Comprehensive Strategies for Identifying X(Twitter) Bots

Arya Dhanesh Department of Computer Science and Engineering(CyberSecurity)
Vimal Jyothi Engineering College
Chemperi, Kannur
koroth.arya@gmail.com

Jyothika K
Department of Computer Science and Engineering(CyberSecurity)
Vimal Jyothi Engineering College
Chemperi, Kannur
jyothikavineesh@gmail.com

Malavika Jayaraj Department of Computer Science and Engineering(CyberSecurity)
Vimal Jyothi Engineering College
Chemperi, Kannur
malavikajayaraj4@gmail.com

Nevin Jose Antony
Department of Computer Science and Engineering(CyberSecurity)
Vimal Jyothi Engineering College
Chemperi, Kannur nevinjose@gmail.com

Anu Treesa George
Assistant Professor Department of Computer Science and Engineering(CyberSecurity) Vimal Jyothi Engineering College Chemperi, Kannur
anuvellackallil@vjec.ac.in

***Abstract*—Twitter is a social network where users interact via text-based posts called tweets, using hashtags, mentions, shortened URLs, and retweets. The growing user base and open nature of Twitter have made it a target for automated programs, known as bots, which can be both beneficial and malicious. This research focuses on detecting and classifying Twitter accounts as human, bot, or cyborg. Given Twitter's open nature, both helpful and harmful bots are prevalent, necessi- tating effective detection strategies. The study analyzes account behavior, content, and properties, introducing a classification system that integrates entropy analysis, spam detection, and account evaluation to improve user interaction transparency and platform security.**

I. INTRODUCTION

Twitter, a widely used platform for social network- ing and micro-blogging, allows users to share and interact through short, text-based posts. While its open and dynamic nature has fueled its popularity, it has also made the platform a hotspot for automated accounts, commonly known as bots. These bots can serve both positive and negative purposes—ranging from sharing valuable information to spreading spam and malicious content. Adding complexity, there are hybrid accounts, known as cyborgs, that blend human and bot activities. Social bots play a major role in spreading misinforma- tion on social media. They can significantly influence public opinion and even pose serious cyber security risks to society.

Given the impact of these automated accounts on user experience and platform integrity, it is crucial to effectively distinguish between humans, bots, and cyborgs. This paper explores this issue by analyzing user behavior, tweet content, and account features, and proposes a novel classification system designed to en- hance transparency and security in online interactions.

II. LITERATURE SURVEY

Khubaib Ahmed Qureshi et al.(2021) [5], The paper proposes a novel method for detecting fake news on Twitter, particularly related to COVID-19. It introduces a source-based approach that analyzes the network patterns and profile features of users who propagate news, both posters and re- tweeters, to classify tweets as true or false. Combining complex network measures with user profile features, the study uses machine learning and deep learning models to perform binary classification. The methodology is validated on a real- world COVID-19 dataset, showing that hybrid features outperform single-feature approaches, achieving an AUC score of 98% with models like CATBOOST and RNN. This approach also proves adaptable to political and entertainment domains.

The study ”Profiling Users and Bots in Twitter through Social Media Analysis” by Pastor-Galindo et al. (2022) [1], investigates social bot's behavior and influence on online networks using data from
2.8M Twitter users and 39M retweets during the 2019 Spanish election. Utilizing the BotoMeterLite algorithm, users were classified as Humans, Semi- Bots, or Bots. Analyses included statistical metrics, network structure, influence, temporal activity, and virality. Semi-bots were central to retweet networks, while bots boosted content visibility but had a limited impact on virality. The study highlights bots' role in content dissemination without disrupting network connectivity. Future work calls for advanced detection,

multi-platform analysis, and real-time monitoring, with applications in politics, cybersecurity, and social media management.

Oliver Beatson et al. study [2], published in Social Science Computer Review (2023), examines the chal- lenges of detecting bots on Twitter and their potential influence on public perception. The authors argue that real-time bot detection is highly difficult, making it un- realistic to expect everyday users to identify bots and assess their impact on online discourse accurately. To explore this issue, the study develops two distinct bot detection methods. The first follows a fixed, criteria- based approach that relies on widely recognized in- dicators of bot activity. The second method takes a more flexible and investigative approach, focusing on identifying bots that work together to influence conversations. Both methods are evaluated against a framework that defines key criteria for effective bot detection, with a particular focus on accuracy. The results reveal significant discrepancies between detec- tion methods, with different approaches often yielding different conclusions about whether an account is a bot. To verify the findings, the study cross-references results with an established bot detection service. How- ever, the authors emphasize that the only definitive way to confirm whether an account is a bot is through official action from Twitter, such as suspensions or public announcements. Ultimately, the study highlights the implications of unreliable bot detection, suggesting that it may undermine public trust in social media information and, by extension, the political process.

N. Chavoshi et al.[9], developed an advanced bot de- tection system that identifies coordinated user accounts on social media platforms like Twitter. Their approach is based on the observation that human users cannot maintain highly synchronized activity for extended periods, whereas bots often exhibit this pattern. Unlike traditional bot detection methods, which rely on su- pervised learning and require large amounts of labeled training data, their system detects bots through activity correlation without needing pre-labeled examples.To achieve their goal, the team came up with a ground- breaking technique known as lag-sensitive hashing. This innovative method allows them to swiftly group related user accounts into clusters in real time. They named their approach DeBot, and it's quite effective identifying thousands of bots each day with an im- pressive accuracy of 94%. By September 2016, DeBot had detected approximately 544,868 unique bots over a year. When comparing their approach to Twitter's sus- pension system and other user-based detection meth- ods, they discovered that some bots evade Twitter's bans and remain active for months. Alarmingly, their results showed that DeBot was detecting bots faster than Twitter was suspending them, highlighting the ongoing challenge of combating automated accounts on social media.

M. Heidari et al.[8], conducted research aimed at improving bot detection on social media using machine learning models. Their approach focuses on analyzing user profiles extracted from tweets to identify patterns that distinguish bots from human users. These profiles include inferred personal details such as age, gender, education, and personality, which are derived from users' online posts. The study makes three key con- tributions: 1. Enhancing bot detection with personal information – Traditional bot detection struggles be- cause bots can mimic human-like posting behaviors. However, this research leverages the similarity in per- sonal information across multiple posts to improve classification accuracy. The proposed model constructs user profiles based on personal characteristics and applies machine learning techniques to detect bots more effectively. 2. Introducing a new public dataset – The researchers compiled a dataset containing over 6,900 Twitter user profiles, extracted from the Cresci 2017 dataset, to support future studies in bot detection.
3. Applying advanced NLP techniques – This study is the first to use ELMO (Embeddings from Language Models), a deeply contextualized word embedding model, for detecting bots on social media. This ap- proach enhances the ability to analyze text patterns and improve classification accuracy. By combining user profiling with advanced NLP models, their method demonstrates high prediction accuracy and provides a more robust approach to identifying social bots in real-world scenarios.

Zi Chu et al.[3], conducted a large-scale analysis of over 500,000 social media accounts to examine the differences inbehaviourr between humans, bots, and cyborgs. Their study focused on key aspects such as tweeting patterns, content, and account characteris- tics to develop a more effective classification system. Based on their findings, they proposed a four-part clas- sification model designed to identify whether an ac- count is operated by a human, bot, or cyborg. The sys- tem includes: 1. Entropy-Based Analysis:– Measures the randomness in user activity to detect automated behavior.2. Spam Detection Component:– Identifies accounts that engage in spam-like posting.3. Account Properties Evaluation:– Examines profile details to distinguish between real users and bots.4. Decision- Making Module:– Integrates data from the previous components to classify an unknown account. Their experimental results demonstrated that this system ef- fectively differentiates between human and automated accounts, offering a reliable method for bot detection on social media platforms.

Fang Zhou et al.[4], introduced a method for as- sessing user credibility based on their past reviews. Their approach evaluates various aspects of a re- view to generate a numerical credibility score for the reviewer. This score can then be used to rank products more effectively, helping users make better- informed decisions. To test the effectiveness of their method, they conducted experiments on social book search databases, which showed that incorporating user credibility significantly improved the accuracy of prod- uct recommendations. This approach provides a more reliable way to filter reviews, ensuring that trustworthy opinions have a greater influence on recommendation systems.

Sina Mahdipour Saravani et al.[7], highlight how advancements in Natural Language Processing (NLP) have made bots more sophisticated and harder to de- tect. The widespread availability of easily deployable bots has increased the risk of malicious activities on social media, making bot detection a critical challenge. While much of the existing research focuses on iden- tifying bot accounts, many detection methods rely on metadata that is not always accessible. Furthermore, these methods struggle to identify cyborg accounts, which are either human-assisted bots or bot-assisted humans. To address these challenges, the researchers focus on detecting bots based on the textual content of their social media posts, which they refer to as fake posts. They emphasize that deep learning-based NLP techniques are among the most effective approaches for identifying deceptive content. Their study intro- duces an end-to-end neural network model designed to detect deepfake text on a real-world Twitter dataset. The results show that their model improves classifica- tion accuracy by 2% compared to previous methods. Beyond improving detection accuracy, this content- based approach has the potential to identify fake posts in real-time, preventing misinformation from spreading. By detecting deceptive content before it gains traction, this method could play a crucial role in mitigating the harmful effects of fake news on social media.

Randall Wald et al.[6], They conducted a detailed study to analyze how Twitter users interact with bots, utilizing a dataset of 610 users who had been contacted by automated accounts. This research aimed to gain deeper insights into user behavior and engagement patterns with social bots. Their research aimed to de- termine which user characteristics were most effective in predicting whether someone would engage with a bot either by replying or following it. To achieve this, They evaluated six different machine learning classifiers to create predictive models. They compared models using all available user features versus those using only the most relevant features. The findings revealed that a user's Klout score, friend count, and follower count were the strongest indicators of whether they would interact with a bot. Among the classifiers tested, The Random Forest algorithm Achieved out- standing performance. when paired with an effective feature-ranking method. However, they also noted that using a poor feature ranking technique could reduce performance, sometimes making it worse than using no ranking at all. Overall, their research provides valuable insights into the characteristics that make users more susceptible to engaging with social bots. This understanding could help in developing strategies to identify and mitigate the influence of bots on social media platforms.

J. Rodriguez-Ruiz et al.[10], examine the pres- ence of automated Twitter accounts, estimating that around 48 million accounts approximately 15% of the platform's total are managed by bots. While some bots serve beneficial purposes, such as sharing news updates, academic research, or providing emergency assistance, others are used for harmful activities, in- cluding spreading malware and manipulating public opinion. Although existing automated bot detection methods are in place, they primarily rely on identifying patterns based on known bot accounts. However, as bot creators develop more sophisticated evasion tech- niques, these detection methods become less effective. To address this challenge, the researchers propose us- ing one-class classification as an alternative approach. This technique improves bot detection by requiring only examples of legitimate accounts, allowing it to identify previously unseen bot behaviors. Experimental results demonstrate that this method can consistently detect various types of bots with high accuracy, achiev- ing an AUC score above 0.89. This suggests that one- class classification could be a valuable tool in enhanc- ing Twitter's ability to identify and combat emerging bot threats without relying on prior knowledge of their characteristics.

TABLE I
Comparison Table

Reference	Description	Advantages	Disadvantages
[1]	Profiling users and bots in Twitter using BotometerLite.	• Better Identification of Bots – By examining user behavior and interaction patterns, this approach improves the detection of automated accounts that spread misinformation or spam.	• Adaptable Bots Can Evade Detection – As bots become more advanced and mimic human activity more effectively, it becomes harder to accurately distinguish them from real users.
[2]	Evaluating Different Methods for Detecting Bots on Twitter: How Effective Are They?	• Evaluates Multiple Detection Methods – The study provides a comprehensive comparison of different bot detection techniques, helping identify the most effective approach for detecting automated accounts on Twitter.	• Challenge of Evolving Bots – As bots continuously improve and adapt, detection methods may become less effective over time, requiring constant updates and refinements.
[3]	Detecting automation of Twitter accounts.	• Distinguishes Between Different Account Types – The study goes beyond basic bot detection by classifying accounts as human, bot, or cyborg, providing a more detailed understanding of automated behavior on Twitter.	• Difficulty in Identifying Cyborg Accounts – Since cyborgs exhibit both human and automated behaviors, accurately distinguishing them from real users or bots can be challenging.
[4]	An assessing method for social network user credibility.	• Improves Trust Assessment – By providing a systematic way to evaluate user credibility, this method helps identify reliable sources and reduce the spread of misinformation on social networks.	• Potential Bias in Scoring – The accuracy of credibility calculations depends on the chosen criteria, which may introduce bias and unfairly label certain users as less credible.
[5]	Classifying COVID-19 Fake News on Twitter Using Complex Networks and Source-Based Analysis.	• Enhanced Fake News Detection – By combining network analysis with user profile features, the approach improves the accuracy of identifying misinformation on Twitter.	• Limited Applicability to Other Platforms – Since the method relies on Twitter-specific data, its effectiveness may be reduced when applied to other social media networks with different structures.
[6]	Predicting susceptibility to social bots on Twitter.	• Detecting users who are vulnerable to social bots can help reduce the spread of misinformation by enabling targeted awareness campaigns and improved content moderation. This helps create a more reliable and trustworthy online environment.	• Analyzing user behavior to predict susceptibility may raise privacy concerns, as it involves tracking online interactions and personal data, potentially leading to ethical issues related to surveillance and data security.
[7]	Detecting Social Media Bots with AI: Using Deepfake Text Analysis for Identification.	• Detecting social media bots through deepfake text analysis improves accuracy by identifying artificial language patterns that may go unnoticed by traditional methods. This helps reduce misinformation and enhances the authenticity of online interactions	• There is a risk of mistakenly flagging real users as bots, especially those with unique writing styles or non-native language use. This could lead to unfair restrictions and limit genuine user engagement.
[8]	Using Deep Contextualized Word Embeddings to Analyze Online Profiles and Detect Social Bots on Twitter.	• Using deep contextualized word embeddings allows for a more precise analysis of language patterns, helping to differentiate between human users and bots based on context and word relationships. This enhances the accuracy of social bot detection on Twitter.	• This method demands substantial computational power and large amounts of data for effective training, making it difficult to implement in real time and less accessible for smaller organizations with limited resources.
[9]	Debot: Twitter bot detection via warped correlation.	• Improved Bot Detection Accuracy – The use of warped correlation enhances the ability to detect bots by capturing subtle behavioral patterns that traditional methods might miss.	• Computational Complexity – The method may require significant processing power, making it challenging to implement efficiently on large-scale Twitter datasets.
[10]	Using a One-Class Classification Method to Detect Bots on Twitter.	• Effective for Imbalanced Data – Since bot detection often involves a scarcity of labeled bot accounts, the one-class classification approach works well by learning patterns from genuine users and identifying deviations.	• Higher False Positive Rate – This method may mistakenly classify some legitimate users as bots, as it primarily focuses on detecting anomalies rather than learning from both human and bot behaviors.

III. CONCLUSION

This study emphasizes the importance of distin- guishing between human, bot, and cyborg accounts on Twitter to tackle the challenges posed by automa- tion on the platform. By examining posting patterns, content, and account characteristics, the research in- troduces a comprehensive classification system. Com- bining tools like entropy analysis, spam detection, and account evaluation, the system demonstrates its ability to accurately identify account types. This approach contributes to safer and more transparent user inter- actions on social media, paving the way for further improvements in automated detection and platform security.

IV. REFERENCES

[1] Javier Pastor-Galindo "Profiling users and bots in Twitter through social media analysis" (2022).
[2] Oliver Beatson, Rachel Gibson, Marta Cantijoch Cunill, and Mark Elliot. 2021. "Automation on Twitter: Measuring the Effectiveness of Approaches to Bot Detection." Social Science Computer Review (2021)
[3] Zi Chu; Steven Gianvecchio; Haining Wang; Sushil Jajodia"'Detecting automation of Twitter accounts: Are you a human, bot, or cyborg?"" (2012).
[4] Fang Zhou;Jianlin Jin;Xiaojiang Du;Bowen Zhang;Xucheng Yin "A calculation method for social network user credibil- ity".(2017).
[5] Khubaib Ahmed Qureshi;Rauf Ahmed Shams Malick;Muhammad Sabih;Hocine Cherifi; " Complex Network and Source Inspired COVID-19 Fake News Classification on Twitter" unpublished. (2021).
[6] Randall Wald; Taghi M. Khoshgoftaar; Amri Napolitano; Chris Sumner, "Predicting susceptibility to social bots on twitter"
.(2013).
[7] Sina Mahdipour Saravani, Indrajit Ray, and Indrakshi Ray "Automated Identification of Social Media Bots Using Deepfake Text Detection".(2021)
[8] M Heidari, JH Jones, O Uzuner; "Deep contextualized word embedding for text-based online user profiling to detect social bots on Twitter".(2020).
[9] N Chavoshi, H Hamooni, A Mueen; "Debot: Twitter bot detection via warped correlation." (2016).
[10] J Rodr´ıguez-Ruiz; JI Mata-Sa´nchez; R Monroy; O Loyola- Gonza´lez; A Lo´pez-Cuevas; "A one-class classification ap- proach for bot detection on Twitter";(2020).