International Journal for Innovative Engineering and Management Research

A Peer Revieved Open Access International Journal

www.ijiemr.org

COPY RIGHT

ELSEVIER

SSRN

IJIEMR Transactions, online available on 11th Jan 2023. Link

:http://www.ijiemr.org/downloads.php?vol=Volume-12&issue=Issue 01

DOI: 10.48047/IJIEMR/V12/ISSUE 01/66

Title **Detection of Malicious Social Bots Using Machine Learning With URL Features in Twitter Network**

Volume 12, ISSUE 01, Pages: 695-702

Paper Authors

A. Shailaja, K. Sridevi

International Journal for Innovative Engineering and Management Research ★

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

Detection of Malicious Social Bots Using Machine Learning With URL Features in Twitter Network

A. Shailaja[1], K. Sridevi[2]

[1]Department of Information Technology, G Narayanamma
Institute ofTechnologyand Science, India
Email: shailaja17achi@gmail.com
[2]Department of Information Technology, G Narayanamma
Institute ofTechnologyand Science, India
Email: kandulasridevia2@gmail.com

Abstract

Emergence of internet as a widespread tool for massive-scale and real-time communication, resulted in extensive use of bots for disingenuous purposes. Twitter has become so dominant to phishers to spread phishing attacks using bots due to its vast propagation but hard to be detected nature unlike other services because of its fast spread inthe network. In this paper, we proposed a bot detection model that has two modules ,the first stage is classifying the tweets as “malicious tweet” and “legitimate tweet” by integrating URL based features using models based on machinelearning and second stage is analyzing the twitter users and classifying them as bot or normal user, using a set of reduced simple counter features that could potentially overcome the limitations of computationally expensive models that require large numbers of features, large labeled datasets for training purposes, and access to the recent history of activity of the account profile to scrutinize. Experimentation has been performed on public data sets using three following classifiers namely, Naive Byes, SVM (Support vector Machine) and MLP classifier and the results illustrate that the proposed system achieves improvement while exploiting a small and interpretable set of features compared with existing approaches for MSBD.

Keywords: Malicious social bots, Machine Learning, MLP, Malicious URL, Online SocialNetworks (OSNs).

Introduction

The rise of smart devices, technology and high-speed internet have resulted in the more usage of Online social networks (OSNs), this leads to interaction between humans on a large volume. As per the research by Kemp [1], 4.2 billion social network accounts are present in the world, which is 53 percent of the total population. The monthly count of active accounts in twitter is 330 million [2]. Open platform of twitter to share the opinions gave it a redundant eventuality to impact the users. The influence of online services on the public resulted in the generation of automated accounts or bots. According to the study by Stefan et al [3], bots participated 66% of all tweeted links to popular websites. Social bots are an advantage when they act as helpers in aggregating and delivering news feeds and disadvantage when misused by creating bots for malicious activities. Bot is aset of software code in social network that function as a actual user in social online services [4], [5]. Moreover, malicious bots generate fake identities, change reviews, circulate spam-messages etc. [4]. Phishing attack, which is a type of social engineering that deceive victims by exposing them to a fraudulent site where their sensitive information is collected for performingillegal crimes.

Phishing attacks have now focused on online services like myspace, twitter, fb, etc., that basically target emailusers. Twitter has become so dominant to phishers to spread phishing attacks due toits vast propagation but hard to be detected nature unlike other services because of its fast spread in the network [6]. A malicious bot share tweets containing malicious links leaving legitimate users unprotected. Due to the presence of malicious bots in the network,twitter services are exposed to vulnerabilities. So, in this paper, a learning model is being proposed which will, with the help of built model, filter out those twitter accounts which have a bot like activity and are sharing malicious URLs. Most, of the methods introduced detect bots at the accountlevel,given a record of activity (e.g., a few hundred tweets posted by a user), the algorithm would determine whether the scrutinized account is a bot or not, they focus on the overall account's activity. Though quite successful, these approaches are expensive as they require significant amounts of data for each user to be scrutinized. So, there is the requirement of a tweet level bot detection that could overcome these limitations. The main objective is to develop a novel architecture that combines text-based and metadata-based features for detecting malicious bots on Twitter using machine learning algorithms. Next to investigate the effectiveness of combining accessiblemetadata with textual data integrating

URL based features when detecting malicious bots on twitter and to benchmark the proposed model with the existing models using URL based features of machine learning applied tobot detection in twitter network.

Related Work

In the approach by Sayyadiharikandeh M [7], the diversities of different kind of bots are handled by training classifiersspecialized for individual group of bots, and a bot-score is calculated for each classifier and the class with the highest bot-score determines the respective class.They have also done a cross-domain analysis of their classifier by testing it on separate datasets. They have used a high- dimensional feature set consisting 1200 features from six categories. Considering a high feature set including an account's actions and social connections improves accuracy but reduces scalability [8].

In [9] the author proposed a bot identification method based on click stream transition probability and clustering approach. In this they have also include feature related to timealong with user click stream pattern and clustering based on semi-supervised model. The experimental results of this approach on real time network showed that the accuracy for detection based on click patterns there is an increase of 12.8% when compared with quantitative behavior analysis.

G. Lingam [10], proposed a model, to detect social bot among legitimate users they used social graph characteristics. Here by using social features of user as state and learning movement among the states as action, they designed a deep Q- graph model in the twittersocial network by updating Q-value function. For this "state-action" function, to build probability values of transition between Q pairs they included all the Q-pairs.

In the paper [11], author considered the bot detection from Chinese Sina Weibo online social network. Here they used deep neural network(DNN) modelconsisting of attention mechanism along with recurrent bidirectionalgated unit and residual unit by including 30 features based on four different sections of timing, metadata, communication and content. This model obtained 0.98 accuracy. Since it has been developed for Sina Weibo, the efficiency of this for twitter has to be checked. The existing models for the identification of twitter bots assumes extensive access to social media data [12]. They detect the bots and block them. However, bots can create a new account and post new malicious tweets again. The malicious bots can manipulate the information present in the tweets. The interaction features are robust but extracting these features required more time because of extremely large volume of the network.

System Methodology

System Model

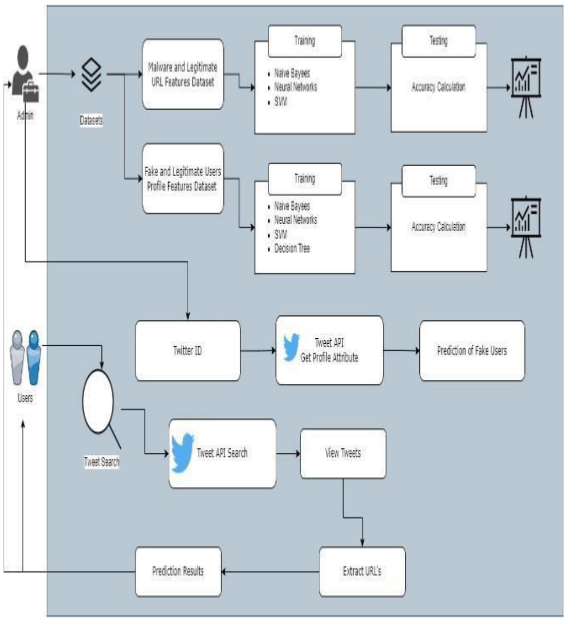


Figure 1: System Model

The figure above depicts the suggested system's technique. The suggested system foranalyzing and detecting malicious bots in twitter is implemented using the following two stages:

1. In the first stage we use three machine learning models to classify the tweets by integrating URL based features, by extracting the urls posted in the tweets into one of the two categories "phishing tweet" or "legitimatetweet".
2. In the second stage, we use a user analysis mechanism based on simple reduced set of counter features to detect bots, which can reduce the spread of phishing links on Twitter.

Classification of the tweets by analyzing URLs present in the tweets

Based on our research, Google's secure browsing is used by twitter to identify and secure from phishing links. The new URLs that have not been added to the database cannot be detected by this approach. These services need four days on average to add new site into the database, but tweets will be accessed within a day after posting. This limits the users from being protected in realtime. In our approach we used machine learning for detection.

Our tweet classification process is divided into two stages building stageand detection stage. In building stage, we used dataset having "malicious urls" and "legitimate urls" including many features out of which we used the most efficient features to build three machine learning models. This building stage is implemented during the learning process. In next detection stage we extract the urls from the tweets collectedusing tweeter API. From these urls we have extracted the features we used in the building stage and used the model that gave us the highest accuracy in the learning process for prediction of malicious tweets based on URLfeatures. We implement this, whenever a tweet is extracted.

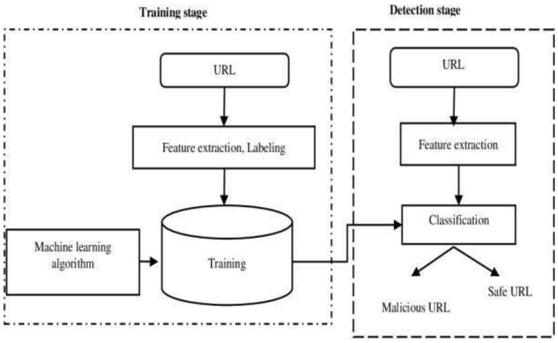
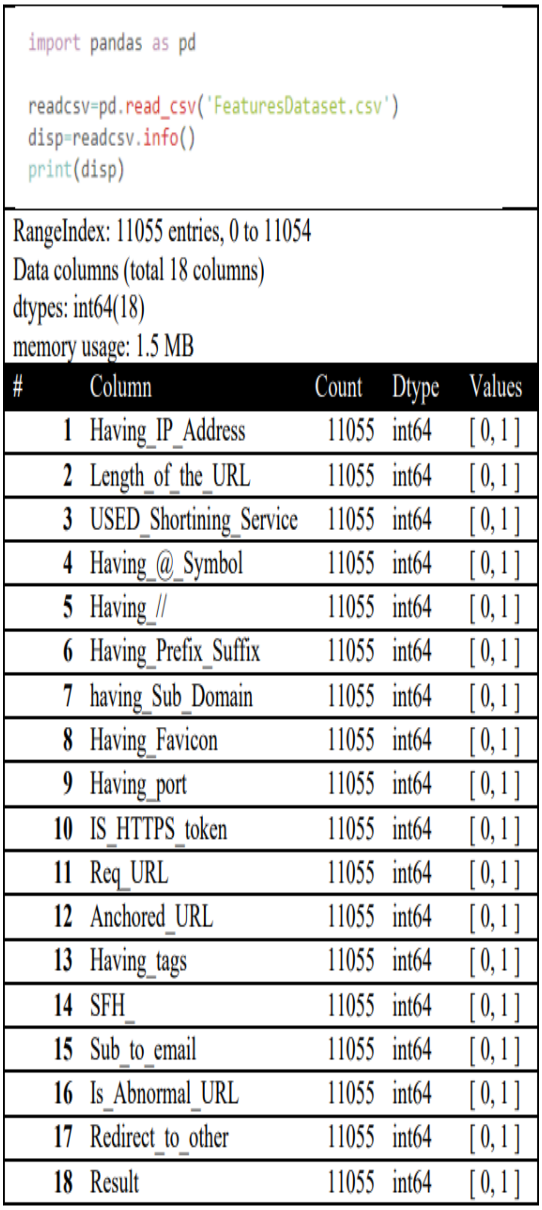


Figure 2: Malicious URL Detection

1) Dataset: This module used the “Phishing Websites Data Set” for the learning and detection of phishing websites. This dataset is published in the UCI repository [Link]. This dataset is accessible in the CSV file format and labeled with two classes“phishing” (0) and “legitimate” (1).
2) Description of the dataset: This section described the “Phishing Websites Data Set” using Pandas Python API, as shown below. This dataset has 18 columns with 17features of URLs and are labeled as “Phishing URL” or “legitimate URL”. Every column has 11055 rowsdata of integer data (0 and 1).
3) Features representation: First we extract the URLs from the retrieved tweets. Next using python, from the URLs we derive the attributes used in thebuilding stage of the models and designate them as integer vector,which is having the following function:

f = 1 presence of the attribute

f = 0 absence of the attribute



```
import pandas as pd

readcsv=pd.read_csv('FeaturesDataset.csv')
disp=readcsv.info()
print(disp)
```

RangeIndex: 11055 entries, 0 to 11054
Data columns (total 18 columns)
dtypes: int64(18)
memory usage: 1.5 MB

#	Column	Count	Dtype	Values
1	Having_IP_Address	11055	int64	[0, 1]
2	Length_of_the_URL	11055	int64	[0, 1]
3	USED_Shorting_Service	11055	int64	[0, 1]
4	Having_@_Symbol	11055	int64	[0, 1]
5	Having_//	11055	int64	[0, 1]
6	Having_Prefix_Suffix	11055	int64	[0, 1]
7	having_Sub_Domain	11055	int64	[0, 1]
8	Having_Favicon	11055	int64	[0, 1]
9	Having_port	11055	int64	[0, 1]
10	IS_HTTPS_token	11055	int64	[0, 1]
11	Req_URL	11055	int64	[0, 1]
12	Anchored_URL	11055	int64	[0, 1]
13	Having_tags	11055	int64	[0, 1]
14	SFH_	11055	int64	[0, 1]
15	Sub_to_email	11055	int64	[0, 1]
16	Is_Abnormal_URL	11055	int64	[0, 1]
17	Redirect_to_other	11055	int64	[0, 1]
18	Result	11055	int64	[0, 1]

Twitter User Account Analysis

The next stage of our model is the interpretation of user accounts for bot detection. In this second module, first we train the model using decision tree, multi-layer perceptron, SVC, NB algorithms with the dataset of bot and human accounts. We use the model that gave the highest accuracyfor the prediction of bot account.

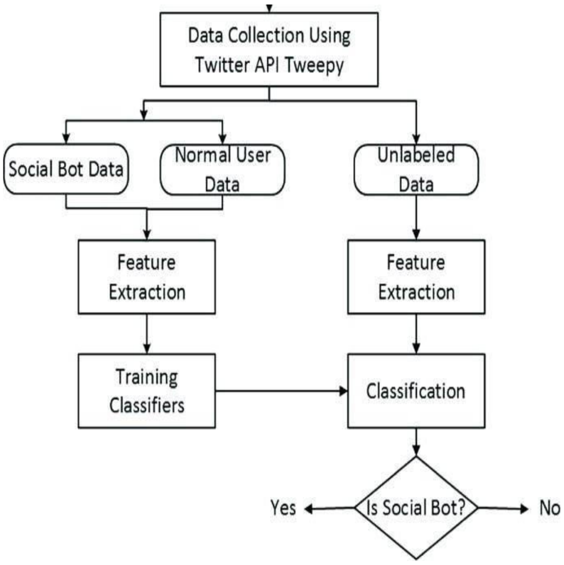


Figure 3: User Account Analysis

4) Dataset: The dataset we used to predict users as "Bot" or "Legitimate Account" contain useraccounts labeled as “bot” or “human”, along with the five counter features that do not require data preprocessing.

Listed_Count
Followers_Count
Favorites_Count
Friends_Count
Statuses_Count

5) Features representation: Once theaccount is identified as posting malicious url, then the five attributes used in dataset are collected in the JSON format using “user” object of the Twitter API and represented as integer vector. These five features illustrate the account usage at basic level. Then using these features and multi-layer perceptron model that gave highest performance, we classify the account.

Experimental Result

In this section, we present the prediction results that we have obtained, with the three classifiers, for malicious URLs, then for bot. Prediction results for malicious bot: Concerning the prediction results of malicious bot, obtained the best accuracy, which is 90.26% with the MLP classifier. As depicted in figure 18, the prediction results of malicious URLs withthe Naive Bayes and SVM (Support Vector Machine) model are acceptable with an accuracy of 80% and 81%respectively, but the obtained results by the MLP classifier are more with an accuracy of 92%. Therefore, the accuracy score produced by the MLP Classifier approach is the greatest.

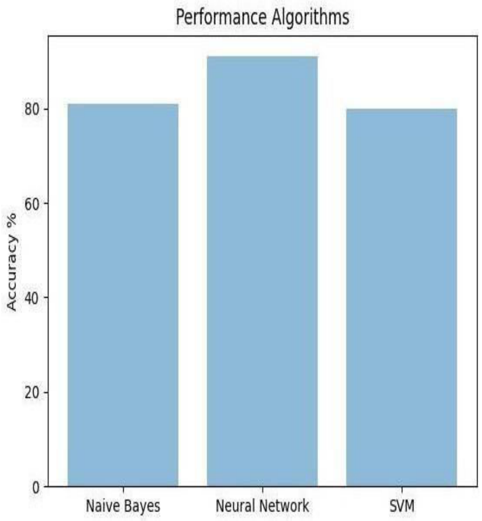


Figure 4: Accuracy

We have implemented our model as atool, which can be used to detect malicious tweets and malicious bots. An interface of the application, for searching the tweets based on the keyword is shownin figure 5, and the latest tweets collectedbased on the searched keyword and the classification of these tweets as“legitimate” and “phishing” is shown in the figure 6. Thus, if a malicious tweetis identified, the user can see which user account has posted the malicious tweet. The last step of our application is the analysis of the user account (see Figure 7), to detect weather the account is a bot account or normal account.

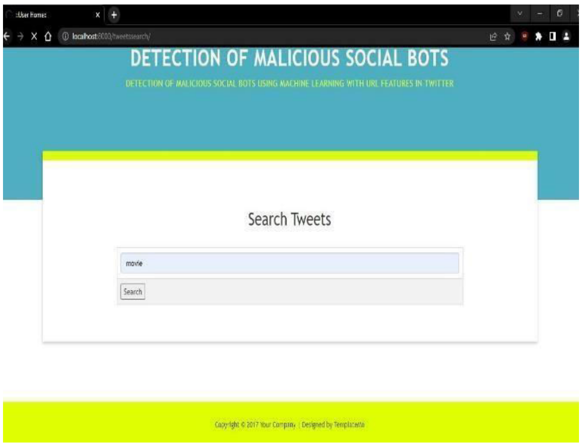
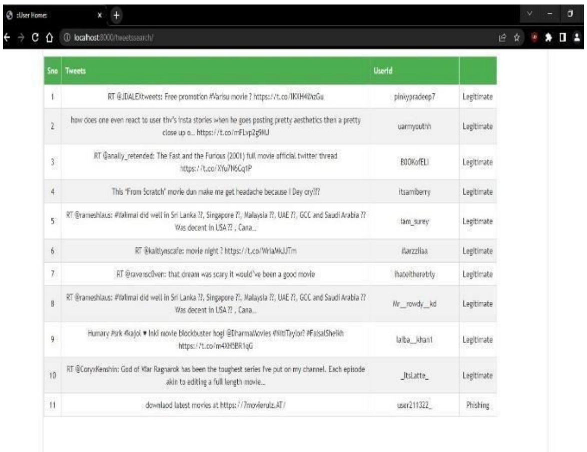

:User Home:
localhost:8000/tweetssearch/
DETECTION OF MALICIOUS SOCIAL BOTS
DETECTION OF MALICIOUS SOCIAL BOTS USING MACHINE LEARNING WITH URL FEATURES IN TWITTER
Search Tweets
movie
Search
Copyright © 2017 Your Company | Designed by Templateste

Figure 5: Search Tweets



Sno	Tweets	UserId	
1	RT @JDALEXtweets: Free promotion #Varisu movie ? https://t.co/IKXH4WzGu	pinkypradeep7	Legitimate
2	how does one even react to user thv's insta stories when he goes posting pretty aesthetics then a pretty close up o... https://t.co/mFLvp2g9MJ	uarmyouthh	Legitimate
3	RT @anally_retended: The Fast and the Furious (2001) full movie official twitter thread https://t.co/XYu7N6Cq1P	BOOKofELI	Legitimate
4	This 'From Scratch' movie dun make me get headache because I Dey cry???	itsamiberry	Legitimate
5	RT @rameshlaus: #Valimai did well in Sri Lanka ??, Singapore ??, Malaysia ??, UAE ??, GCC and Saudi Arabia ?? Was decent in USA ?? , Cana...	iam_surey	Legitimate
6	RT @kaitlynscafe: movie night ? https://t.co/WrIaMkJJTm	Marzzliaa	Legitimate
7	RT @ravenscOven: that dream was scary it would've been a good movie	ihateitheretrly	Legitimate
8	RT @rameshlaus: #Valimai did well in Sri Lanka ??, Singapore ??, Malaysia ??, UAE ??, GCC and Saudi Arabia ?? Was decent in USA ?? , Cana...	Mr__rowdy__kd	Legitimate
9	Humary #srk #kajol ♥ Inki movie blockbuster hogi @DharmaMovies #NitiTaylor? #FaisalSheikh https://t.co/m4XH5BR1qG	laiba__khan1	Legitimate
10	RT @CoryxKenshin: God of War Ragnarok has been the toughest series I've put on my channel. Each episode akin to editing a full length movie...	_ItsLatte_	Legitimate
11	download latest movies at https://7movierulz.AT/	user211322_	Phishing

Figure 6: Collected Tweets

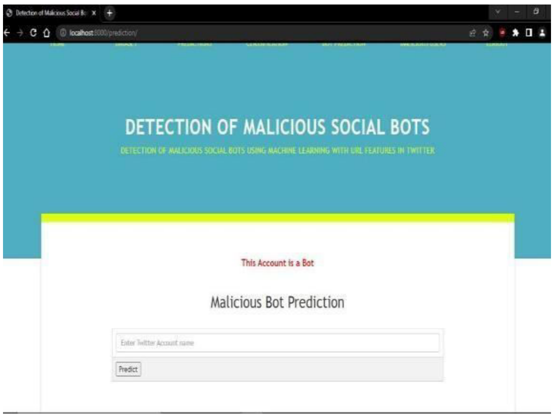

Detection of Malicious Social Bo
localhost:8000/prediction/
DETECTION OF MALICIOUS SOCIAL BOTS
DETECTION OF MALICIOUS SOCIAL BOTS USING MACHINE LEARNING WITH URL FEATURES IN TWITTER
This Account is a Bot
Malicious Bot Prediction
Enter Twitter Account name
Predict

Figure 7: Bot Prediction

Conclusion

In this work, we have tackled the problem of malicious bots in twitternetworks. Twitter bots are very dangerous if they are broadcasting malicious links. So, we proposed an approach comprising of two steps, which are analysis of links, next the analysis of accounts using machine learning that can accurately detect bots which are posting malicious URLs. The prediction module extracts real-time tweets from theTwitter server and implanted the developed model for prediction of malicious bots which shares malicious websites. This model can help in reducing APT attacks, phishing attacks etc. as the beginning point for these types of attack is generally the malicious links circulated through social networking sites. Once if the source of these malicious URLs isdetected then these attacks can be prevented. The results illustrated that the proposed system achieves improvement while leveraging a small and interpretableset of features compared with existing approaches.

References

[1]Kemp S, “The latest insights into the‘state of digital’ approaches,” (2021).

[2]Tankovska H, “Number of monthly active twitter users worldwide from 1st quarter 2010 to 1st quarter 2019,” (2021).

[3] Stefan W, Solomon M, Aaron S et al, “Bots in the twitter sphere,” in pewresearch 2018.

[4] P. Shi and K. K. R. Choo, “Detecting malicious social bots based on clickstream sequences,” IEEE Access, 2019.

[5] R. R. Rout, and G. Lingam, “Adaptive deep Q-learning model for detecting social bots and influential users in online socialnetworks,” in Application Intelligence,volume-49, Nov-2019.

[6] Nair MC, Prema S, “A distributed system for detecting phishing in Twitter stream,” in International Journal of Science Innovation Technology 2014.

[7] Sayyadiharikandeh M, Varol O, YangK- C et al, “Detection of Novel Social Bots by Ensembles of SpecializedClassifiers,” 2020

[8] Yang K-C and Menczer F, “Scalable and generalizable social bot detection through data selection,” in conference of AAAI on artifcial intelligence, pp 1096– 1103, (2020).

[9] K. K. R. Choo and P. Shi, “Detecting malicious social bots based on clickstream sequences,” in 2019.