The Nucleus ***60***, *No. 2 (2023) 199-204*

www.thenucleuspak.org.pk

The Nucleus

ISSN 0029-5698 (Print)
ISSN 2306-6539 (Online)

A Holistic Approach for Detecting Socialbots on Twitter: Integration of Diverse Features

Muhammad Owais*, Muhammad Shoaib, Muhammad Waseem

Computer Science, University of Engineering & Technology 54890, Punjab, Pakistan

ABSTRACT

The usage of social media platforms has grown, offering individuals diverse avenues for communication, expressing opinions and sharing online content. However, this surge has also given rise to the emergence of social bots, which are programmed accounts designed to imitate human behavior. Such bots possess the capability to disseminate false information, manipulate financial markets, aid terrorism, and disrupt democratic processes. To tackle this issue, various approaches have been utilized to detect social bots, including approaches based on profiles, time patterns, content analysis, behavior, and network characteristics. However, neither of the approaches effectively combines all these features to implement social bot detection comprehensively. This paper introduces an ensemble methodology that merges profile, behavioral, temporal, network, graph, and content-based attributes, culminating in a comprehensive model for discerning social bots on the Twitter platform. We utilize the Twibot-22 dataset for conducting experiments and evaluate the performance of our approach against benchmark models. The XGBoost model, with an accuracy of 0.898, exhibited superior performance compared to the benchmark models. This research contributes to the continuous endeavor focused on safeguarding the authenticity of tweet content and mitigating the risks associated with social bots on social networks.

Keywords: *Machine Learning, Social bot, XGBoost, Twitter, detection*

1. Introduction

Online social networking platforms have revolutionized communication by providing a platform for individuals to share their thoughts and personal information regarding current events [1]. A significant portion of the global population actively engages with one or more OSNs, with Twitter and Facebook being particularly popular choices. In fact, as of December 2022, Twitter boasted a staggering user base of over 368 million monthly active users worldwide, and this number is projected to continue growing in the coming years [2]. However, the widespread adoption of Twitter has also brought about an upsurge in the presence of Twitter bots, which are automated accounts designed to mimic human behavior [3]. These social bots, comprising approximately 9 to 15 percent of active Twitter accounts, serve various purposes but are often associated with malicious intentions [4]. Their activities encompass spreading false information [5], stock market manipulation [6], endorsing terrorist activity [7], disseminating explicit content [8] and even interfering in the 2016 United States presidential election [9]. Consequently, the proliferation of social bots undermines the authenticity of online trends, erodes democratic processes and poses significant harm to society.

Socialbots use contemporary techniques to imitate human qualities to hide their true identity online and avoid being detected. Therefore, it is essential to detect Twitter bots to maintain the genuineness of tweet content. In pursuit of this objective, experts utilize various methods to distinguish between human users and programmed accounts. Machine learning algorithms have quickly advanced, giving researchers the ability to detect the presence of bots and suspect sources. In the past, machine learning algorithms, specifically supervised learning, utilized various manually created characteristics to identify bots [10]. Prior studies have shown enhanced performance in social bot identification through recent advancements in machine learning methods such as profile metadata-based methods [1], graph network-based approaches [11], behavior-based techniques [12], as well as temporal modeling-based methods [3]. Nevertheless, these previous approaches have certain limitations. For instance, the profile-based technique is vulnerable to spoofing, as it can be manipulated to resemble human behavior more closely [13]. Similarly, the graph partitioning-based method primarily focuses on analyzing the network information and the behavioral patterns of programmed accounts, which makes it ineffective in detecting bots that successfully establish connections with regular users [14]. Moreover, the limitations of these techniques stem from their dependence on textual, temporal, and profiling data. On the other hand, behavioral and temporal-based research methodologies have been utilized to detect bots. Nevertheless, unlike alternative methods, these models solely detect a particular bot category based on user information [15].

Beforehand, different methods including profile metadata-based, behavioral-based, temporal-based, network-based, graph-based and content-based techniques have been separately utilized to detect social bots. However, to the best of our knowledge, none of these approaches have integrated all the mentioned features and capabilities in combination.

In our proposed approach, the model was trained on the Twibot-22 dataset. The unique aspect of this research lies in the incorporation of various features such as profile-based, behavioral-based, temporal-based, graph network-based and content-based techniques to identify social bots. When evaluating the model's performance on the Twibot-22 dataset, it exhibited superior results compared to benchmark models.

Section 2 provides an overview of the relevant research and the motivation behind detecting social bots on Twitter. Section 3 offers a sophisticated method for locating social bots on Twitter using a thorough multi-feature analysis. Transitioning to Section 4, we provide the resources and

*Corresponding author: owaiskhana635@gmail.com

techniques used in this investigation. In Section 5, we explore the results obtained from the conducted experiments, assessing the efficacy of the proposed approach. Lastly, Section 6 delivers a thorough summary of the article.

2. Related Work

Online social media platforms' programmed profiles mimic human traits to manipulate public perception or spread inaccurate information. Several machine-learning techniques have been suggested to curb the impact of social bot profiles. The profile-based framework was created to detect social bots, achieving an accuracy of 89% through the utilization of a random forest algorithm and six hybrid-selected features [16]. Similarly, the detection of a bot account through profile metadata, incorporating an extra tree classifier for feature selection and the weight of evidence encoding attained an impressive 88% accuracy with random forest [17].

Furthermore, to detect social bots, the authors [3] analyzed users' online behavior as if it were DNA sequences. They utilized the Term Frequency-Inverse Document Frequency (tf-idf) a numerical method adopted to compute entropy on segments, aiming to identify the pattern linked to bots. This approach achieved an accuracy of 94%. GANBOT [12] employed LSTM as a common interface to link the discriminator and, generator allowing for the assessment of online users' behavioral patterns. This approach demonstrated promising results on the Cresci dataset. Similar to the other bot detection methods, [18] utilized a Network graph approach to identify social botnet communities, taking into account behavior similarity and participant trust values. This strategy employed deep autoencoder technology. Additionally, the study delved into the interrelation between the social and collaborative network of the account as part of the endeavor to identify social bots.

This involved applying a semi-supervised learning algorithm that incorporated label propagation and an iterative traversal of the users' proximity graph in sequence [11].

Moreover, the adoption of content-based identification for social bots has also taken place. The researcher [19] employed CNN and LSTM in conjunction with BERT to classify tweets as either generated by programmed accounts or originating from real users. In addition, a fabricated prediction algorithm was utilized to calculate the sentiment score of each piece of content, enabling the identification of social bot accounts with a 97% accuracy rate [20].

In the study by Kosmajac et al. [21], a temporal-based approach to detect social bot accounts has been implemented. The approach utilizes statistical diversity measures to examine the variation in user behavior during a specific period. Additionally, they developed a method to identify malicious tweets by analyzing the pairwise similarity of individual retweeting behavior, focusing on temporal features [22]. While various techniques have been utilized for social bot detection, encompassing aspects such as profile analysis, timestamps, content examination, behavioral assessment, network relationships, and graph analysis, a research gap exists in the lack of emphasis, in prior studies, on integrating multiple features like metadata, content, interactions, and timing for accurate social bot identification [14]. This study uniquely addresses this gap by comprehensively incorporating all available features - profiles, content, timestamps, behavior, and network connections - to significantly advance the precision of social bot identification.

3. Proposed Methodology

Social media platforms often employ social bot detection techniques to distinguish between real human users and automated bots. The diverse functions of these bots, including spreading propaganda and overwhelming users with irrelevant content, present a challenge to preserving the integrity of tweet content. As a result, detecting social bot accounts would greatly benefit the preservation of content integrity. To address the issue of social bot detection, we have put forth an approach aimed at identifying programmed accounts. The system architecture is depicted in Figure 1.

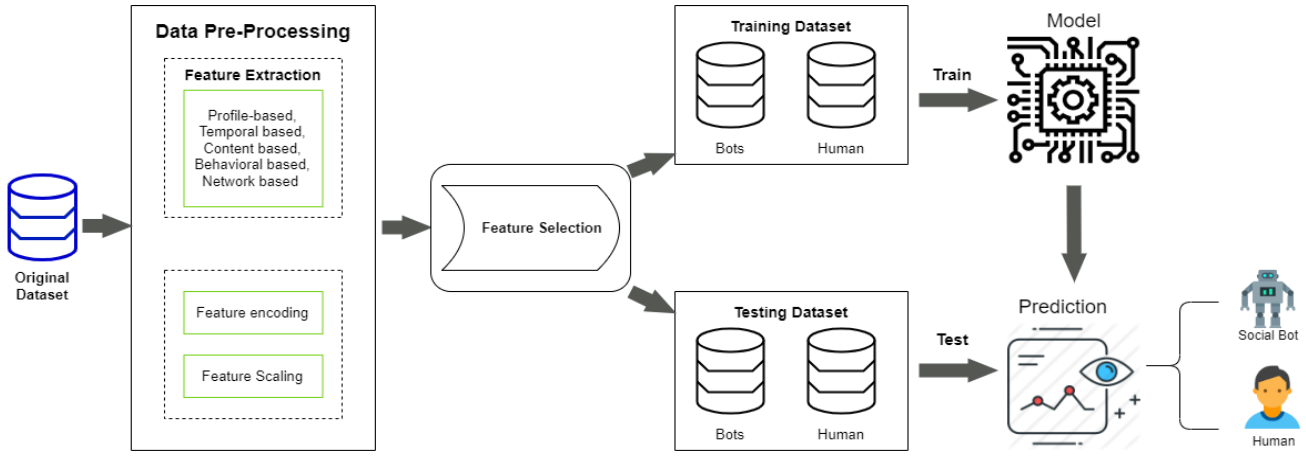

Figure 1: Architectural Framework

The model was trained and social bots were identified using a hybrid features-based method. The model was trained using a publicly accessible dataset. Before training, the features were extracted to analyze their high correlation. The closely interrelated attributes, obtained by integrating a comprehensive set of features - profiles, content, timestamps, behavior and network connections - to markedly enhance the accuracy of social bot identification, were subsequently chosen for categorical encoding and feature scaling. The chosen classifier utilized these selected features as input and the evaluation parameters were employed to assess their performance.

4. Materials and Methods

The paper's section discusses an approach for identifying social bot accounts, consisting of multiple phases outlined below.

4.1) Dataset

This research paper addressed the problem of Twitter bot localization by utilizing the publicly accessible Twibot-22 datasets [23]. The datasets consist of 1,000,000 diverse users, comprising 860,057 humans and 139,943 bots. Table 1 illustrates that the dataset contains 4 types of entities and 14 types of relations. These entities include the user, hashtag, tweet, and list and the relations encompass pin, contain, retweet, follower, post, mention, like, reply, quote, own, follow, member, discuss, and following.

Table 1: Dataset details

Twibot-22	
Human	860,057
Bot	139,943
User	1,000,000
Tweet	86,764,167
Edge	170,185,937
Entity	User, Tweet, List, Hashtag
Relation	Follower, following, post, pin, like, mention, retweet, quote, reply, own, member, follow, contain, discuss

4.2 Data Pre-processing

Data pre-processing entails the preparation of data for analysis. The initial step consists of feature encoding; wherein categorical data is transformed into machine-learning data with significant meaning. During this process, Boolean values into a binary representation, and the encoding of the account attribute is 0 for humans and programmed accounts is 1. After encoding the features, the machine learning algorithm's efficiency is improved by scaling the numeric attributes. Scaling is an important step because numeric attributes with a diverse set of values can hinder the algorithm's speed or result in convergence on a local minimum.

To enhance data training and remove biases, numerical attributes are scaled using the standard scaler. Equation (i) enables the computation of the standardized value "y" for a certain numeric feature value "x".

$$Y = \frac{x - mean}{std} \quad (1)$$

In this context, "mean" signifies the numerical attribute's average, while "std" indicates its standard deviation.

4.3 Feature Selection

Feature selection is employed to identify and select a subset of relevant features, aiming to improve the model's performance and remove unnecessary features.

Using correlation feature selection [24], a commonly employed method to discern pertinent features, we chose the features depending on how well they correlated with the target variable. Features with a high correlation were thought to be more important and preserved, while others were dropped. As a result, we obtained a list of the top 51 features with the highest correlation, shown below.

1) Profile description: Is there a description in the profile?
2) Profile Location: Is there a location specified in the profile?
3) Profile URL: Is there a URL associated with the profile?
4) Verified: Is the profile confirmed to be authentic?
5) Bot word in name: Is there a word in the username that relates to the bot?
6) Screen name containing a bot word: Is the screen name chosen from a list of words to identify the bot?
7) Bot word in description: Is the word related to the bot from a list of words present in the description?
8) Username size: How long is the username?
9) Screen name size: How long are screen names?
10) Description length: What is the size of a description?
11) Followees count: User's account follows count
12) Followers count: User's follower count
13) Follower to Followee Ratio: Followee ratio refers to the proportion of people a user follows compared to the number of people following that user on social media platforms
14) Tweets count: Total number of tweets by user
15) Listed count: User's membership in public lists
16) Username digit count: The number of digits in the username as a whole
17) Screen name numeral count: Total digits in the screen name
18) Hashtags in username: Hashtag count within the username
19) Hashtags in description: The description's hashtag count
20) URLs in description: URLs count within the description

21) Def image: Is there a default profile image associated with the account?

22) URLs: Tweets' URLs are taken into account.

23) Protected: Does the profile have protection?

24) Pinned tweet: Is there a pinned tweet on the account?

25) Counting mentions: Mentions in the description are taken into account

26) Average tweets: Daily tweet count

27) Entropy of screen names: It is the measure of how unpredictable or random a screen name's characters are.

28) Listed growth: User membership in public lists grows daily.

29) Follower' s growth: Follower growth indicates the rate at which a user's or account's follower count increases over a specific period on social media platforms.

30) Characters: Characters count within the tweets

31) Hashtags: Hashtag count within the tweets

32) Words: Tweet word count

33) Hashtags to words: Word-to-hashtag proportion

34) Mentions: The mention count in tweets

35) Numeric characters: The numeric characters count in tweets

36) URLs to words: ratio of URLs to words

37) Times like Times like refers to the number of times a tweet has been marked as liked by other users on Twitter.

38) Times retweeted: The tweet's retweet count per post

39) Links ratio: The proportion of links within a tweet

40) Ratio of distinct linkages: The proportion of unique URLs in the tweets

41) Mention ratio: The proportion of mentions within a tweet

42) Unique mention ratio: The number of distinct mentions contained within the tweets

43) Eigen centrality: A node's importance is assessed by taking into account its links to other important nodes

44) Harmonic centrality: Calculates the mean distance to every other node within a network

45) Authority: A node's expertise or trustworthiness within a network can be quantified

46) Hub score: A node's ability to connect with other important nodes is measured

47) Degree centrality: Evaluates the number of connections associated with a node

48) Eccentricity: Calculates the farthest distance to every other node in a network

49) Page Rank: A node's significance is determined by the significance of the nodes that are connected to it

50) Coreness: measures the degree of connectedness between a node and the network's most important nodes

51) Symbols: Symbols count in tweets

4.4 Training Model

The detection of Twitter bots was accomplished by utilizing various classification algorithms, each of which was closely monitored for its performance throughout the process. In this inquiry, ineffective algorithms were removed, while retaining the effective ones. The algorithms employed included Extra-Trees Classifier, AdaBoost, Decision Trees, Random Forest, and XGBoost. Table 1 demonstrates that XGBoost surpassed the other algorithms.

Using a stratified technique, we separated the dataset into training, test, and validation sets. The dataset as a whole was made up of 70% training data, 10% test sets, and 20% validation data, respectively. The XGBoost algorithm was subsequently utilized to create decision trees sequentially, with an overall of 500 trees established. Each tree had a maximum depth of 4, while the learning rate, responsible for regulating the individual tree's impact within the sequence, was set to 0.1. The objective function was specifically configured for binary logistics to optimize the classification problem. Moreover, to avoid overfitting, the incorporation of a regularization term with an alpha parameter of 10 was crucial for managing the model's complexity. This parameter effectively penalized significant coefficients within the decision trees. In general, these contributions played a vital role in establishing a strong model that possesses the ability to effectively classify binary data while minimizing the likelihood of overfitting.

4.5 Evaluation metrics

We employ established performance metrics, including Precision, F1-score, Accuracy, and Recall, to assess the model's effectiveness. Using the formula, accuracy indicates the model's percentage of correct predictions out of all of its forecasts:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

The number of true negatives and true positives in the context is denoted by TN and TP, respectively, while the number of false negatives and false positives is denoted by FN and FP. Precision is the percentage of correct positive predictions among all positive predictions made by the model. The formula below can be used to determine the precision.

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

The recall represents the proportion of accurate positive predictions made by the model compared to the total number of positive cases. To calculate recall, use the following formula:

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

The F1-score is a calculation that merges precision and recall, determining their overall performance. The harmonic mean of recall and precision is used to derive it. A higher F1 score indicates improved recall and precision. The following formula is used to determine the F1 score.

$$F1 - score = \frac{2 * Precision * Recall}{Precision + Recall} \quad (5)$$

5. Experimental Results

The Twibot-22 training dataset was used to train the XGBoost classification model in this study and it exhibited remarkable performance. The model successfully classified 89.8% of the input data, achieving an accuracy of 0.898. The F1 score, which provides a balanced measure of precision and recall, achieved a value of 87.5%. The precision of our model stood at 87.6%, indicating a high proportion of correctly identified social bots compared to the total predicted positive cases. Additionally, the recall, or true positive rate, reached 89.9%, suggesting that the model effectively captured a significant portion of the actual positive cases in the dataset. The effectiveness of the XGBoost algorithm in social bot detection, which is crucial for preserving the integrity of social media platforms, is emphasized by these findings. A comparison of the XGBoost model's performance with those of the other models used in this investigation is shown in Table 2.

Table 2: Model performance comparison

Models	Accuracy	F1 score	Precision	Recall
Decision Tree	0.8892	0.8670	0.8701	0.8872
Extra-Trees Classifier	0.8906	0.8660	0.8722	0.8915
Random Forest	0.8921	0.8683	0.8748	0.8934
AdaBoost	0.8888	0.8648	0.8684	0.8895
XGBoost	0.8986	0.8751	0.8767	0.8997

The Gini index served as the splitting criterion in the decision tree. It selected 300 trees without any depth restrictions, yielding an accuracy of 0.8892. The Random Forest method utilized 500 trees to find the optimal split node, resulting in an accuracy of 0.8921. Similarly, the Extra Tree Classifier employed 300 trees and achieved an accuracy of 0.8906. Utilizing 1000 weak learners, a learning rate of 0.1, and using alpha as the regularization parameter, the AdaBoost model achieved an accuracy of 0.8888.

The comparison section analyzes the performance of the model and compares it to other baseline models, such as the Extra-Trees Classifier, AdaBoost, Decision Trees, Random Forest and XGBoost. We evaluated the performance of these algorithms when applied to hybrid features, including profile metadata-based, behavioral-based, temporal-based, graph network-based, and content-based features. The results demonstrate the efficiency of these particular algorithms in accurately detecting programmed accounts.

The performance comparison was conducted in terms of precision, accuracy, F1 score, and recall. XGBoost demonstrated superior performance compared to the previous classification algorithms, exhibiting impressive accuracy and making minimal errors in detecting and classifying the social bots.

6. Conclusion

To sum up, the utilization of online social networks has brought about a noteworthy revolution in people's communication methods and their ability to express personal and public viewpoints. However, the rise of social bots on websites like Twitter has become a major problem since they compromise the veracity of tweets, spread false information, and hurt both people and society as a whole. Therefore, it is crucial to identify and locate social bots on Twitter to maintain the credibility of tweet content and protect democratic principles.

Past research has utilized a range of methods, including profile-based, behavioral-based, temporal-based, graph network-based, and content-based approaches, to detect social bots; however, each of these approaches has its limitations and lacks sufficient effectiveness in accurately identifying social bots. Consequently, this study introduces a novel approach that synergistically integrates all of the mentioned attributes to definitively and accurately identify social bots.

After testing our XGBoost model on hybrid features using the Twibot-22 dataset, we discovered that it outperforms other classifiers in identifying Twitter social bots, attaining an accuracy of 89.8. This research offers an efficient solution for identifying and finding social bots on Twitter, which helps protect the authenticity of tweet content and support democracy. In conclusion, while our proposed approach holds promise for enhancing social bot detection accuracy on Twitter, there are avenues for further refinement. Future research endeavors could involve synergizing our method with additional techniques such as topic modeling, sentiment analysis, and natural language processing. Additionally, it is important to acknowledge the limitations of our current research, which primarily focused on Twitter and may require adaptation for application on other platforms.

Acknowledgment

I would like to express my sincere gratitude to all those who have contributed to the completion of this research paper. I extend my heartfelt appreciation to my supervisor for their guidance, support, and valuable insights throughout the entire process. I would also like to thank the participants who generously provided their time and cooperation for data collection. Additionally, I am grateful to my colleagues and friends for their encouragement and assistance during this endeavor. Finally, I acknowledge the funding agencies that have supported this research, enabling its realization.

References

[1] K. Hayawi, S. Mathew, N. Venugopal, M.M. Masud, and P.H. Ho, "DeeProBot: a hybrid deep neural network model for social bot detection based on user profile data," Social Network Analysis and Mining, vol. 12, pp. 43, 2022.

[2] J. León-Quismondo, "Social Sensing and Individual Brands in Sports: Lessons Learned from English-Language Reactions on Twitter to Pau Gasol's Retirement Announcement," International Journal of Environmental Research and Public Health, pp. 895, 2023.

[3] R. Gilmary, A. Venkatesan, and G. Vaiyapuri, "Detection of automated behavior on Twitter through approximate entropy and sample entropy," Personal and Ubiquitous Computing, pp. 1-15, 2021.

[4] B. Wu, L. Liu, Y. Yang, K. Zheng, and X. Wang, "Using improved conditional generative adversarial networks to detect social bots on Twitter," IEEE Access, pp. 36664-36680, 2020.

[5] M. Heidari, S. Zad, P. Hajibabaee, M. Malekzadeh, S. HekmatiAthar, O. Uzuner, and J.H. Jones, "Bert model for fake news detection based on social bot activities in the COVID-19 pandemic," IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), 2021.

[6] M. Kolomeets, O. Tushkanova, D. Levshun, and A. Chechulin, "Camouflaged bot detection using the friend list," In 2021 29th Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP), 2021.

[7] E. Puertas, L.G. Moreno-Sandoval, F.M. Plaza-del Arco, J.A. Alvarado-Valencia, A. Pomares-Quimbaya and L. Alfonso, "Bots and gender profiling on Twitter using sociolinguistic features," CLEF (Working Notes), pp. 1-8, 2019.

[8] L. Ilias and I. Roussaki, "Detecting malicious activity in Twitter using deep learning techniques," Applied Soft Computing, pp. 107360, 2021.

[9] R. Bailurkar and N. Raul, "Detecting bots to distinguish hate speech on social media," in International Conference on Computing Communication and Networking Technologies, 2021.

[10] L. Alkulaib, L. Zhang, Y. Sun and C.T. Lu, "Twitter Bot Identification: An Anomaly Detection Approach," IEEE International Conference on Big Data (Big Data), 2022.

[11] M. Mendoza, M. Tesconi and S. Cresci, "Bots in social and interaction networks: detection and impact estimation," ACM Transactions on Information Systems (TOIS), vol. 39, pp. 1-32, 2020.

[12] S. Najari, M. Salehi and R. Farahbakhsh, "GANBOT: a GAN-based framework for social bot detection," Social Network Analysis and Mining, vol. 12, pp. 1-11, 2022.

[13] M.M. Pingili, M.J. Lakshmi, M. Divya, D. Abhishek, K. Naresh and P. Sparsha, "Detection of malicious social bots using learning automata with URL features in twitter network," IEEE Transactions on Computational Social Systems, vol. 7(4), pp. 1004-1018, 2020.

[14] Y. Wu, Y. Fang, S. Shang, J. Jin, L. Wei and H. Wang, "A novel framework for detecting social bots with deep neural networks and active learning," Knowledge-Based Systems, vol. 211, pp. 106525, 2021.

[15] M. Mazza, S. Cresci, M. Avvenuti, W. Quattrociocchi and M. Tesconi, "Rtbust: Exploiting temporal patterns for botnet detection on twitter," Proceedings of the 10th ACM conference on web science, 2019.

[16] E. Alothali, K. Hayawi and H. Alashwal, "Hybrid feature selection approach to identify optimal features of profile metadata to detect social bots in Twitter," Social Network Analysis and Mining, vol. 11, pp. 1-15, 2021.

[17] H. Shukla, N. Jagtap and B. Patil, "Enhanced Twitter bot detection using ensemble machine learning," 6th International Conference on Inventive Computation Technologies (ICICT), 2021.

[18] G. Lingam, R.R Rout, D.V. Somayajulu and S.K. Das, "Social botnet community detection: a novel approach based on behavioral similarity in twitter network using deep learning," Proceedings of the 15th ACM Asia Conference on Computer and Communications Security, 2020.

[19] S. Kumar, S. Garg, Y. Vats and A.S. Parihar, "Content based bot detection using bot language model and bert embeddings," 5th International Conference on Computer, Communication and Signal Processing (ICCCSP), 2021.

[20] C. Monica and N. Nagarathna, "Detection of fake tweets using sentiment analysis," SN Computer Science, vol. 1, pp. 1-7, 2020.

[21] D. Kosmajac and V. Keselj, "Twitter bot detection using diversity measures," Proceedings of the 3rd International Conference on Natural Language and Speech Processing, 2019.

[22] N. Vo, K. Lee, C. Cao, T. Tran and H. Choi, "Revealing and detecting malicious retweeter groups," in Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2017.

[23] S. Feng, Z. Tan, H. Wan, N. Wang, Z. Chen, B. Zhang, Q. Zheng, W. Zhang, Z. Lei, S. Yang and X. Feng, "TwiBot-22: Towards graph-based Twitter bot detection," 2022.

[24] M.A. Hall, "Correlation-based feature selection for machine learning," The University of Waikato, 1999.