International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 4 April 2023, pp: 1217-1219 www.ijaem.net ISSN: 2395-5252

Use of VGG19 to Identify Abusive Twitter Bots

P.Subba Rao[1a], M.shabarish[2b], S.V.Kishore[3c], S.M.M.Mansoor[4d], T.Varun kurmar[5e], B. Chandra Sekhar[6f]

[1]*Assistant Professor, Department of Computer Science and Engineering, Santhiram Engineering College, Nandyal*

[2,3,4,5,6] *Student, Department of Computer Science and Engineering, Santhiram Engineering College, Nandyal*

Date of Submission: 15-04-2023 Date of Acceptance: 25-04-2023

ABSTRACT

Nowadays, Twitter is used often and has significant meaning in the lives of many people, including businesspeople, media professionals, politicians, and others. Twitter, a top social media platform, allows users to express their thoughts and ideas on a wide variety of topics, including politics, sports, the economy, pop culture, and more. It's one of the quickest ways to share data with others. The way people think is profoundly impacted by it. Twitter has seen an increase in the number of users who hide their identities for nefarious purposes. Twitter bots should be easily identifiable due to the potential threat they represent to other users. That's why it's so important for actual individuals, not bots, to publish tweets. The Twitter feed is being spammed by a bot. This means that recognizing bots may help you spot spam emails. Attributes of users' Twitter accounts are utilized as Features in machine learning algorithms to determine whether they are authentic or bot accounts. As part of this research, we used Logistic Regression and VGG19, two machine learning techniques, to establish trustworthiness of social media profiles. We evaluate the accuracy and classification abilities of several methods. Twitter achieves a 73% accuracy with Logistic Regression, a 72% accuracy with URL prediction, and a 90% accuracy with VGG19. Hence, it is clear that VGG19 provides more accuracy than Logistic Regression.

Keywords: Malicious bots, Deep Learning, Logistic Regression, VGG19

I. INTRODUCTION

Twitter is rapidly becoming one of the most popular social networking sites. Sharing information, expressing opinions, and discussing timely topics are all made possible. Users may "follow" other people who they think are interesting or who have similar views. Users may instantly broadcast tweets to their followers. Through re-tweeting, the original tweet is exposed to a new audience. There is a dramatic increase in the volume of tweets during live events like sports games and award shows. Twitter may be accessed from both smartphones and computers.

In addition to potentially increasing product sales, paid marketing may provide new streams of money. Twitter may be a great resource for students interested in learning more about the topics discussed in class. A "tweet" is a message sent to one's followers.

The tweet should be succinct and to the point, no more than 140 characters in length. A search for and subsequent tracking of a particular topic may be made using the hashtag. When a certain hashtag is widely used, it is said to be trending. Due to the two-way nature of Twitter interactions, one user may have both followers and followers. If someone follows you on Twitter and their account is public, you will be able to see all of their tweets, but that doesn't mean they can see yours. If you follow someone and they follow you back, the person you followed can see your tweets.

II. RELATED WORK

"Using machine learning to detect fake identities: bots vs humans."

The number of individuals who use fake profiles on SMPs to spread misinformation and harm others is rising rapidly. Regrettably, little work has been done to far to identify human-created fraudulent identities, particularly on SMPs. In contrast, there are several instances when machine learning models effectively uncovered false accounts generated by bots or computers. Artificial factors like the 'friend-to-follower ratio' were crucial to the success of these machine

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 4 April 2023, pp: 1217-1219 www.ijaem.net ISSN: 2395-5252

learning models when applied to bots. Attributes like 'friend-count' and 'follower-count' found in user profiles on SMPs were mined to create these functions. This study discusses research that uses these similar manufactured qualities to a dataset of false human accounts in an effort to improve the accuracy with which fake identities produced by humans can be uncovered in SMPs.

"Real-time detection of content polluters in partially observable Twitter networks."

It is well-known that content polluters, or bots that hijack a discussion for political or commercial objectives, offer a dilemma when trying to foresee events, predict elections, and tell true news from false news using social media data. Current best practises for detecting this sort of bot rely on using massive amounts of network data as features for machine learning models, which presents a significant computational challenge. Most apps that use real-time social media data for prediction lack access to such datasets. In this study, we create a system for monitoring live social media datasets in search of content polluters. We apply our approach to the challenge of predicting events of civil disturbance in Australia, and we do it by identifying content pollutants from individual tweets without gathering social network or history data from individual accounts. We highlight some of the out-of-the-ordinary bot traits present in our dataset and provide measures for spotting fake profiles. Following this, we ask a number of research questions on bot detection, such as how effective Twitter is in identifying content polluters and how well state-of-the-art algorithms do at detecting bots in our dataset.

III. METHODOLOGY

Some strategies rely on feature extraction from user profiles, blog posts, or social networks, while others employ machine learning to do so. These methods have shown promise, but they call for extensive user input as well as substantial time and energy investments. In recent years, deep-learning techniques have excelled their predecessors in speed and efficiency, all while requiring no manual input from the user. Recently, research into deep learning has been preoccupied on extracting characteristics from text rather than traditional data. This occurs because deep-learning algorithms can identify and follow subtle patterns in the accessible texts, which may elude a more conventional method. Provided a recent assessment and analysis of studies relating to spam filtering on Tweets and the information supplied within. Took use of a paid Twitter API with access to premium features to collect a large collection of tweets. An example dataset with URL categorization and object recognition from real-world data might be used as a reference for other academics. Making use of machine learning techniques, a new framework is being developed to identify rogue accounts on Twitter. This framework will integrate textual elements with metadata traits. We looked at the efficacy of merging easily accessible information with textual data by integrating URL-based features to spot spam Twitter accounts. Using URL-based characteristics of deep learning and machine learning algorithms VGG19 and logistic Regression applied to spam filtering in the Twitter network, a comparison is made between the proposed framework and the most renowned models.

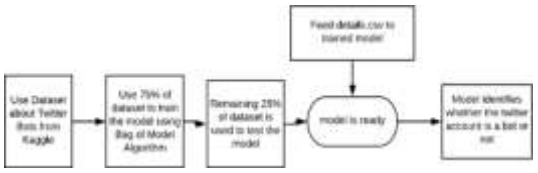

Fig.1 : System Architecture

IV. RESULT AND DISCUSSION

Just load the Tweets dataset after uploading it. Choose all tweets in a dataset and read them using the "Run Module 1 (Extract Tweets)" button.

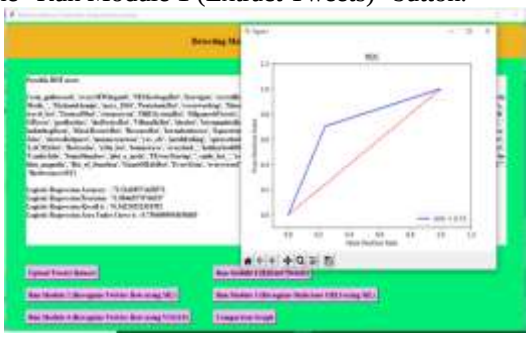

Fig.2: Extracting the tweets

'Run Module 2 (Recognize Twitter Bots with ML)' button to begin identifying BOTS users and then using logistic regression. ML Logistic Regression ML Accuracy was Determined to Be 73%.

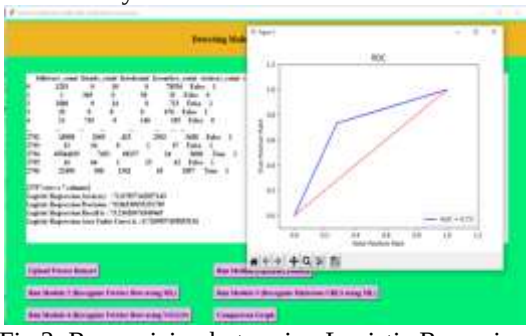

Fig.3: Recognizing bots using Logistic Regresion

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 4 April 2023, pp: 1217-1219 www.ijaem.net ISSN: 2395-5252

Above, we see that the accuracy of our URL predictions is 72%, and in the last column, 1 denotes a non-malicious URL, while a 0 suggests a malicious URL.

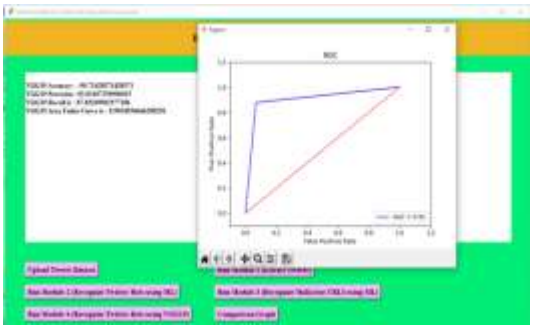

Fig.4: Recognizing the Bots Using VGG19

"Run Module 4 (Twitter Recognition with VGG19)" as well as VGG19's Accuracy Is 90%

Fig.5: Performance Comparision
Compared to other bots, VGG19 BOT & URL Accuracy is high

V. CONCLUSION

In our study, we created a method for automatically spotting Twitter bots. As compared to logistic regression, the best model for train data was VGG19's bag of words approach due to its superior accuracy. Consequently, the Twitter bots were successfully recognized by applying word algorithms to real-time data.

REFERENCES

[1]. P Subba Rao. "Applications of machine learning in environmental engineering" JOURNAL OF Science,Technology and evelopment (2021): 643-648

[2]. P Subba Raol. "Using an Energy Efficient Extended Leach Work with Multilevel Clustering Approach" from JOURNAL of Research Publication and Reviews (2020): 60-64

[3]. Mahammad, Farooq Sunar, et al. "Prediction Of Covid-19 Infection Based on Lifestyle Habits Employing Random Forest Algorithm." JOURNAL OF ALGEBRAIC STATISTICS 13.3 (2022): 40-45.

[4]. Devi, M. Sharmila, et al. "Machine Learning Based Classification and Clustering Analysis of Efficiency of Exercise Against Covid-19 Infection." JOURNAL OF ALGEBRAIC STATISTICS 13.3 (2022): 112-117.

[5]. Bhaskar, P., Mahammad, F. S., Kumar, A. H., Kumar, D. R., Khadar, S. A., Khan, P. M., & Reedy, P. V. S. (2022). Machine Learning Based Predictive Model for Closed Loop Air Filtering System. JOURNAL OF ALGEBRAIC STATISTICS, 13(3), 609-616.

[6]. Gowthami, V., et al. "Knowledge Based System for Immunity Improvement Against Covid-19 Infection." JOURNAL OF ALGEBRAIC STATISTICS 13.3 (2022): 01-07.

[7]. Mahammad, Farooq Sunar, et al. "Heuristics Approach Based Expert System for Covid-19 Infection Susceptibility." JOURNAL OF ALGEBRAIC STATISTICS 13.3 (2022): 46-51.

[8]. Reddy, E. Madhusudhana, and P. Bhaskar. "Able Machine Learning Method for classifying Disease-Treatment Semantic Relations from Bio-Medical Sentences." vol 5 (2018): 5..