ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 18, Issue 2, 2024

TWEET BASED BOT DETECTION USING BIG DATA

P. PAVAN KUMAR[1], M. VIKAS KUMAR[2], P. BHUVANESWARI[3], SABHAVATH RAMADEVI[4]

ASSISTANT PROFESSOR[1], UG SCHOLAR[2,3&4]

DEPARTMENT OF CSE, CMR INSTITUTE OF TECHNOLOGY, KANDLAKOYA VILLAGE, MEDCHAL RD, HYDERABAD, TELANGANA 501401

ABSTRACT— With millions of users, Twitter is one of the most widely used microblogging social media sites. Because of its popularity, Twitter has been the subject of various attacks, including spyware, phishing links, and the dissemination of falsehoods. Because they can initiate extensive attacks and manipulation efforts, tweet-based botnets pose a significant risk to users. Big data analytics approaches, including shallow and deep learning techniques, have been used to address these risks by accurately differentiating between tweet-based bot accounts and human accounts. In this research, we present a taxonomy that categorizes the state-of-the-art in tweet-based bot identification methods and explore current approaches. Along with their performance outcomes, we also outline the shallow and deep learning methods for tweet-based bot detection. The dynamic nature of bot activity on Twitter must be addressed in addition to the several detection methods. It gets harder to tell bots apart from real people as they get more complex since they can imitate human behavior more accurately. To improve detection accuracy, researchers are investigating the application of cutting-edge techniques like ensemble learning, natural language processing (NLP), and graph-based algorithms. Furthermore, the sustainability of existing bot detection techniques is seriously threatened by the ongoing evolution of adversarial tactics meant to evade detection systems. Future research will probably concentrate on developing more flexible systems that can react in real time to novel forms of bot behavior and learn from new patterns.

Index Terms— Social media, Twitter, big data analytics, shallow learning, deep learning, tweet-based bot detection.

I. INTRODUCTION

Nowadays, social media is one of the most popular tools used by people to communicate with one another. It is also largely used by organizations to reach out to customers. In [1], it has been reported that there are 3.5 billion active social media users globally. Facebook, Twitter, LinkedIn, and other social media networks are used by organizations to improve brand visibility and boost their sales. Twitter is one of the most popular social media platforms. It has 340 million active users who are allowed to communicate at a large scale and share their opinions about different topics. Twitter could be targeted by various kinds of attacks. For example, a spear phishing attack in July 2020 led to the hijack of high-profile Twitter accounts [2]. Also, fraudulent accounts could be created to impersonate legitimate users and organizations. Twitter can also be exploited by botnet, which is a set of malicious accounts that operate under a botmaster, and are controlled by software programs rather than human users. The tweet-based social media bots pose serious security risks to Twitter users. These bots are used to spread fake contents, phishing links, and spams. Although they are not used as bots to launch DDoS attacks, they could be utilized as Command and Control (C&C) infrastructure to coordinate DDoS attacks [3], [4]. They are capable of interacting with human accounts to deceive the users and hijack their accounts. These bots are also used as tools to launch large-scale manipulation campaigns to influence public opinions. According to a study [5], 52% of

online traffic is generated by botnets, and the rest is produced by actual users. It is also worthy to note that some bots are found with over 350,000 fake followers. To deal with the above issues, there is a need to develop detection systems that can accurately distinguish between Twitter bot accounts and human accounts. Twitter data represent one of the examples of big data as around 500 million tweets are generated every day, i.e., 6,000 tweets every second [6]. Big data analytics has been widely used in different fields [7]–[11] to process large amount of data, discover hidden patterns, and find correlations among data points. Artificial intelligence techniques are increasingly leveraged by big data analysis. In particular, shallow (conventional) and deep learning techniques have received considerable attention from the academia and industry due to their success in dealing with heterogeneous and complex data, automatic learning of models, revealing unseen patterns, identifying dependencies, and getting insights from analyzing data. Artificial intelligence has been extensively used by Twitter to determine tweet recommendations for users. In fact, deep neural networks are applied on Twitter data to determine the relevant content for users, and hence improve their experience on the platform [12]. Artificial intelligence has played an important role in fighting inappropriate content. In 2017, about 300,000 accounts were suspended and identified with the help of artificial intelligence tools rather than humans. This review aims at providing an overview of different tweet-based bot detection methods that use shallow and deep learning techniques to distinguish between human accounts and bot accounts.

II. LITERATURE SURVEY

A) Social media bot detection with deep learning methods: a systematic review by Kadhim Hayawi, Susmita Saha, M. Masud, S. Mathew, M. Kaosar, Published in Neural computing & applications (6 March 2023)- Social bots are automated social media accounts governed by software and controlled by humans at the backend. Some bots have good purposes, such as automatically posting information about news and even to provide help during emergencies. Nevertheless, bots have also been used for malicious purposes, such as for posting fake news or rumour spreading or manipulating political campaigns. There are existing mechanisms that allow for detection and removal of malicious bots automatically. However, the bot landscape changes as the bot creators use more sophisticated methods to avoid being detected. Therefore, new mechanisms for discerning between legitimate and bot accounts are much needed. Over the past few years, a few review studies contributed to the social media bot detection research by presenting a comprehensive survey on various detection methods including cutting-edge solutions like machine learning (ML)/deep learning (DL) techniques. This paper, to the best of our knowledge, is the first one to only highlight the DL techniques and compare the motivation/effectiveness of these techniques among themselves and over other methods, especially the traditional ML ones. We present here a refined taxonomy of the features used in DL studies and details about the associated pre-processing strategies required to make suitable training data for a DL model. We summarize the gaps addressed by the review papers that mentioned about DL/ML studies to provide future directions in this field. Overall, DL techniques turn out to be computation and time efficient techniques for social bot detection with better or compatible performance as traditional ML techniques.

B) Fake Profile Detection Using Deep Learning Ms.B. Gunasundari, M. Baráth, M. Hariharan, Published in 2023 - Online social networks (OSN) have greatly improved communication, information exchange, and enjoyment in modern society. However, because of their accessibility and anonymity, OSNs have provided a favorable

environment for a variety of harmful practices like spamming, trolling, fake news, and astroturfing. One of the primary threats to OSNs is socialbots, which are computer programs that perform various illicit activities. To address the threat of socialbots, researchers have been developing various detection methods. One of the latest and most advanced methods is SBRidAPI, which stands for SocialBot RID (Rapid Identification) using deep API learning. The goal of SBRidAPI is to detect socialbots by analyzing a user's behavior on OSNs.SBRidAPI models a broad range of profile, temporal, activity, and content information for user behavior representation using deep learning techniques. Profile information includes user's name, age, location, and other similar data, whereas temporal information is about the frequency and timing of user activities. Activity information includes the type of activities, such as likes, comments, and shares. Content information includes the text, images, and videos shared by the user. SBRidAPI represents profile, temporal, and activity information as sequences in order to analyze the sequential nature of this information that is supplied to a two-layers stacked BiLSTM. Deep CNN is fed content data in order to analyze the text content and learn the visual characteristics of images and videos. Once SBRidAPI has analyzed a user's behavior, it assigns a score that reflects the likelihood of the user being a socialbot. The user is labeled as a socialbot if their score is higher than a threshold that is then used to compare scores. SBRidAPI is the first method that jointly models a complete collection of profile, temporal, activity, and content information for user behavior representation, making it an effective tool for identifying socialbots on OSNs.

C) A Deep Learning Approach for Robust Detection of Bots in Twitter Using Transformers by David Martín-Gutiérrez, Gustavo Hernández-Peñaloza, Alberto Belmonte Hernández, Alicia Lozano-Diez, Federico Álvarez, Published in IEEE (2021) – During the last decades, the volume of multimedia content posted in social networks has grown exponentially and such information is immediately propagated and consumed by a significant number of users. In this scenario, the disruption of fake news providers and bot accounts for spreading propaganda information as well as sensitive content throughout the network has fostered applied research to automatically measure the reliability of social networks accounts via Artificial Intelligence (AI). In this paper, we present a multilingual approach for addressing the bot identification task in Twitter via Deep learning (DL) approaches to support end-users when checking the credibility of a certain Twitter account. To do so, several experiments were conducted using state-of-the-art Multilingual Language Models to generate an encoding of the text-based features of the user account that are later on concatenated with the rest of the metadata to build a potential input vector on top of a Dense Network denoted as Bot-DenseNet. Consequently, this paper assesses the language constraint from previous studies where the encoding of the user account only considered either the metadata information or the metadata information together with some basic semantic text features. Moreover, the Bot-DenseNet produces a low-dimensional representation of the user account which can be used for any application within the Information Retrieval (IR) framework.

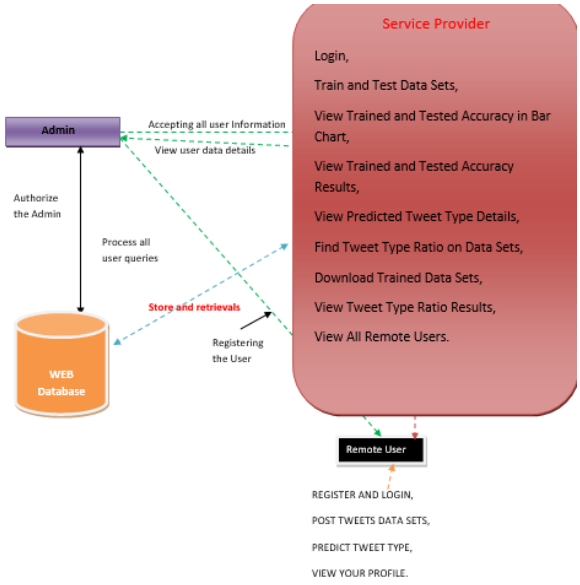
III. PROPOSED SYSTEM

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 18, Issue 2, 2024



Modules

Service Provider

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Login, Browse Data Sets and Train & Test, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results, View All Antifraud Model for Internet Loan Prediction, Find Internet Loan Prediction Type Ratio, View Primary Stage Diabetic Prediction Ratio Results, Download Predicted Data Sets, View All Remote Users.

View and Authorize Users

In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

Remote User

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like REGISTER AND LOGIN, PREDICT PRIMARY STAGE DIABETIC STATUS, VIEW YOUR PROFILE.

CONCLUSION

One of the most widely used social media sites for connecting people and assisting businesses in communicating with their clientele is Twitter. A botnet based on tweets has the ability to infiltrate Twitter and establish malicious accounts in order to initiate extensive attacks and manipulation efforts. In order to combat tweet-based botnets and effectively differentiate between human and tweet-based bot accounts, we have concentrated on big data analytics in this research, particularly shallow and deep learning.
In addition to discussing related polls, we have offered a taxonomy that categorizes the most advanced tweet-

based bot detection methods as of 2020. Furthermore, the performance outcomes of the shallow and deep learning methods for tweet-based bot detection are explained. Lastly, we gave a presentation and talked about the unresolved problems and upcoming research difficulties.

REFERENCES

[1] M. Mohsin. (2020). 10 Social Media Statistics You Need to Know in 2021. [Online]. Available: https://www.oberlo.com/blog/social-mediamarketing-statistics

[2] I. Arghire. (2020). Twitter Hack: 24 Hours From Phishing Employees to Hijacking Accounts. https://www.securityweek.com/twitter-hack24-hours-phishing-employees-hijacking-accounts

[3] The Rise of Social Media Botnets. Accessed: Feb. 21, 2021. [Online]. Available: https://www.darkreading.com/attacks-breaches/the-rise-ofsocial-media-botnets/a/d-id/1321177

[4] M. Imran, M. H. Durad, F. A. Khan, and A. Derhab, "Toward an optimal solution against denial of service attacks in software defined networks," Future Gener. Comput. Syst., vol. 92, pp. 444–453, Mar. 2019.

[5] M. S. Savell. (2018). Protect Your Company's Reputation From Threats by Social Bots. [Online]. Available: https://zignallabs.com/blog/protect-yourcompanys-reputation-from-threats-by-social-bots/

[6] S. Aslam. (2021). Twitter by the Numbers: Stats, Demographics & Fun Facts. [Online]. Available: https://www.omnicoreagency.com/twitterstatistics/

[7] A. Aldweesh, A. Derhab, and A. Z. Emam, "Deep learning approaches for anomaly-based intrusion detection systems: A survey, taxonomy, and open issues," Knowl.-Based Syst., vol. 189, Feb. 2020, Art. no. 105124.

[8] S. Mahdavifar and A. A. Ghorbani, "Application of deep learning to cybersecurity: A survey," Neurocomputing, vol. 347, pp. 149–176, Jun. 2019.

[9] E. B. Karbab, M. Debbabi, A. Derhab, and D. Mouheb, "MalDozer: Automatic framework for Android malware detection using deep learning," Digit. Invest., vol. 24, pp. S48–S59, Mar. 2018.

[10] F. A. Khan, A. Gumaei, A. Derhab, and A. Hussain, "A novel twostage deep learning model for efficient network intrusion detection," IEEE Access, vol. 7, pp. 30373–30385, 2019.

[11] A. Derhab, A. Aldweesh, A. Z. Emam, and F. A. Khan, "Intrusion detection system for Internet of Things based on temporal convolution neural network and efficient feature engineering," Wireless Commun. Mobile Comput., vol. 2020, pp. 1–16, Dec. 2020.

[12] B. Marr. (2020). How Twitter Uses Big Data and Artificial Intelligence (AI). [Online]. Available: https://www.bernardmarr.com/ default.asp?contentID=1373

[13] A. T. Kabakus and R. Kara, "A survey of spam detection methods on Twitter," Int. J. Adv. Comput. Sci. Appl., vol. 8, no. 3, pp. 29–38, 2017.

[14] M. Chakraborty, S. Pal, R. Pramanik, and C. R. Chowdary, "Recent developments in social spam detection and combating techniques: A survey," Inf. Process. Manage., vol. 52, no. 6, pp. 1053–1073, Nov. 2016.

[15] E. Alothali, N. Zaki, E. A. Mohamed, and H. Alashwal, "Detecting social bots on Twitter: A literature review," in Proc. Int. Conf. Innov. Inf. Technol. (IIT), Nov. 2018, pp. 175–180.

[16] M. Latah, "Detection of malicious social bots: A survey and a refined taxonomy," Expert Syst. Appl., vol. 151, Aug. 2020, Art. no. 113383.

[17] K. Yang, O. Varol, C. A. Davis, E. Ferrara, A. Flammini, and F. Menczer, "Arming the public with artificial intelligence to counter social bots," Hum. Behav. Emerg. Technol., vol. 1, no. 1, pp. 48–61, Jan. 2019.

[18] Z. Guo, J.-H. Cho, I.-R. Chen, S. Sengupta, M. Hong, and T. Mitra, "Online social deception and its countermeasures: A survey," IEEE Access, vol. 9, pp. 1770–1806, 2021.

[19] S. B. Abkenar, M. H. Kashani, M. Akbari, and E. Mahdipour, "Twitter spam detection: A systematic review," 2020, arXiv:2011.14754. [Online]. Available: http://arxiv.org/abs/2011.14754

[20] W. Daffa, O. Bamasag, and A. AlMansour, "A survey on spam URLs detection in Twitter," in Proc. 1st Int. Conf. Comput. Appl. Inf. Secur. (ICCAIS), Apr. 2018, pp. 1–6.

[21] C. Besel, J. Echeverria, and S. Zhou, "Full cycle analysis of a large-scale botnet attack on Twitter," in Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM), Aug. 2018, pp. 170–177.

[22] S. C. Woolley and P. N. Howard, Computational Propaganda: Political Parties, Politicians, and Political Manipulation on Social Media. Oxford, U.K.: Oxford Univ. Press, 2018.

[23] Data Dictionary: Standard V1.1. Accessed: Feb. 21, 2021. [Online]. Available: https://developer.twitter.com/en/docs/twitter-api/v1/datadictionary/object-model/tweet

[24] S. Kudugunta and E. Ferrara, "Deep neural networks for bot detection," Inf. Sci., vol. 467, pp. 312–322, Oct. 2018.