Spammer Detection and Fake User Identification

T. Amulya Reddy[1], M. Sandhya[2], P. Sai Kiran Reddy[3], R. Sandeep Naik[4], Mr.S.T. Saravanan[5], Mr.G.Rakesh Reddy [6]

[1,2,3,4]Students, Department of CSE(CYBER SECURITY), Sphoorthy Engineering College, Hyderabad, India.

[5]Assistant Professor, Department of CSE(CYBER SECURITY), Sphoorthy Engineering College, Hyderabad, India.

[6]Professor & HOD, Department of CSE(CYBER SECURITY), Sphoorthy Engineering College, Hyderabad, India.

***Abstract*—Social Networking sites engage millions of users around the world. The user's 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 therefore allows an unreasonable amount of spam fake users send undesired tweets to users to promote services or websites but not only effect legitimate users but also disrupt resource assumption. Moreover, the possibility of expanding invalid information to users through fake identities has increased the results in the unrolling of harmful content. Recently, the detection of spammers and identification of fake users and twitter and face book has become a common area of research in contemporary online social networks (OSNs).**

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. 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. With the evolution of OSNs, the need to study and analyze users' behaviors in online social platforms has intensity. 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. 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.

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.

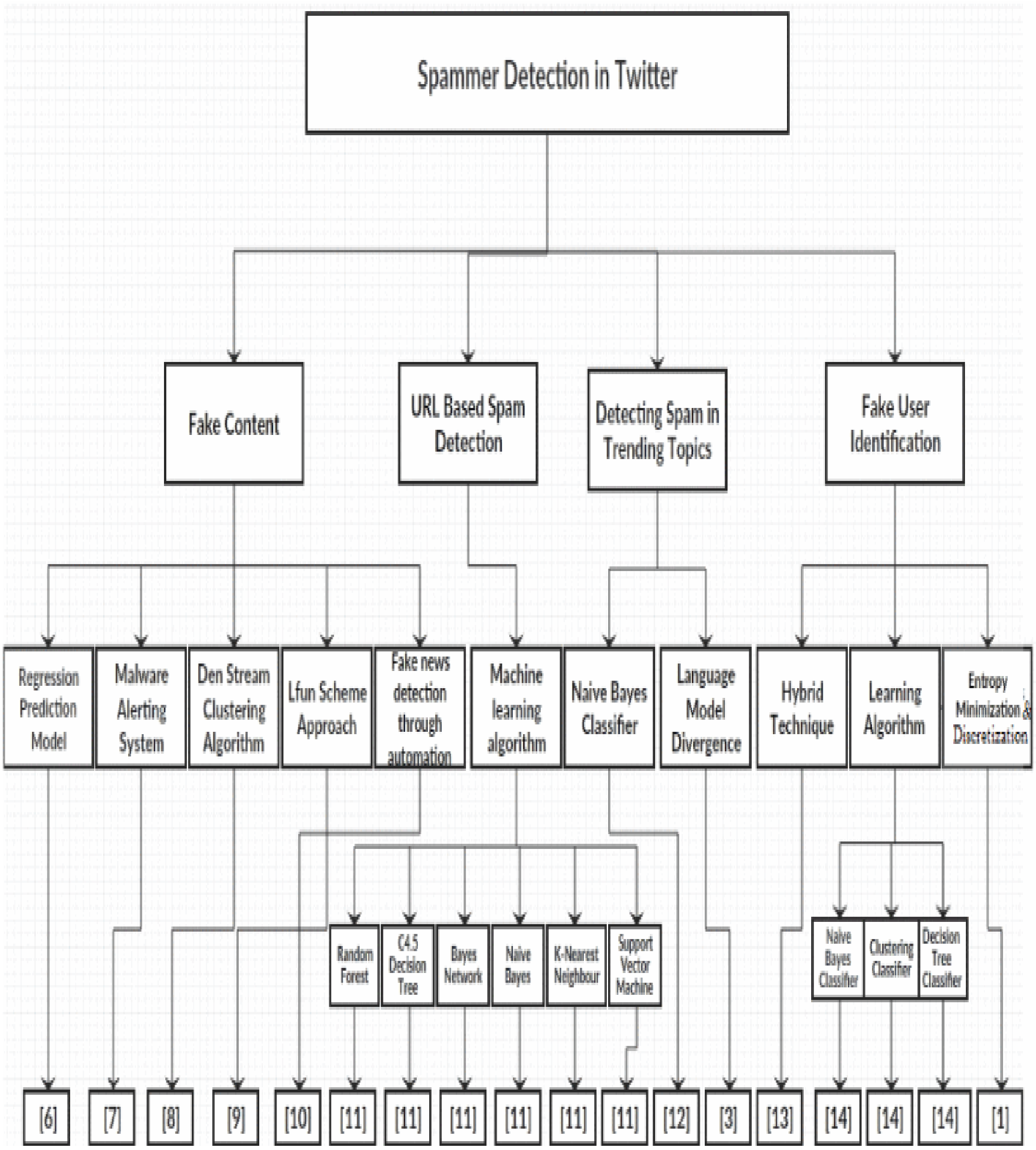

FIG. ARCHITECTURE

II. RELATED WORK

In recent years, a considerable amount of research has been conducted in the area of spammer detection and fake user identification, especially in the context of Online Social Networks (OSNs) like Twitter and Facebook. As these platforms are increasingly exploited by malicious users to spread fake content, advertisements, phishing attacks, and disinformation, detecting spammers and fake users has become a critical cybersecurity challenge.

Benevenuto et al. (2010) proposed a feature-based classification approach that considers user behavior such as tweet frequency, follower-following ratio, and use of URLs in tweets to distinguish between spammers and legitimate users. Similarly, Erçahin et al. (2017) developed a machine learning-based method that focuses on metadata and user activity patterns for fake account detection.

Wu et al. (2018) presented a comprehensive survey of Twitter spam detection techniques, categorizing them into graph-based, content-based, and hybrid approaches. Their study emphasized the effectiveness of combining network structure with textual features for better detection accuracy.

Another approach by Gupta et al. (2013) focused on identifying fake content and analyzing its spread during crisis events such as the Boston Marathon bombings. They demonstrated how spammers exploit trending hashtags to reach larger audiences and mislead users.

More recently, studies have incorporated Artificial Neural Networks (ANNs) and deep learning techniques to improve spammer classification accuracy. These models learn complex patterns from user data, such as posting behavior, interaction with other users, and account metadata. Meda et al. applied a Random Forest classifier with non-uniform feature sampling, achieving high accuracy in distinguishing fake users based on selected behavioral and structural features

III. OBJECTIVE

The primary objective of this project is to design and implement a robust and intelligent system capable of detecting spammers and identifying fake user accounts on online social networking platforms such as Twitter and Facebook. The project aims to address the increasing security threats posed by fake profiles, spam messages, and malicious user activities by leveraging machine learning techniques—specifically, Artificial Neural Networks (ANNs).

The specific goals of the project are:

1.To analyze behavioral and profile-based features (e.g., account age, follower/following ratio, posting frequency, URL usage) to distinguish between genuine and fake users.

2.To detect and classify spam content in tweets or posts using natural language and metadata analysis.

3.To develop a classification model using ANN that can learn from historical data and accurately predict whether a user is legitimate or a spammer.

4.To implement an admin-user web application where admins can monitor users, view spam detection results, and track fake login attempts.

5.To enhance the security and trustworthiness of social platforms by reducing the influence of spammers and preventing misinformation or fraudulent interactions.

IV. PROPOSED METHODOLOGY

The proposed system aims to detect spammers and identify fake users on social media platforms using Artificial Neural Networks (ANN). The methodology involves the following key steps:

1.Data Collection: User data such as account age, user activity, follower/following count, post content, and IP address is collected from social media sources.

2.Data Preprocessing: The collected data is cleaned and transformed. Categorical variables are converted into numerical format, and missing values are handled to prepare the dataset for model training.

3.Feature Selection: Important features like number of tweets, presence of URLs, profile completeness, friend count, and post frequency are selected to improve model accuracy.

4.Model Development: An ANN model is built and trained using labeled data to classify accounts as either genuine or fake. The model learns patterns from training data and is later tested on unseen data for validation.

5.System Integration: A web-based interface is developed using Python Flask. It allows users to register, login, and view their spam detection results. Admins can upload datasets, train the model, and monitor fake login attempts.

6.Fake Login Monitoring: IP addresses and login timestamps are recorded. Flask-Limiter is used to detect and block suspicious or repeated login attempts.

7.Result Analysis: The performance of the system is evaluated using accuracy, precision, recall, and F1-score. The results help in understanding how effectively the model identifies fake accounts and spam behavior.

V. RESULT

The proposed system successfully detects spammers and identifies fake users on social media platforms using an Artificial Neural Network (ANN). After training the model with a dataset containing both genuine and fake user accounts, the system was tested on new data to evaluate its performance.

The results show that the ANN model can accurately classify users based on their behavior, such as account age, number of followers, post frequency, and presence of spam-related content. The following outcomes were observed:

High Accuracy: The trained model achieved a high accuracy rate in detecting fake users.

Effective Detection: The system could differentiate between real and fake accounts using profile-based and activity-based features.

Real-Time Monitoring: Fake login attempts were successfully recorded with IP address and timestamp, and rate limiting was used to block repeated suspicious access.

User & Admin Interfaces: Both users and admins could interact with the system via a web-based dashboard. Users could register, log in, and view their detection status, while admins could monitor system usage and view fake login attempts.

VI. CONCLUSION

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, there are still certain open areas that require considerable attention by the researchers. The issues are briery highlighted as under: False news identification on social media networks is an issuethat 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.

REFERENCES

[1] B. Erçahin, Ö. Akta³, D. Kilinç, and C. Akyol, ``Twitter fake account detection,'' in Proc. Int. Conf. Comput. Sci. Eng. (UBMK), Oct. 2017, pp. 388392.

[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. 435438.

[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. 265284, Jul. 2018.

[5] S. J. Soman, ``A survey on behaviors exhibited by spammers in popular social media networks,'' in Proc. Int. Conf. Circuit, Power Comput. Tech- nol. (ICCPCT), Mar. 2016, pp. 16.

[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. 112.