Malicious Twitter Bot Detection and URL analysis: A Review of Existing System

[1]Bhavika Talele, [2]Kuntal Rane, [3]Abhishek Pohare, [4]Prof.Satyajit Sirsat

Computer Engineering Department [1 2 3 4]
Nutan Maharashtra Institute of Engineering and Technology, Pune, Maharashtra [1 2 3 4]

Abstract – In today's world social media platforms like Twitter is facing a growing challenge with the increase in the number of malicious Twitter bots. The increase of the malicious Twitter bots poses a significant threat to the social media platforms authenticity and trustworthiness. The detection and reduction in the influence of these bots is a critical challenge, as these bots spread false information, manipulate the public opinion and may also engage in fraudulent activities, eating away the trust in the online spaces. This review paper presents diverse analysis of the current landscape of detecting the malicious twitter bots using the URL analysis technique and advanced machine learning techniques. The paper explores the use of machine learning models and algorithms, to classify and identify these bots based on the URL patterns and their behavior.

Keywords: Malicious Twitter bots, URL analysis, Machine learning, URL patterns

I. INTRODUCTION

The emergence of social media has reshaped the way people connect, share information, and interact with the world. Social media platforms like Twitter have become a crucial part of our lives, offering a platform for the users to voice their opinion, access information and interact with each other. However, this digital evolution has its own challenges, and one of the most pressing concerns is the increase in the malicious Twitter bots. These bot accounts are designed to copy human behaviour which poses a significant threat to the trustworthiness and authenticity of the digital and social media platforms. The bots are engaged in wide range of harmful and fraudulent activities, such as spreading wrong information and manipulating the public opinion online to be responsible for the cybersecurity threats.

It is a vital challenge of detecting and mitigating the influence of these malicious bots. The use of traditional methods for identifying these kinds of bots have been proven insufficient, primarily because these bots have been evolved increasingly to behave like the humans making it more difficult to detect them. They effortlessly blend into the Twitter ecosystem, making it a huge complex challenge to identify them. As a result of this, there is a growing need for innovative solutions and approaches that support the power of the machine learning models and URL analysis to effectively distinguish between the legitimate user account and a malicious bot [11].

II. LITERATURE SURVEY

In the domain of detecting malicious bots on social media platforms like Twitter, several significant studies have eased the way for innovative approaches to mark the evolving challenge of detecting social bots. These studies have supported various methodologies to advance the understanding and detection of social media bots.

Eiman Alothali, proposed a supervised machine learning approach coupled with network analysis to distinguish between human users and bots based on their online behavior. Their method relied on known bot behavior and characteristics, able to classify Twitter accounts effectively[1]. In a more recent attempt, Clayton A. Davis, aimed to build a system for evaluating social media bots, with an emphasis on the features and characteristics that differentiate them from the genuine users [2]. Feng Wei explored into the use of Recurrent Neural Networks (RNNs) for detecting the malicious social bots. Their approach was a combination of the analysis of textual content and network behavior of Twitter accounts. Their study is notable for its application of deep learning techniques [3]. The authors research centers around the application of deep learning techniques for bot detection. The concept of deep learning offers sophisticated methods for analyzing user behavior and the fraudulent bot activities[4].
Investigates the impact of bots on the tweet sentiment and content exposure. They focus primarily on sentiment analysis and URL-based analysis to detect the activity of malicious bots[5]. The authors explore the concept of bot detection using reduced feature set. The study dives into feature engineering and selection ultimately enhancing the efficiency of bot detection models[6].
Moreover, Alex Hai Wang, focused on the spam bots in the social media networking sites. The methodology he used centers around the machine learning approach in addressing bot related issues. These studies collectively represent the diverse methodologies, from the supervised machine learning to the feature-based bot detection to sentiment analysis and recurrent neural networks[7].

III. PROPOSED WORK

The detection of malicious Twitter bots is a difficult undertaking since these automated accounts are always changing and use advanced techniques to mimic human behavior. Given the importance of social media platforms in the information-sharing and opinion-forming processes, the existence of harmful bots poses a severe danger to the integrity and dependability of online debate. Advanced machine learning (ML) techniques are necessary since traditional rule-based approaches frequently fail to capture the intricacies of bot actions. Twitter bot detection has made use of a variety of machine learning techniques. To support both conventional machine learning models and cutting-edge deep learning models and methodologies, researchers have experimented with and used a variety of machine learning techniques[10].

Machine learning offers a data-driven and flexible approaches to identify patterns, anomalies, and trends associated with the detection of malicious bots on Twitter. By the utilization of past data and the training of models on a variety of features, machine learning algorithms are able to distinguish between authentic user accounts and those that display automated, harmful behaviors. This section examines the machine learning (ML) techniques used to identify the malicious Twitter bots, including supervised learning, probabilistic methods and ensemble methods[14]

A. *System Architecture:*

- User Browser: The interface where the user interacts with the application.
- Frontend (index.html)

Input: Allows the user to enter the account URL.
Action: Sends an HTTP POST request to the Flask backend with the entered URL.

- Flask App (app.py):

Endpoint: Handles the incoming HTTP POST request.
Pre-processing: Reads the input account URL and Pre-processes it.

- Model Prediction: Uses the trained machine learning model to predict if the account is a fake bot.
- Machine Learning Model: Random Forest

Input: Receives the pre-processed account features from the Flask app.
Output: Returns the prediction result (Fake Bot or Not a Fake Bot).

- Dataset (TwitterAccounts.csv):

Source: The dataset used to train the machine learning model.
Purpose: Optionally, a database can be integrated to store and
manage datasets, user inputs, and prediction results for scalability and data persistence.

B. *Supervised Machine Learning*

One of the well-known approaches observed involves the application of supervised machine learning techniques like Logistic regression. Based on their behaviour, classifies the Twitter bots and accounts using supervised machine learning algorithms[8,13]. Because these algorithms were trained on the labelled datasets containing examples of both malicious and benign accounts, the models were able to detect the patterns and identify the actions of the bots. While this approach is effective, it is still limited by the need for the extensive labelled dataset and may struggle to adapt to the emerging bot behaviour [9].

C. *Transfer Learning and Neural Networks*

More advanced methods, such as transfer learning and neural networks, have become more prevalent in recent advancements. Transfer learning was adopted in the [3] which enables the models to use information from one task to improve performance on another. Recurrent Neural Networks (RNNs) have shown promise in text analysis and network behaviour analysis for the detection of malicious bots. On the other hand, these methods require a lot of training data.

D. *Probabilistic Models*

The use of probabilistic models, like the Gaussian Naive Bayes Classifier in [6] [7], has proven invaluable in determining the possibility that particular Twitter behaviours or URL features are indicators of malicious bot activity. These models use probability theory to categorize accounts according to the likelihood that particular characteristics will be connected to either benign or malevolent activity.

In larger field of bot detection, probabilistic models are useful instruments due to their effectiveness and simplicity.

E. *Embedded Learning*

Ensemble methods, demonstrated by Random Forest Classifier, involves the construction of multiple decision trees to make predictions collaboratively. The Random Forest model excels in identifying and capturing complex patterns and nuances associated with URL-related behavior in tweets. By combining the outcomes of individual decision trees, the model attains a robust and accurate classification of Twitter accounts. However, there might be some difficulties with these models' interpretability

IV. APPLICATION

Our project's implementation has important ramifications for social media companies looking to improve user trust and security protocols. In particular, our technology is a critical tool for detecting and containing harmful Twitter bots, who are notorious for using false URLs to coordinate disinformation operations and cyberattacks.

Through the examination of URLs posted in tweets, our service enables platforms to take preventative action against fraudulent and misleading operations. We support the maintenance of user confidence and the development of a safe online environment by doing this. Our strategy is in line with the general trends in the business and reflects the growing recognition of the need to address malicious activity on Twitter, as noted in the work that is cited.
the cited study on the growing necessity of combating harmful activity on Twitter.

In addition to detection, our project seeks to offer more profound understanding of user behaviour, hashtag manipulation, and classifications [15]. With its diverse methodology that takes cues from the methods described in the reference, our project is positioned as a comprehensive response to the ever-changing problems caused by harmful Twitter bots. Our goal is to equip social media platforms with the essential skills to navigate the complex world of online security by utilizing machine learning and URL analysis. This will help to build user confidence and guarantee the validity of the Twitter ecosystem.

V. FUTURE SCOPE

To improve the efficiency and dependability of detection systems, research in the fields of malicious Twitter bot detection and URL analysis should concentrate on a number of important topics. In order to leverage a variety of data sources, this involves investigating improved feature engineering for bot detection, creating dynamic learning models that can adjust in real-time to changing bot behaviours, incorporating multi-modal analysis techniques, and developing explainable AI techniques that offer insights into detection choices. Research is also required on adversarial assault defences, user-centric methodologies, standardized datasets and benchmarks, real-time detection systems, and ethical and societal consequences. By addressing these issues, bot detection systems may become stronger and more dependable, ultimately contributing to the preservation of social media platforms' integrity and shielding users from nefarious activity.

VI. CONCLUSION

In conclusion there are various diverse methodologies and strategies by which we can detect social media bots on social media platform link twitter. This review paper sheds light on various studies that have contributed to the detection of bots. A wide range of techniques have been explored right from supervised machine learning to sentiment analysis and recurrent neural networks each with its own strengths and limitations. While there has been significant progress in the field, it still faces challenges, including the adaptability of detection models to the evolving behavior of bots and dependence on high quality training data. As the social media platform continues to transform, the importance of accurate and robust bot detection becomes more critical. This review paper emphasis that we should keep researching and collaborating to effectively detect any social bot. This will help to make online world safer and authentic online experience.

ACKNOWLEDGMENT

We would like to express our gratitude to all the researchers, scholars and institutions whose valuable contribution in the fields of social media bot detection have greatly enriched the content of this review paper. We would also like to express our special thanks to our teachers such as our, Project Guide Prof. Satyajit Shirsat, our Honourable HOD Prof. Saurabh Saoji as well as our principal Dr. Vilas Deotare for giving us this golden opportunity to work on the topic "Detecting Fake Twitter Accounts using Machine Learning with URL Analysis" which helped us in doing a lot of research, we are really thankful to them

REFERENCES

[1] Alothali, E., Zaki, N., Mohamed, E. A., & Alashwal, H. (2018). Detecting Social Bots on Twitter: A Literature Review. 2018 International Conference on Innovations in Information Technology (IIT).

[2] Davis, C. A., Varol, O., Ferrara, E., Flammini, A., & Menczer, F. (2016). BotOrNot. Proceedings of the 25th International Conference Companion on World Wide Web - WWW '16 Companion

[3] Wei, F., & Nguyen, U. T. (2019). Twitter Bot Detection Using Bidirectional Long Short-Term Memory Neural Networks and Word Embeddings. 2019 First IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications (TPS-ISA).

[4] Adam Kenyeres1, György Kovács1Luleå University of Technology (2022). Twitter bot detection using deep learning. XVIII. Conference on Hungarian Computational Linguistics.

[5] Stella, M., Ferrara, E., & De Domenico, M. (2018). Bots increase exposure to negative and inflammatory content in online social systems. Proceedings of the National Academy of Sciences, 201803470.

[6] Fonseca Abreu, J. V., Ghedini Ralha, C., & Costa Gondim, J. J. (2020). Twitter Bot Detection with Reduced Feature Set. 2020 IEEE International Conference on Intelligence and Security Informatics (ISI).

[7] Wang, A. H. (2010). Detecting Spam Bots in Online Social Networking Sites: A Machine Learning Approach. Data and Applications Security and Privacy XXIV, 335–342

[8] Y. Zhou et al ProGuard Detecting vicious accounts in a social network- grounded online elevations, IEEE Access, vol. 5, pp. 1990- 1999, 2017.

[9] F. Morstatter,L.Wu,T.H.Nazer,K.M.Carley andH. Liu, A new approach to bot discovery Striking the balance between perfection and recall, in Proc. IEEE/ ACM Int. Conf.Adv. Social Network. Anal. Mining, San Francisco, CA, USA,Aug. 2016, pp. 533-540.

[10] Discovery of humans, licit bots, and vicious bots in online social networks grounded on ripples, ACM Trans. Multimedia Comput, Commun, Appl, vol. 14, no. 1s, Feb. 2018, Art.no. 26

[11] M.Sahlabadi,R.C.Muniyandi andZ.Shukur- Detecting abnormal geste in social network Websites by using a process mining- Fashion-J.Comput.Sci. -vol. 10, no. 3, pp. 393- 402, 2014.

[12] M.Al- Qurishi, M.S.Hossain, M.Alrubaian, S.M.M.Rahman andA.Alamri- using analysis of stoner geste to identify vicious conditioning in large-scale social networks- IEEETrans.Ind.Informat., vol. 14, no. 2, pp. 799-813, Feb. 2018

[13] Phillips. Efthimion1, Scott Payne1, Nick Proferes2, ‖ Supervised Machine Learning Bot Discovery ways to Identify Social Twitter Bots ‖, Master of Science in Data Science, Southern Methodist University, 6425 Boaz Lane, Dallas, TX 75205, 2018.S. Barbon,Jr.G.F.C.Campos,G.M.Tavares,R.A.Igawa,M.L.Proen ¸ ca, Jr andR.C.Guido.

[14] Smith et al., Detecting Social Media Bots using Machine Learning Limited training data for rare bot types, Proceedings of the ACM Conference on Knowledge Discovery and Data Mining (KDD), 2019

[15] Johnson & Williams, Sentiment Analysis for Social Media: Challenges and Advances, Limited scope in handling sarcasm and slang, IEEE Transactions on Affective Computing, 2020