ICEST-2K24 [International Conference on Engineering, Science and Technology]
In Association with
International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN - 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)

Tool for Identifying and Categorizing the Twitter Bots Using Machine Learning

Prof. Jalindar Nivrutti Ekatpure[1], Mr. HarshadPrakashBarve[2], Mr. Aditya Umesh Bhange[2], Mr. Sanket Sunil Patil[2], Mr. Aadit Suresh Yadav[2]

[*1]Assistant Professor, S. B. Patil College of Engineering, Indapur, Maharashtra, India
[2]S. B. Patil College of Engineering, Indapur, Maharashtra, India

ABSTRACT

Developing a Twitter tool to distinguish between useful and harmful bots aims to enhance user safety by categorizing bots and enabling users to utilize beneficial bots while safeguarding against potential privacy threats This paper aims to ensure Twitter bot users' data privacy. There are very few existing systems with very low accuracy. The tool aims to enhance user safety by categorizing Twitter bots into two distinct groups using machine learning: Useful Bots and Harmful Bots. The classification system serves as a safeguard, allowing users to interact with beneficial bots for their purposes while avoiding harm. The paper aims to create a safer and more efficient Twitter experience by allowing users to navigate and interact with bots responsibly.

Keywords: Summarization, Classification, Identification, Machine Learning, analysis, Data Analysis

I. INTRODUCTION

In the vast landscape of social media, Twitter stands as a dynamic platform for global conversations, news publishing, etc. However, some software, commonly known as bots, can interact with systems or users. The nature of these bots, coupled with their potential to influence narratives and manipulate tweets, makes it necessary to develop advanced tools for their identification and categorization. This research introduces a robust machine-learning framework designed to reveal the difficulty of Twitter's bot ecosystem. With a focus on both identification and categorization, our tool provides platform administrators, cybersecurity experts, and researchers with the means to discern between genuine users and automated bots, while also revealing the purposes behind the activities of these bots. This research helps us in the following task:

- Identification: Develop an advanced machine learning model capable of identifying Twitter bots by analysing behavioural patterns, engagement count and other key features.
- Our approach combines machine learning algorithms, natural language processing techniques, and data analytics to sift through the massive volume of Twitter data. Leveraging labelled datasets and a meticulous training regimen, the model endeavours to discern subtle distinctions between human and bot behaviour.

- Extract relevant features from user profiles, tweets and engagement data to improve the accuracy of bot detection. Using machine learning to detect unusual behaviour patterns that are characteristic of bot accounts, such as high-frequency posting or identical content sharing are several existing systems which also help the users to do the same work. However, those systems are unable to generate the desired output. Because those systems had a lack of accuracy.

II. LITERATURE SURVEY

As per the paper[1] author stated that The consideration of images in detecting Twitter bots because images can contain valuable information and analysing them could potentially improve the accuracy of bot detection models.

As per the paper[2] author stated that moving beyond traditional feature-based detection, future systems may focus on analysing the behavioural patterns of accounts to detect bots more effectively. This could include looking at posting frequency, content similarity, and engagement patterns.

As per the paper[3] author stated that we need to develop a few more models and use some other features that help to find the bot in a more précised and accurate way.

As per the paper[4] author stated that Future implementations could provide real-time data, which would allow Twitter to incorporate this function into their app. Additionally, it can be integrated among all other market-available social networking programs. in this paper, the dataset used for detection is provided through us, it is entirely manual. However, in future, I may upgrade the paper so that the model can use the dataset needed for bot detection on its own.

As per the paper[5] author stated that Graph-based bot detection methods demand significantly more computation resources and execution time than feature-based models. Given that the Twitter network is rapidly expanding, we aim to further explore scalable and graph-based bot detection methods.

As per the paper[6] author stated that Retrainable models through real-time processing would be another solution to this issue. Finally, most of the models are confined to Twitter now. Cyberattacks surge. Cybercriminals seek efficient channels to spread malware via images. JPEGVigilant, a machine learning method, identifies malicious JPEGs using 10 derived properties.[16] Leveraging the DL solutions to overcome similar issues in other platforms may potentially increase the usability and impact of this research to a great extent.

As per the paper[7] author stated that we will use those features as heuristics of an unsupervised system aimed at ranking Twitter accounts from more human to less human. Paper innovates plant species classification using Deep Learning and leaf vein features, aiming to automate identification, accelerate research, aid conservation, and foster education in botany and technology.[15] This ranked list of accounts will be revised by annotators so that a reliable gold-standard dataset is obtained at the end.

III.PROPOSED SYSTEM

A. Problem Statement

A tool that will identify harmful Twitterbots to ensure the user's privacy

B. Architecture Diagram

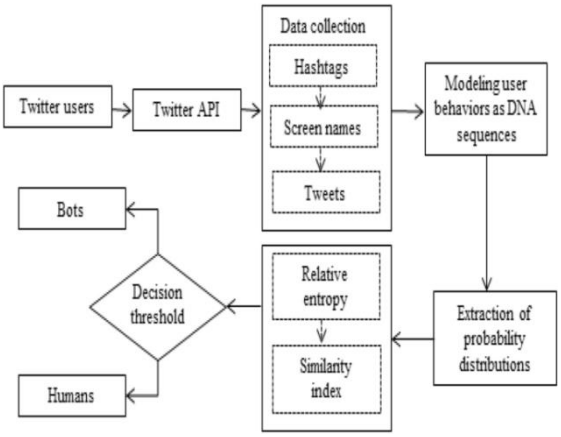

Figure 1: Architecture Diagram

C. Requirements

1) Hardware Requirements:

- Processor - Intel i3/i5/i7
- Speed - 1.1 GHz
- RAM - 4Gb(Min)
- Hard Disk - 256GB
- Keyboar- Standard Keyboard
- Mouse -Two or Three-Button Mouse

2) Software Requirements:

- Operating System- Windows 7/8/10
- Application Server - Apache Tomcat 7
- Front End - HTML, CSS, Bootstrap
- Language -Python
- IDE - Visual Studio Code

IV. RESULT AND DISCUSSION

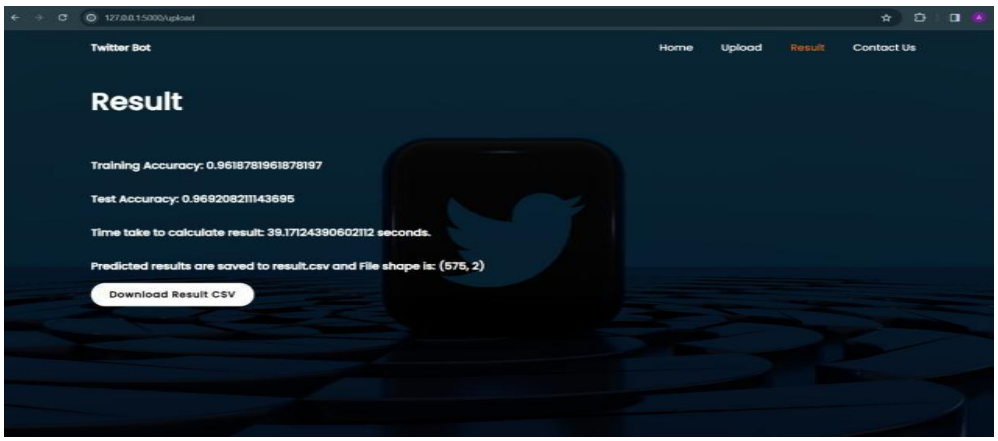

Figure 2:

Our machine-learning model can accurately classify Twitter bots when we test it on a second dataset based on follower counts without the need for manual feature engineering. Screenshots of the classification results, including test accuracy, are available. However, training is currently delayed due to the unavailability of training datasets.

V. CONCLUSION

The development and implementation of bot detection & classification systems represent a significant step forward in the safe use of Twitter, offering promising solutions to address the challenges posed by Twitter bots. Bots adapt, evolve, and find new ways to mimic authentic user behaviour. As we navigate the complexities of identifying and categorizing Twitter bots, we acknowledge the dynamic nature of these digital bots. In conclusion, As we look to the future, the path forward involves not only refining our technological tools but also a collaborative and ethically grounded approach to face the challenges that lie ahead. We're always looking to make it better based on what you and other users tell us. We also take your privacy and security very seriously. You can access our tool on the web, designed to be user-friendly.

VI. REFERENCES

[1]. F.K. Alarfaj, H. Ahmad, H.U. Khan, A.M. Alomair, N. Almusallam, M. Ahmed " Twitter Bot Detection Using Diverse Content Features and Applying Machine Learning Algorithms" 2023.

[2]. Hrushikesh Shukla, BalajiPatil, NakshatrJagtap "Enhanced Twitter bot detection using ensemble machine learning"2021.

[3]. P. SaiKarthik Reddy, P. SaiNath, Dr. J. Vijayashree "Twitter Bot Detection Using Machine Learning Algorithms" 2023.

[4]. Sopinti Chaitanya Raj, B. Srinivas, S.P. Kumar "Detecting Malicious Twitter Bots Using Machine Learning"2022.

[5]. ShangbinFeng, Zhaoxuan Tan, Herun Wan1, Ningnan Wang, Zilong Chen, Binchi Zhang "TwiBot-22: Towards Graph-Based Twitter Bot Detection" 2022.

[6]. Kadhim Hayawi, Susmita Saha, Mohammad Mehedy Masud, Sujith Samuel Mathew, Mohammed Kaosar "Social media bot detection with deep learning methods" 2023.

[7]. Pablo Gamallo and SattamAlmatarneh "Naive-Bayesian Classification for Bot Detection in Twitter" 2019.

[8]. Rajnish K. Prince, Snehal S. Thube, Rahul Ranjan, AkashL.Sakat, V. A. Yaduvanshi. "Detection Of Bots In Twitter Network Using Machine Learning Algorithm" 2022.

[9]. AshkanDehghan, KingaSiuta, AgataSkorupka, AkshatDubey, Andrei Betlen, David Miller, Wei Xu, PawełPrałat "Detecting bots in social-networks using node and structural embeddings" 2022.

[10]. Yicong Chen and Jiahe Ling "Online Twitter Bot Detection: A Comparison Study of Vectorization and Classification Methods on Balanced and Imbalanced Data" 2023.

[11]. Aaglave, K. N., Shivanjali Santosh Jadhav, Amaan Firoj Khatib, and Rohini Laxman Khurangale. "A Survey on the Web Scraping: In the Search of Data." (2023).

[12]. Karve, S. M, Kakad , S, Swapnaja Amol, Gavali, A. B. ., Gavali , S. B. ., & Shirkande, S. T. . (2024). An Identification and Analysis of Harmful URLs through the Application of Machine Learning Techniques.

International Journal of Intelligent Systems and Applications in Engineering, 12(17s), 456–468. Retrieved: https://www.ijisae.org/index.php/IJISAE/article/view/4905.

[13]. Kale, R, Shirkande, S. T., Pawar, R., Chitre, A, Deokate, S. T, Rajput, S. D, & Kumar, J. R. R(2023). CR System with Efficient Spectrum Sensing and Optimized Handoff Latency to Get Best Quality of Service. International Journal of Intelligent Systems and Applications in Engineering, 11(10s), 829–839.

[14]. Ajinath, B. S., Sunil, H. S., Digambar, K. S., Anandkumar, B. P., Nalawade, V. S., &Sayyad, G. G. (2018). Optimizing Information Leakage and Improve Security over Public Multi-Cloud Environment. Journal of emerging technologies and innovative research.

[15]. Ekatpure, J. N., Kamble, Y. P., More, P. T., & Patankar, S. S. (2023). A Survey On Leaf Vein Morphometrics: A Deep Learning Approach to Plant Classification.

[16]. Ekatpure, J. N., Kharade, N., Korake, D., Kshirsagar, D., & Mind, R. (2023). JPEG Vigilant: AI-Powered Malware Image Detection.