute to the development of more effective strategies for combating spam and

enhancing the integrity of the Twitter social network. Such a review would provide valuable insights into emerging trends and challenges, guiding future research efforts in mitigating the detrimental effects of spam on online social platforms like Twitter.

DISADVANTAGES OF EXISTING SYSTEM:

- ❖ No efficient methods used.
- ❖ No real time datas used.
- ❖ More complex

IV PROPOSED SYSTEM:

- This paper aims to comprehensively explore various approaches to detecting spam on Twitter and establish a taxonomy by categorizing these approaches into distinct classifications. To facilitate this classification, four primary means of identifying spammers have been identified: (i) analyzing fake content, (ii) employing URL-based spam detection methods, (iii) detecting spam within trending topics, and (iv) identifying fake user accounts.
- Furthermore, the analysis conducted in this study reveals that several machine learning-based techniques show promise in effectively identifying spam on Twitter. However, the selection of the most suitable techniques and methods is contingent upon the availability and quality of the data being utilized. Therefore, this paper not only elucidates the different avenues for spam detection on Twitter but also underscores the importance of leveraging appropriate machine learning methodologies tailored to the specific characteristics of the available data. Through this comprehensive examination, the paper aims to contribute to the development of robust strategies for combating spam and enhancing the integrity of the Twitter platform.

ADVANTAGES OF PROPOSED SYSTEM:

- ❖ This study includes the comparison of various previous methodologies proposed using different datasets and with different characteristics and accomplishments.
- ❖ Tested with real time data.

V. SYSTEM DESIGN

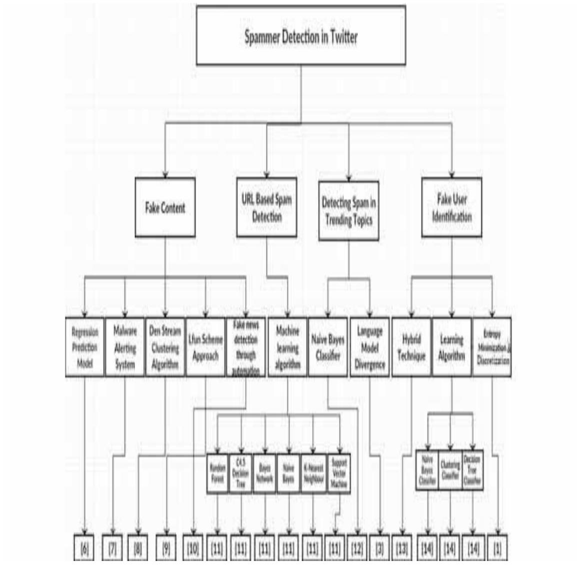


Fig1: Architecture of system.

VI.MODULE DESCRIPTION:

Admin Module:

Home: Dashboard providing an overview of system activities.

View Users: Access to view registered users and their details.

Add Filters: Ability to set filters or criteria for detecting fake users or data.

View Fake User: View a list of identified fake users.

View Fake Data: Access to data flagged as fake or suspicious.

Logout: Option to log out from the admin panel.

User Module:

Home: User dashboard displaying relevant information and updates. Profile: Manage user profile settings and information.

Compose Tweet: Create and post tweets or messages on the platform. Trending: View trending topics or hashtags within the platform.

Following: Manage users or accounts followed by the user. Follower: View users who are following the user's account

. Logout: Option for users to log out from their accounts.

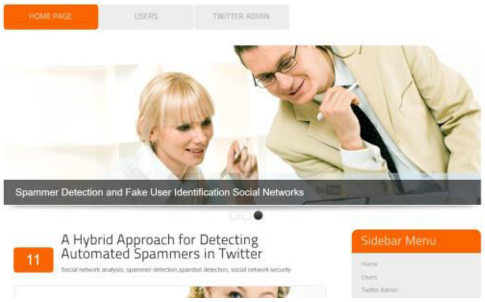
Spammer Module:

Compose Tweet: Ability to create and post tweets or messages.

View Tweets: Access to view all tweets posted by the spammer account.

VII. RESULT:

Home Page



Spammer Detection and Fake User Identification Social Networks

11 A Hybrid Approach for Detecting Automated Spammers in Twitter

Social network analysis, spammer detection,spambot detection, social network security

User Page

HOME PAGE
USERS
TWITTER ADMIN
social media
Spammer Detection and Fake User Identification Social Networks

VIII. CONCLUSION

In this study, we conducted a review of methods for Twitter spam detection. Furthermore, we unveiled a taxonomy of Twitter spam detection strategies, classifying them into four categories: spam identification in trending topics, URL-based spam detection, false content detection, and fake user detection methods. Additionally, we contrasted the methods that were offered according to a number of factors, including user, content, graph, structure, and temporal aspects. Additionally, a comparison of the methods' stated objectives and datasets was conducted. The review that is being given is expected to assist academics in finding information on the most recent methods for detecting spam on Twitter in a centralised manner. Even while methods for identifying false users and detecting spam on Twitter have become more successful and efficient [34], there are still some unresolved issues that need more study. The following is a basic summary of the issues:The detection of false news on social media networks is a problem that requires investigation due to the grave consequences that such news may have on both an individual and a group level [25]. Finding the origins of rumours on social media is another related issue worth looking into. While some statistically based research have previously been carried out to identify the origins of rumours, more advanced techniques, such as social network-based approaches, may be used due to their shown efficacy.

IX. FUTURE ENHANCEMENT

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. 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.

X. 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, AntiAbuse 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.