



IJIRCCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Issue 5, May 2023

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.379



9940 572 462



6381 907 438



ijircce@gmail.com



www.ijircce.com

BOT Detection in Twitter Using ML

Prof. Mrs. A.R. Mate¹, Umbare Amit Prashant², Mule Tushar Dnyaneshwar³,
Garad Abhishek Madan⁴

Professor, Dept. of Information Technology, PREC, Loni, India¹

Student, Dept. of Information Technology, PREC, Loni, India²

Student, Dept. of Information Technology, PREC, Loni, India³

Student, Dept. of Information Technology, PREC, Loni, India⁴

ABSTRACT: Social bots are the most common malware on social platforms. They will spread fake news, spread rumors and even manipulate public opinion. Social robots are used to perform automated analysis services and provide better service to users. However, malicious social bots are also used to spread misinformation with real-world consequences. The hardest part of detecting bots on social media is understanding what modern social bots are capable of and determining the value of their behavior. A method for detecting malicious actors, including two unique options, provides an understanding of the psychology behind bot creation and custom distribution. The most important malicious social bot available finds ways to analyze the characteristics of its behavior. This activity is carried out only by social bots; therefore resulting in low analysis accuracy. A method of identifying malicious people in the community, which includes the selection of the two attributes that support the best truth and the classification rules. This distinguishes regular users from social bots. Experiments show that social content analysis is often used to identify social robots on online social platforms.

KEYWORDS: Social Media, Supervised Classification, Social Bots-Class Classifier, Machine Learning

I. INTRODUCTION

In online social networks, social bots are social accounts managed by programs that can act as programs. The increasing use of mobile devices such as Android and iOS devices has also led to an increase in the level and quality of user interaction through networks. It is important to detect and eliminate malicious social bots on online social networks. The most up-to-date methods of searching for malicious social robots describe the characteristics of their behavior. These bots can easily act as social bots. This reduces the accuracy of the analysis.

II. PROBLEM STATEMENT

In online communication, social robots are social accounts managed by machine-oriented services that perform social tasks that support the process. We expect to see changes in malicious bots and regular users by collecting and analyzing the consciousness state of the user's behavior to distinguish social bots from normal users, detect bad bots and cut harm from bad bots. . Therefore, it is necessary to detect and remove malicious social bots from online social networks. One of the biggest challenges in searching for bots on social media is understanding what a modern social bot will do and analyzing its behavioral characteristics. Therefore, we developed a unique method to support supervised learning to identify robots from large and uneven real-life data.

A. Motivation

- Bots behave like real users. i don't know the difference
- Understanding the psychology behind building robots to analyze.
- Use different learning systems.
- Treat the best model as a trade-off.
- Prevent malicious bots from contaminating social media sites.

B. Scope

• The detection method will be expanded and optimized to identify specific intents and targets of various malicious social robots.

- This request attempts to evaluate many aspects of the latest bad social bot behavior.
- Create maximum accuracy and discrimination between real users and bots.

C. Purpose

- Bot search aims to distinguish bots from humans and increase interest.
- For real-time social bot search on online social networking platforms.

III. METHODOLOGY

System Architecture

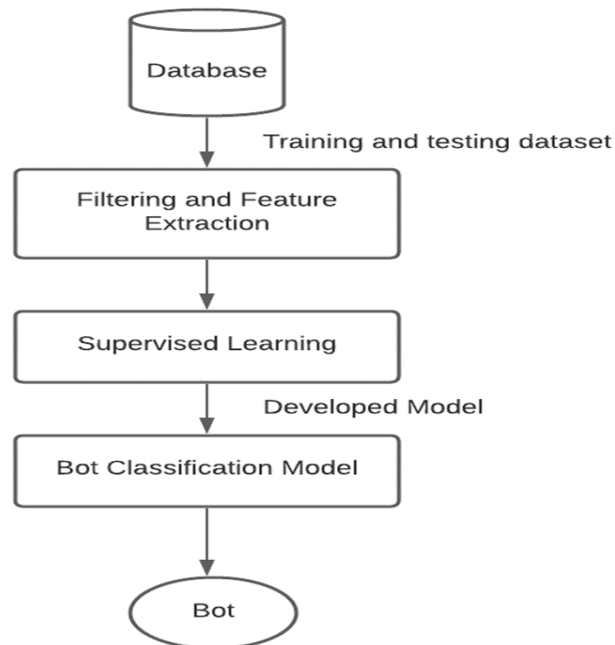


Fig1. System Architecture

- Before moving on to any method, you need to do some simple data cleaning tasks of each machine learning. Data cleaning refers to the detection of faulty, incomplete, incorrect, inaccurate or incomplete data sources and the necessary updating, modification or deletion operations accordingly.
- Feature extraction - It is a method of dimensionally reduction by which an initial set of raw data is reduced to additional manageable groups for further data processing.
- Supervised Learning - The training knowledge given to the machine is effective because the supervisor teaches the machine to predict results correctly. The purpose of the supervised learning algorithm is to find a function in the graph to map the input variable (x) to the output variable.
- Bot Classification Model - Bot Classification Model divides Twitter users into two groups as bots and regular users.

IV. TECHNIQUES

1. Data cleaning

In this model, we clean data and remove blank rows. And white papers don't buy anything. And for data consistency, where there are words like null or NaN they are replaced with None.

2. Feature Selection/Extraction

Feature selection is the process of selecting (relatively, roughly) a subset of features for the design. There are many reasons to use custom options: simplifies the structure, makes it easier for researchers/users to interpret, reduces training time, well avoids the curse, improves details by reducing overfitting (design, reduces variance)

Feature extraction creates the first data measure set and data produces the resulting results (features) designed to process rather than reproduce, which makes it easier to learn and expand on later, and sometimes better for human interpretation. Feature extraction is about size reduction.

V. RESULTS



Fig2 – Home Page

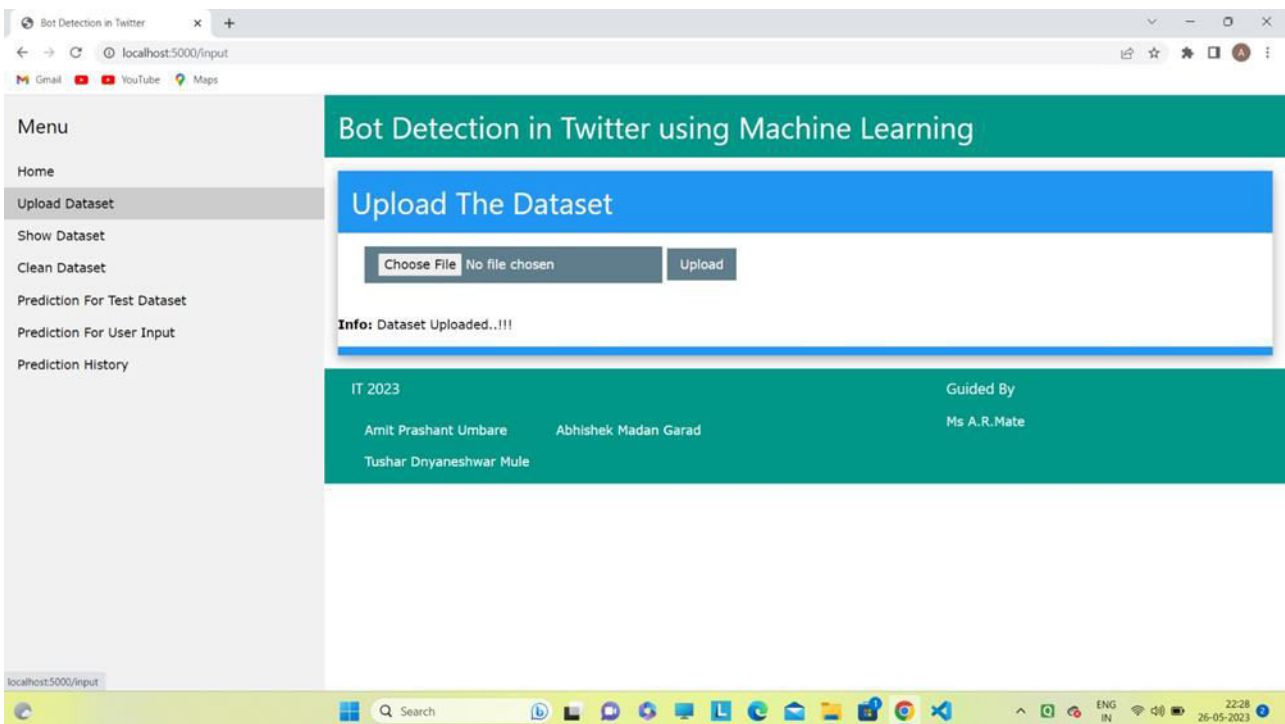
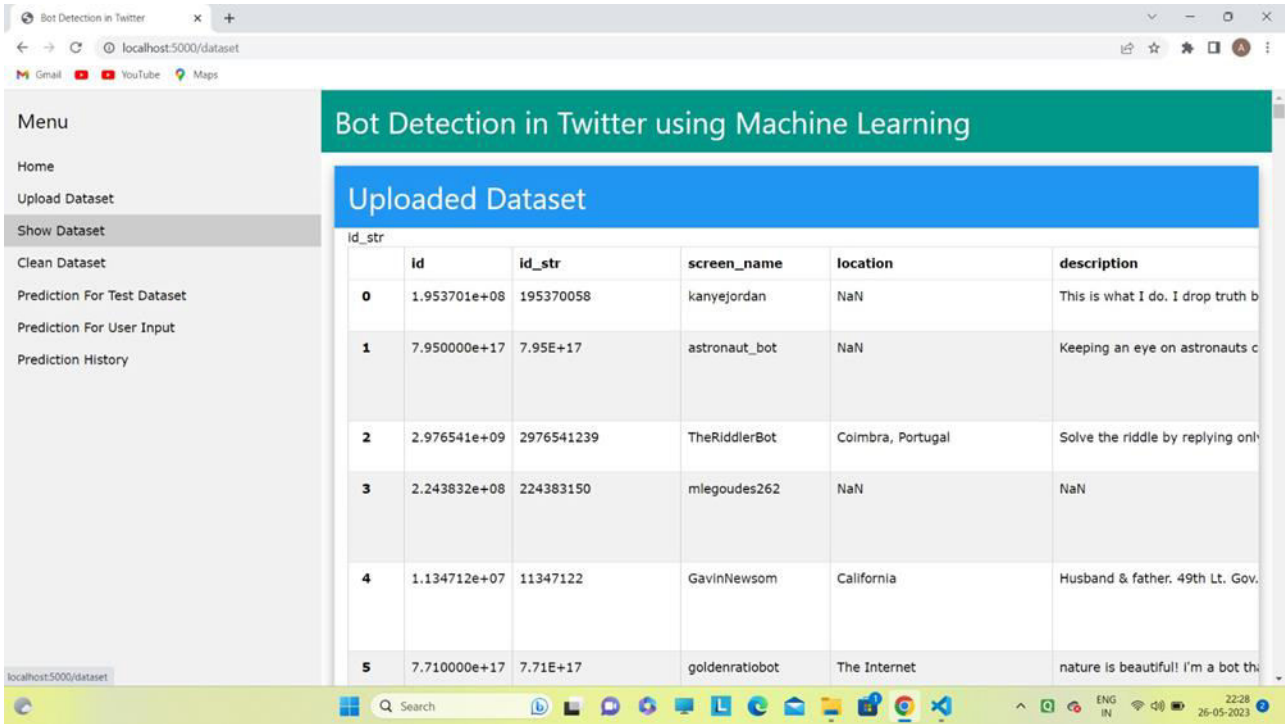


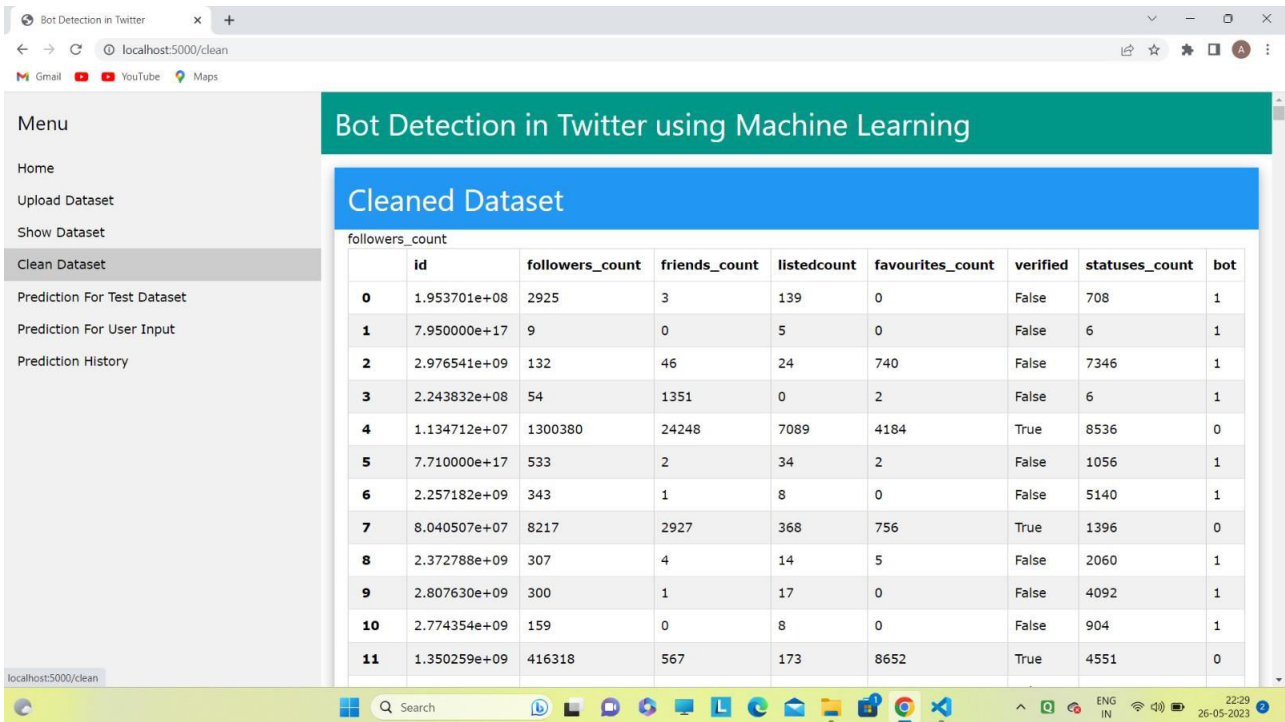
Fig3- Upload Dataset



The screenshot shows a web browser window with the URL localhost:5000/dataset. The page title is 'Bot Detection in Twitter using Machine Learning'. A sidebar menu on the left includes options like Home, Upload Dataset, Show Dataset, Clean Dataset, Prediction For Test Dataset, Prediction For User Input, and Prediction History. The main content area displays a table titled 'Uploaded Dataset' with the following data:

id	id_str	screen_name	location	description
0	1.953701e+08	kanyejordan	NaN	This is what I do. I drop truth b
1	7.950000e+17	astronaut_bot	NaN	Keeping an eye on astronauts c
2	2.976541e+09	TheRiddlerBot	Coimbra, Portugal	Solve the riddle by replying onl
3	2.243832e+08	mlegoudes262	NaN	NaN
4	1.134712e+07	GavinNewsom	California	Husband & father. 49th Lt. Gov.
5	7.710000e+17	goldenratiobot	The Internet	nature is beautiful! I'm a bot th

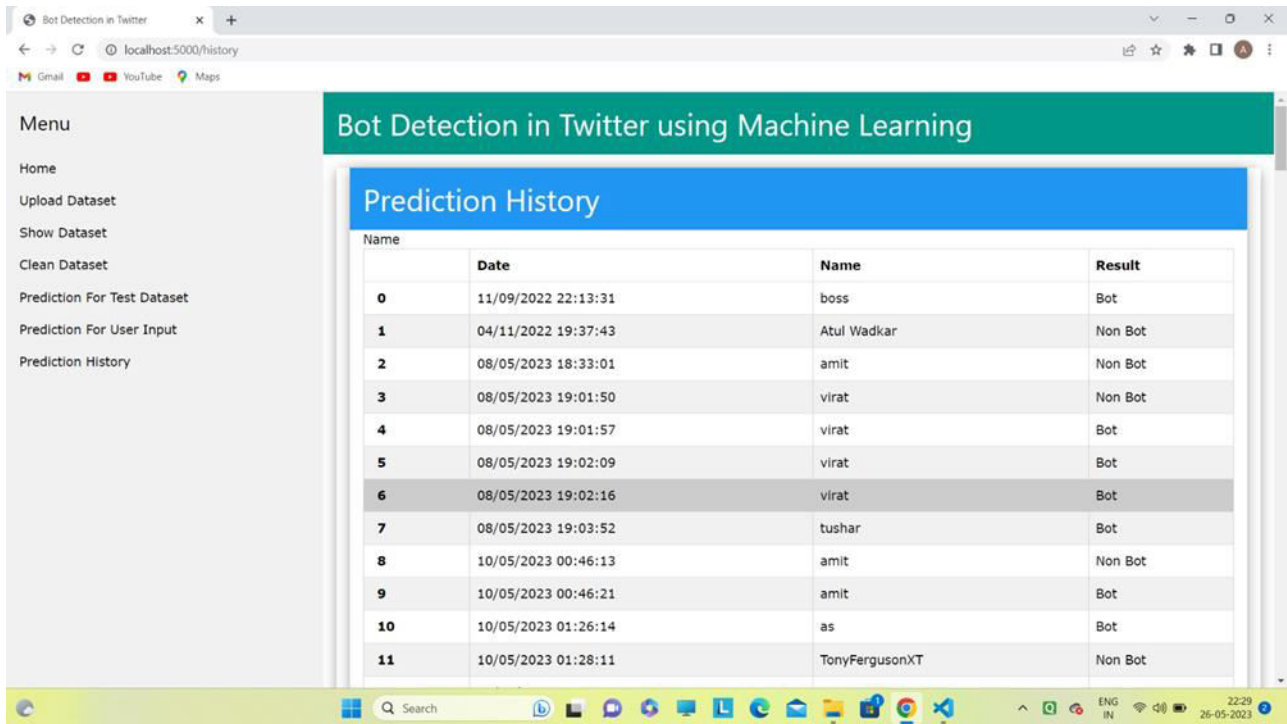
Fig4- Uploaded Dataset



The screenshot shows the same web application but with the URL localhost:5000/clean. The sidebar menu is identical. The main content area displays a table titled 'Cleaned Dataset' with the following data:

id	followers_count	friends_count	listedcount	favourites_count	verified	statuses_count	bot
0	2925	3	139	0	False	708	1
1	9	0	5	0	False	6	1
2	132	46	24	740	False	7346	1
3	54	1351	0	2	False	6	1
4	1300380	24248	7089	4184	True	8536	0
5	533	2	34	2	False	1056	1
6	343	1	8	0	False	5140	1
7	8217	2927	368	756	True	1396	0
8	307	4	14	5	False	2060	1
9	300	1	17	0	False	4092	1
10	159	0	8	0	False	904	1
11	416318	567	173	8652	True	4551	0

Fig5- Cleaned Train Data



Name	Date	Name	Result
0	11/09/2022 22:13:31	boss	Bot
1	04/11/2022 19:37:43	Atul Wadkar	Non Bot
2	08/05/2023 18:33:01	amit	Non Bot
3	08/05/2023 19:01:50	virat	Non Bot
4	08/05/2023 19:01:57	virat	Bot
5	08/05/2023 19:02:09	virat	Bot
6	08/05/2023 19:02:16	virat	Bot
7	08/05/2023 19:03:52	tushar	Bot
8	10/05/2023 00:46:13	amit	Non Bot
9	10/05/2023 00:46:21	amit	Bot
10	10/05/2023 01:26:14	as	Bot
11	10/05/2023 01:28:11	TonyFergusonXT	Non Bot

Fig5- Result data

VI. CONCLUSION

Bot detection is a major challenge in network security management. There are many methods and techniques used to track and capture bot activity. For new generation bots, this method may not be used at all, but this method can be effective in detecting bots on social platforms. We developed a new method based on machine learning and supervised learning

REFERENCES

- 1] M. AL-QURISHI, M. S. HOSSAIN, M. ALRUBAIAN, S. M. M. RAHMAN, AND A. ALAMRI, “LEVERAGING ANALYSIS OF USER BEHAVIOR TO IDENTIFY MALICIOUS ACTIVITIES IN LARGE- SCALE SOCIAL NETWORKS “ IEEE TRANS. IND. INFORMAT., VOL. 14, NO. 2, PP. 799813, FEB. 2018.
- 2] Z. ZHANG, R. SUN, X. WANG, AND C. ZHAO, “ A SITUATIONAL ANALYTIC METHOD FOR USER BEHAVIOR PATTERN IN MULTIMEDIA SOCIAL NETWORKS,” IEEE TRANS. BIG DATA, TO BE PUBLISHED. DOI: 10.1109/TBDDATA.2017.2657623.
- 3] Y. LIU, C.WANG, M. ZHANG, AND S. MA, “ USER BEHAVIOR MODELLING FOR BETTER WEB SEARCH RANKING,” FRONT. COMPUT. SCI., VOL. 11, NO. 6, PP. 923936, DEC. 2017.
- 4] G. Wang, X. Zhang, S. Tang, C. Wilson, H. Zheng, and B. Y. Zhao, “ Clickstream user behavior models,” ACM Trans. Web, vol. 11, no. 4, Jul. 2017, Art. no. 21



Impact Factor: 8.379



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 9940 572 462  6381 907 438  ijircce@gmail.com



www.ijircce.com

Scan to save the contact details