Section A-Research paper
ISSN 2063-5346

Detecting Malicious Twitter Bots Using Deep Learning

DR. B. Narendra Kumar
Professor, HOD, Department of IT
Sridevi Women's Engineering College, Telangana
swecnarendra@gmail.com

Daram Lavalika
BTech Student, Department of IT
Sridevi Women's Engineering College, Telangana
lavalika40@gmail.com

Sathyaboina Sanjusha
BTech Student, Department of IT
Sridevi Women's Engineering College, Telangana
satyaboinasanjusha@gamil.com

Kandhukurthi Sindhu
BTech Student, Department of IT
Sridevi Women's Engineering College, Telangana
sindhukandhukurthi@gmail.com

ABSTRACT— Today, Tweet is used often and has significant meaning in the lives of many people, including businesspeople, media professionals, politicians, and others. Twitter, one of the most widely used social networking services, allows its users to express their views 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 become a breeding ground for criminals who use anonymity to commit crimes. Recognizing Twitter bots is crucial because of the potential threat they represent to other users. Thus, it is essential that tweets be posted by actual individuals and not by Twitter bots. The Twitter feed is being spammed by a bot. Therefore, recognizing bots helps in recognizing spam communications. Features extracted from Twitter accounts are utilized by machine learning algorithms to determine whether or not a user is authentic. In this research, we used three different machine learning techniques—a Decision Tree, a Random Forest, and a Multinomial Naive Bayes—to establish whether or not an account was genuine. We evaluate the classification performance and accuracy of several methods. Approximately 89% accuracy is achieved by the Multinomial Naive Bayes approach, 90% by the Random Forest algorithm, and 93% by the Decision Tree method. It follows that Decision tree achieves higher accuracy than both Random Forest and Multinomial Nave Bayes..

INTRODUCTION

Twitter's rapid expansion makes it one of the most promising social media sites. Users may share news and opinions on current events and other topics. Individuals who a

user has shown interest in or who shares their views may be followed. Users may instantly tweet to their followers. By retweeting, your message will reach a wider audience. The volume of tweets naturally rises during major live performances like sports games and award presentations. Twitter may be accessed from a computer or a mobile device. Product sales may be boosted and large funds can be generated via paid advertising. Students may use Twitter to find out more about the topics discussed in class. Tweets are messages sent to someone who has chosen to follow you. Tweets are limited to 140 characters and should be succinct. to locate and keep track of a certain topic, a hashtag (#) may be used. When a hashtag gets widely used, it is said to be trending. Links on Twitter go in both directions, enabling users to follow and be followed. Even if the person you're following on Twitter also has a public account, only you are going to be able to view your tweets until you choose to make them public. If someone follows you on Twitter, they will be able to see your tweets.

Twitter followers are inundated with automated tweets. The ability to identify bots is crucial for identifying fake users and protecting genuine users from malicious information. A Twitter chatbot is a piece of software that can send out tweets on a user's behalf. Spam is only one of many activities that bots are programmed to do.

One) Twitter bots are programmed to spread false information and rumors.

2) in the direction of damaging someone's reputation.

Third, forged letters may be used to steal credentials.

4) People are directed to malicious sites.

5) aiming to change someone's or an organization's opinion, maybe by swaying how well-liked they are

RELATED WORK

Using artificial intelligence to distinguish between robot and human identities

More and more individuals are using SMPs to harass, threaten, and otherwise hurt others while hiding their true identities behind fake profiles. There hasn't been a lot of work done so far to figure out how to spot human-made fake identities, especially with reference to SMPs. However, machine learning models have been successfully used to identify countless cases of fake accounts created by bots or computers. Many of these machine learning systems relied on fabricated variables, such as the "friend-to-followers ratio," to detect and avoid being fooled by bots. These functions make use of information that is already present in user profiles on SMPs, such as "friend-count" and "follower-count." Using the same intended qualities on a collection of fake human accounts, the work in this article aims to enhance accurate recognition of false human identities on SMPs.

Pollutant material in partly viewable Twitter networks detected in real time

Identifying real news from false news in social network data is a recognized challenge, and content pollution or bots which hijack a debate for political or commercial reasons are a well-known concern when trying to foresee events or predict the outcomes of elections. Identifying this kind of bot is exceedingly challenging due to the fact that modern approaches utilise massive quantities of network data as inputs for machine learning models. However, such datasets are often inaccessible in systems that use streaming internet data for immediate form event prediction. In this research, we develop a method for tracking content trolls across live-streamed social media data sets. Our method allows us to anticipate when civil disturbance will occur in Australia based on tweet content without access to a user's social network or browsing history. Our dataset reveals certain peculiarities in these automated accounts, and we provide measures for spotting them. We then pose a number of research questions related to bot identification, including how successful Twitter is at detecting content spammers and how well state-of-the-art approaches perform on our dataset.

Using Deep Learning to Spot Fake Twitter Followers

Because of the rise in popularity of Twitter, a new spam industry has developed. Among the many services available in this marketplace are the selling of phony accounts, the promotion of Twitter spam via affiliate programs, and the coordination of large-scale spam campaigns. Users of Twitter have also started buying fake followers to boost their popularity. In this project, we showcase our developed machine learning approaches for spotting fake Twitter followers. On a phony account we created for research purposes, we found 13,00 paid fake followers and 5,382 genuine ones. We then identified a set of characteristics that make fake followers easy to spot. Using these characteristics, we were able to train machine learning algorithms to distinguish between fake and genuine users. We have used sure machine learning methods with the goal of achieving high detection accuracy and others with the goal of achieving low accuracy.

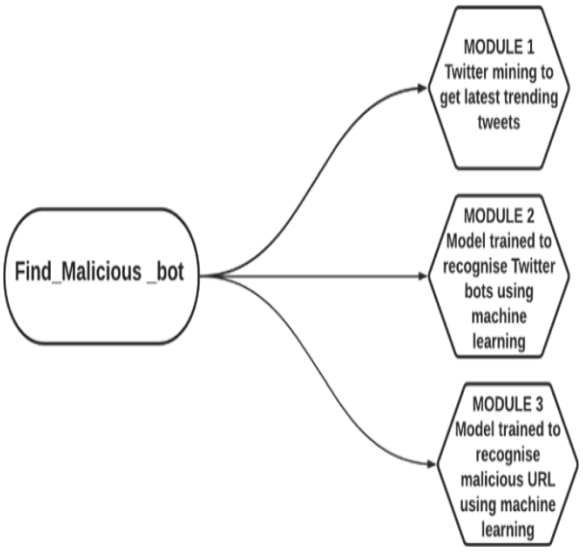
I See a Bot: Developing a Two-Stage Classifier to Identify Twitter Bots.

It is estimated that as much as half of all Twitter activity is created by bots [1] – algorithmically programmed accounts set up to promote products, propagate spam, or sway public opinion. It has been shown that accounts suspected of being bots accounted for as much as 20% of all Twitter activity relating to the 2016 U.S. presidential election. Bots were also employed to propagate false material about French presidential candidate Emmanuel Macron and to aggravate the ongoing dispute in Qatar, according to available evidence. Detection of bots is necessary for identifying bad actors in the "Twitterverse" and protecting legitimate users from

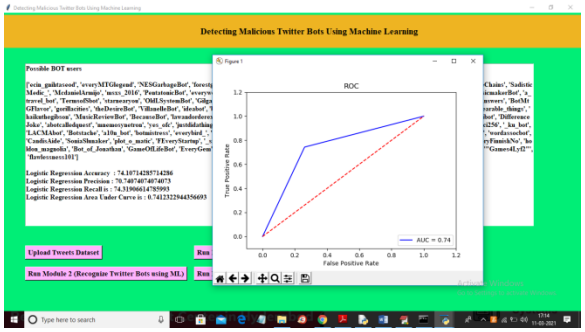
misinformation and malicious intent. Despite years of study in this area, current algorithms are not nearly as effective as humans [2]. Our study aimed to develop a binary classifier that, given a Twitter user's profile and tweet history, could identify whether the person was a bot or a human. The final product of this kind of classifier (see examples on page 5) would be a browser extension that could instantly assess a user's account. Our method, together with a working prototype, can identify an anonymous Twitter account using data provided by the Twitter API. The check screenname.py program is able to swiftly categorize any Twitter user by using the Twitter API.

METHODOLOGY

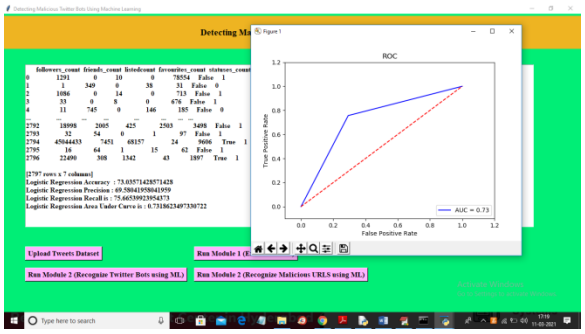
Several methods rely on the use of machine learning to extract information from user profiles, postings, or social networks. Although these methods have shown promising results, they call for extensive user input, a large investment of resources, and additional processing time. In recent years, deep-learning technologies have outperformed their predecessors in speed and efficiency, all while requiring no manual input from the user. In current deep-learning-based research, feature extraction from text has replaced ordinary data as the primary focus. This occurs as a result of deep-learning algorithms' capacity to identify and follow latent patterns within the accessible texts, patterns that may elude a more conventional method.Provided a recent overview and analysis of studies relating to Twitter spam detection and the information it contains.Utilizing a premium Twitter API with access to capabilities that aren't accessible to the general public, we were able to collect a sizable tweet dataset. The data gathered in the field might be used to create a benchmark set for other academics to use in their own work on topics such as URL categorization and feature extraction.In order to identify rogue accounts on Twitter, researchers have developed a unique framework that blends text-based and metadata-based elements using machine learning techniques.We looked at the efficacy of mixing open metadata and textual data by integrating URL-based features in spotting Twitter spam accounts.Comparison of the proposed framework to state-of-the-art models for Twitter spam detection utilizing URL-based characteristics of neural networks and deep learning.



RESULT AND DISCUSSION



The y-axis shows the proportion of correct predictions made compared to those made incorrectly (the False Positive Rate), while the x-axis shows the proportion of incorrect predictions made (the False Positive Rate), and the red line shows the False Rate.



The incorrect prediction rate (in red) is larger than the rate of malicious URLs (blue line) in the ROC plot.

CONCLUSION

In our study, we created a method for automatically spotting Twitter bots. When comparing the accuracy of the Decision Tree, Naive Bayes, and Random Forest models on the test data, the Bag of Words technique came out on top with a score of 96.7%. As a result of applying word algorithms to real-time data, Twitter bots were successfully identified.

REFERENCES

[1] Van Der Walt, Estée, and Jan Eloff. Using machine learning to detect fake identities: bots vs humans. IEEE Access 6 (2018): 6540-6549.

[2] Sever Nasim, Mehwish, Andrew Nguyen, Nick Lothian, Robert Cope, and Lewis Mitchell. Real-time detection of content polluters in partially observable Twitter networks. arXiv preprint arXiv:1804.01235 (2018).

[3] Khalil, Ashraf, Hassan Hajjdiab, and Nabeel Al-Qirim. Detecting Fake Followers in Twitter: A Machine Learning Approach. International Journal of Machine Learning and Computing 7, no.6(2017).

[4] Wetstone, Jessica and Sahil R. Nayyar. I Spot a Bot: Building a binary classifier to detect bots on Twitter. (2017).

[5] Karataş, Arzum, and SerapŞahin. A Review on Social Bot Detection Techniques and Research Directions. In Proc. Int. Security and Cryptology Conference Turkey, pp. 156-161. 2017.

[6] Chavoshi, Nikan, Hossein Hamooni, and Abdullah Mueen. Identifying correlated bots in twitter. In International Conference on Social Informatics, pp. 14- 21. Springer, Cham, 2016.

[7] Perdana, Rizal Setya, Tri HadiahMuliawati, and Reddy Alexandro. Bot spammer detection in Twitter using

tweet similarity and time interval entropy. JurnalIlmuKomputer dan Informasi 8, no. 1 (2015): 19-25.

[8] Haustein, Stefanie, Timothy D. Bowman, Kim Holmberg, Andrew Tsou, Cassidy R. Sugimoto, and Vincent Larivière. Tweets as impact indicators: Examining the implications of automated bot accounts on T witter. Journal of the Association for Information Science and Technology 67, no. 1 (2016): 232-238.