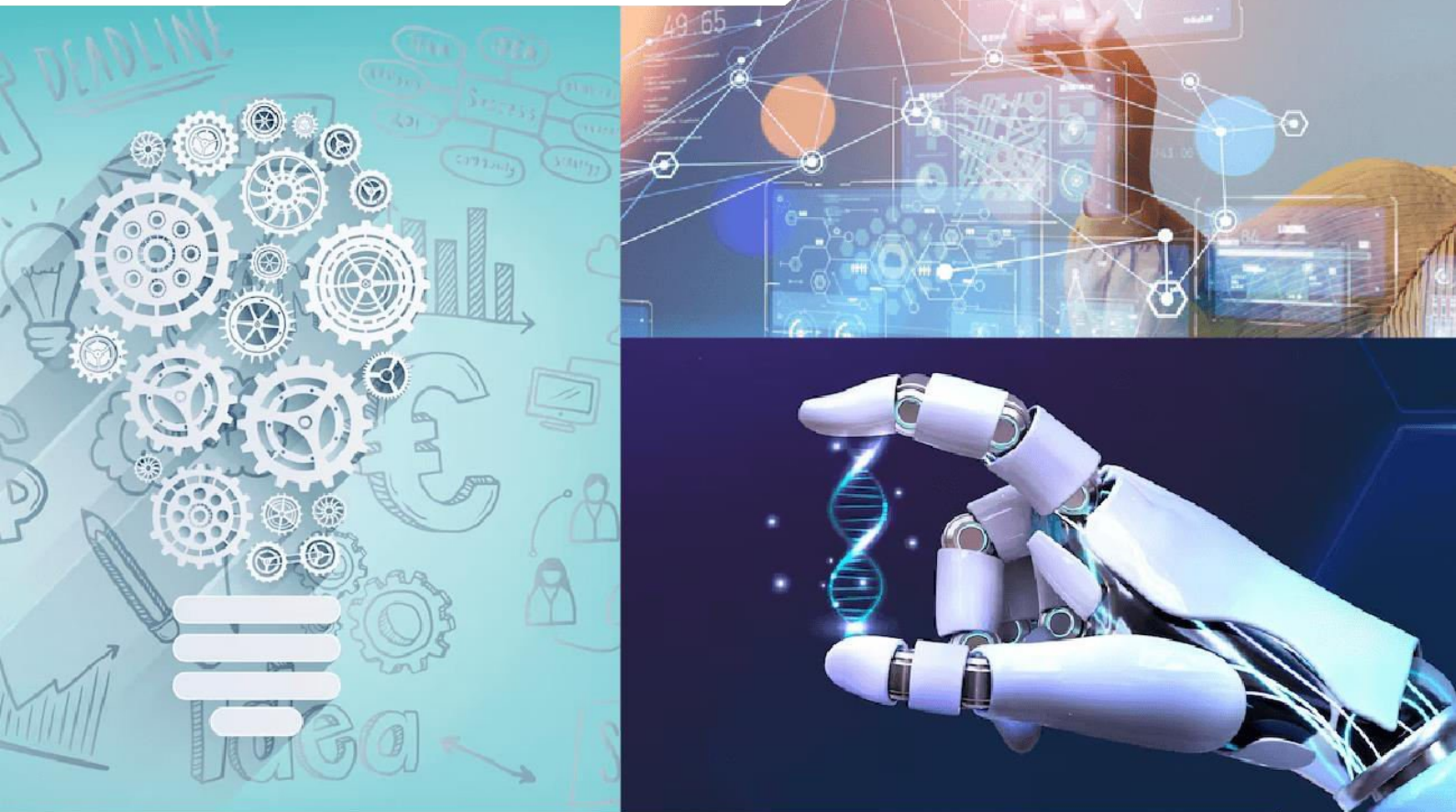




International Journal of Advanced Research in Education and Technology (IJARETY)

Volume 13, Issue 2, March - April 2026

Impact Factor: 8.152



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



Detecting Social Bots using Graph Metrics and User Profile Features

Nuha Aquil

Master of Computer Applications, CMR Institute of Technology, Bengaluru, India

ABSTRACT: The increasing influence of social media platforms has made the detection of automated bot accounts crucial to maintaining the authenticity of online interactions. This study proposes a machine learning-based framework that integrates user metadata with graph-based structural features derived from a user-user mention network. A mention graph was built from user interactions, and features such as in-degree, out-degree, clustering coefficient, and community membership were extracted. These features were combined with metadata to train Logistic Regression and XGBoost models. Results showed that XGBoost achieved 94% accuracy and 0.969 ROC-AUC, outperforming Logistic Regression. The findings highlight that combining metadata with graph-based features enhances bot detection performance.

KEYWORDS: Social Bot Detection, Graph-Based Features, Machine Learning, Community Detection, XGBoost

I. INTRODUCTION

The rapid growth of social media platforms such as Twitter (now X), Facebook, and Instagram has transformed the nature of online communication, information dissemination, and public discourse. These platforms facilitate global interaction by allowing users to exchange ideas, participate in discussions, and engage in civic and political movements. However, this openness has also enabled the rise of malicious entities, particularly automated bot accounts. [4] Social bots are algorithmically controlled profiles that imitate human behaviour to spread misinformation, manipulate trends, or influence public opinion [4]. [7] Their increasing sophistication, characterized by realistic activity patterns and content generation, makes them difficult to distinguish from legitimate users, posing serious challenges to the integrity of online ecosystems [7].

[9] Traditional bot detection methods have primarily focused on user profile and content-based attributes such as account creation date, follower-following ratio, tweet frequency, and linguistic cues [9]. While these techniques have achieved moderate success, they often fail to capture the underlying relational dynamics between users. [12] Bots frequently operate in coordinated groups, forming dense clusters that amplify certain narratives or artificially boost engagement metrics. As a result, approaches relying solely on metadata or textual information overlook the network-level dependencies that characterize coordinated behaviour [12]. Recent studies have emphasized that incorporating network-based features, such as degree centrality, clustering coefficients, and community structure, can substantially enhance detection performance by revealing hidden interaction patterns and group-level coordination.

The motivation for this research arises from the limitations of existing metadata-driven models and the growing need for holistic frameworks that integrate both profile information and structural network characteristics. This study aims to develop a machine learning-based bot detection framework that combines user metadata with graph-derived features extracted from a user-user mention network. By integrating structural metrics such as in-degree, out-degree, clustering coefficient, and community membership - identified using the Greedy modularity algorithm - this approach seeks to capture latent behavioural patterns that are otherwise undetectable using profile data alone.

The contributions of this work are twofold. First, it presents a scalable and interpretable framework that fuses metadata with graph-based structural features for improved social media bot detection. Second, it empirically demonstrates that including community-level information significantly enhances classification accuracy, offering deeper insights into user coordination and influence dynamics. The proposed system utilizes Logistic Regression and XGBoost models for classification, with experimental results showing that XGBoost achieves superior performance.

The remainder of this paper is organized as follows. Section 2 reviews related work on social bot detection and graph-based methods. Section 3 describes the proposed methodology, including data preprocessing, network construction, and

DOI:10.15680/IJARETY.2026.1302066

feature extraction. Section 4 presents the experimental setup and results, while Section 5 discusses key findings and implications. Finally, Section 6 concludes the study and outlines future research directions.

II. LITERATURE REVIEW

2.1 Machine Learning-Based Social Bot Detection

Early research in social bot detection primarily focused on applying machine learning algorithms using user profile and activity-based features. These approaches rely on attributes such as follower-following ratio, account age, tweet frequency, and content characteristics to classify users as bots or humans. Sethurajan and N.K. demonstrated that supervised learning models can effectively detect bots using carefully engineered metadata features [1]. Similarly, Ng et al. emphasized that the reliability of classification improves with sufficient temporal data, highlighting the importance of activity history in bot detection [2]. While these methods achieve reasonable accuracy, they are limited in capturing complex interaction patterns, as they treat users as independent entities without considering network relationships.

2.2 Deep Learning and Hybrid Detection Approaches

To improve detection performance, recent studies have incorporated deep learning and hybrid models that combine multiple feature types. Rohini et al. proposed a hybrid architecture integrating fuzzy systems, gradient boosting, and Bi-LSTM networks, achieving high accuracy in bot classification tasks. Cai et al. and Nasser et al. explored multimodal approaches that integrate textual, visual, and contextual features using deep learning techniques such as Transformers and RNNs [3]. Additionally, models like TweezBot leverage AI-driven architectures to detect bots in large-scale social networks [6]. Although these approaches enhance detection accuracy, they often suffer from high computational complexity and reduced interpretability, making them less suitable for real-time deployment.

2.3 Graph-Based and Network-Oriented Detection Methods

Recent research has shifted towards incorporating graph-based features to capture relational and structural patterns among users. Bots often operate in coordinated groups, forming dense clusters within social networks. Dehghan et al. demonstrated that node embeddings and structural graph features significantly improve classification performance [8]. Lakshmi et al. highlighted the effectiveness of node embedding techniques such as Node2Vec and Struc2Vec in capturing network dynamics [11]. Similarly, the DARPA Twitter Bot Challenge revealed that large-scale bot activity can be better understood through network-level analysis [12]. These studies show that features like degree centrality, clustering coefficient, and community structure provide valuable insights into coordinated bot behavior that cannot be captured using metadata alone.

2.4 Feature Engineering and Hybrid Metadata-Graph Models

Combining metadata with graph-based features has emerged as a promising direction in bot detection research. Pastor-Galindo et al. demonstrated that integrating behavioral and profile-based features improves classification accuracy [14]. Lopez-Joya et al. emphasized the importance of feature selection strategies in enhancing model reliability [10]. Recent studies also highlight that community detection plays a crucial role in identifying coordinated bot clusters, as bots tend to interact within specific groups. However, while hybrid models improve performance, challenges remain in balancing feature complexity, scalability, and interpretability.

Summary synthesis:

The literature reveals a clear shift toward integrating diverse feature types to improve social bot detection. Traditional machine learning approaches effectively utilize metadata and activity-based features for classification but often fail to capture interaction-level dependencies [1], [2], [9]. Deep learning and hybrid models enhance detection accuracy by incorporating multimodal data and complex feature representations, though they introduce challenges related to computational cost and interpretability [3], [6]. More recent studies emphasize graph-based techniques, demonstrating that structural features such as degree centrality, clustering coefficient, and community structure provide deeper insights into coordinated bot behavior [8], [11], [12]. Hybrid approaches that combine metadata with graph-based features further improve classification performance by capturing both individual and relational characteristics of users [10], [14]. However, despite these advancements, most existing systems rely on static datasets, lack scalability for real-time detection, and struggle with cross-platform generalization [17]. Collectively, these limitations highlight a critical gap in developing scalable, interpretable, and real-time capable frameworks that effectively integrate metadata and network-based features to detect evolving bot behavior.

DOI:10.15680/IJARETY.2026.1302066**III. REVIEW BASED METHODOLOGY**

This section presents a review-oriented approach to examine existing techniques for social bot detection and to define a structured methodology for building an effective detection framework. The analysis focuses on understanding how different approaches - ranging from traditional machine learning models to graph-based techniques - contribute to identifying automated accounts. The goal is to derive a unified methodology that combines multiple feature representations to improve detection performance and robustness.

3.1 Study Selection and Evaluation Criteria

The methodology relies on a comparative synthesis of machine learning-based bot detection models, graph-based approaches, and hybrid frameworks integrating multiple feature types. The selection process prioritized studies that utilizes user metadata such as follower count, activity frequency, and profile attributes for classification, incorporates graph-based structural features derived from user interaction networks and demonstrates scalable workflows applicable to real-world social media datasets.

Prominent studies, including metadata-driven models and graph-based approaches using node embeddings and community detection, provide a comparative baseline for evaluating how different systems extract, combine, and utilize features for bot detection [1]–[5], [8], [11]. These works establish the foundation for analyzing the effectiveness of individual and combined feature sets in identifying automated accounts.

Further research into hybrid models provides additional insight into integrating structural and behavioral features, particularly in capturing coordinated bot activity through network interactions [10], [14]. Such categorization lays the methodological groundwork for feature-level integration, emphasizing the importance of combining independent and relational data representations.

3.2 Analysis of Feature Utilization and Model Workflows

Existing studies demonstrate that bot detection systems often focus on isolated feature categories, leading to variations in performance and generalizability. Metadata-based models effectively capture individual user characteristics but fail to identify coordinated behavior across users [1], [9]. In contrast, graph-based approaches highlight relational patterns, such as clustering and community formation, which are indicative of bot networks [8], [12].

Hybrid frameworks attempt to bridge this gap by integrating metadata with structural features. These models demonstrate improved classification accuracy by capturing both user-level and network-level information. However, a key limitation across most approaches is the lack of balanced feature contribution, where metadata often dominates model performance while graph-based features provide only incremental improvements.

Additionally, many detection systems rely on static datasets, limiting their ability to adapt to evolving bot behavior. This highlights the need for a structured workflow that ensures effective feature extraction, integration, and evaluation within a consistent detection pipeline.

3.3 Network-Centric Insights for Behavioral Pattern Discovery

Recent advancements emphasize the role of graph-based features in uncovering hidden behavioral patterns in social networks. Structural metrics such as in-degree, out-degree, clustering coefficient, and community membership enable the identification of coordinated activity among users.

Community detection algorithms, such as Greedy modularity, group users based on interaction patterns, revealing clusters that may correspond to bot networks. These techniques introduce key methodological components such as extraction of structural metrics representing user connectivity and influence, integration of graph features with metadata for enhanced classification, identification of interaction-based communities within user networks. Despite these advantages, graph-based methods alone are insufficient for accurate detection, as they may not capture individual user characteristics. Therefore, combining graph-based insights with metadata ensures a more comprehensive representation of user behavior.

3.4 Consolidated Observations and Design Implications

Existing literature highlights two primary methodological approaches: metadata-driven classification and graph-based structural analysis [1]–[5], [8], [11]. While both approaches contribute to bot detection, neither independently provides

DOI:10.15680/IJARETY.2026.1302066

a complete solution. Metadata-based models lack relational context, whereas graph-based methods may overlook individual user attributes.

Therefore, a review-based synthesis defines an integrated methodology where metadata captures individual user behavior and profile characteristics, graph-based features represent interaction patterns and community structures and machine learning models combine both feature types for classification and validation.

This integrated framework forms the research foundation for a scalable and effective bot detection system capable of identifying both individual and coordinated bot activity, thereby addressing the limitations of existing single-feature approaches.

IV. RESULT AND OUTCOME

The dataset used in this study was obtained from a labeled Twitter dataset consisting of user metadata (kaggle.com/datasets/charvijain27/training-data-2-csv-utfcsv), along with tweet-level and network-level information. It includes features such as user verification status, followers and friends count, listed count, profile description, and screen name. Additionally, a user-user mention graph was constructed from tweet interactions to derive graph-based structural features. These include in-degree, out-degree and clustering coefficient.

Exploratory Data Analysis (EDA) revealed observable differences in graph metrics between bot and non-bot accounts. For instance, bots tend to have higher out-degrees and cluster together within specific communities. These patterns were visualized through boxplots and network structure diagrams, offering early insights into the relevance of graph-based signals.

Figure 2 illustrates a sampled portion of the user-user mention graph, showing how users interact and cluster within the network. This figure visualizes a sampled subgraph of the user-user mention network. It highlights the structure of user interactions and the potential for community separation.

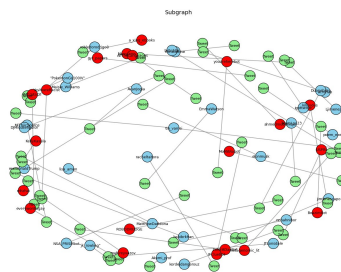


Fig. 2: User-User Mention Graph Subgraph

As shown in Figure 3, the Greedy modularity algorithm identifies communities within the mention graph based on interaction patterns. Nodes in this graph are coloured based on their community membership detected using the Greedy algorithm. The visual shows how users cluster into distinct communities based on interaction patterns.

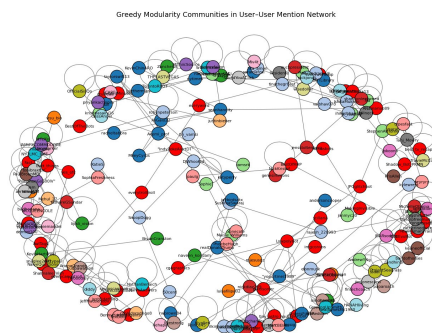


Fig. 3: Greedy Community Detection in Mention Graph

DOI:10.15680/IJARETY.2026.1302066

Figure 4 presents the feature importance scores from the XGBoost model, highlighting the dominance of metadata and the contribution of graph-based features. This bar chart presents the most influential features used by the XGBoost classifier. It shows that while metadata dominates, graph-based features like community ID also contribute to bot prediction.

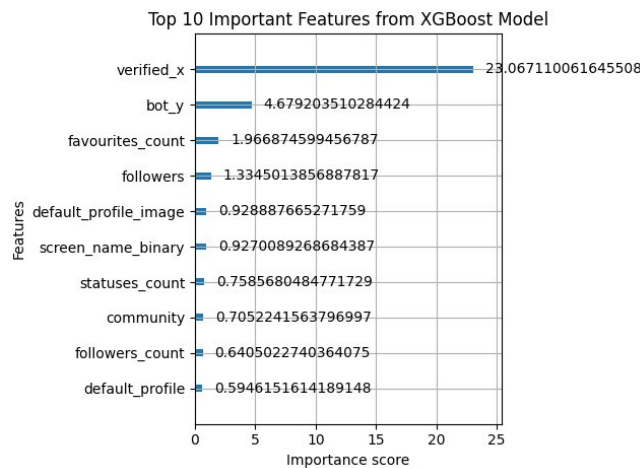


Fig. 4: Feature Importance from XGBoost Model

These tables compare the performance metrics of Logistic Regression and XGBoost models on the test set. While both models achieved strong classification results, XGBoost outperformed Logistic Regression in all key metrics, reaching a higher accuracy (94%) and F1-score (0.95) for bot detection.

Figure 5 shows the confusion matrices of both classifiers, providing insights into correct predictions and misclassifications. These matrices illustrate the number of true and false predictions made by each model. They provide insight into where each model succeeds or misclassifies bot and non-bot users.

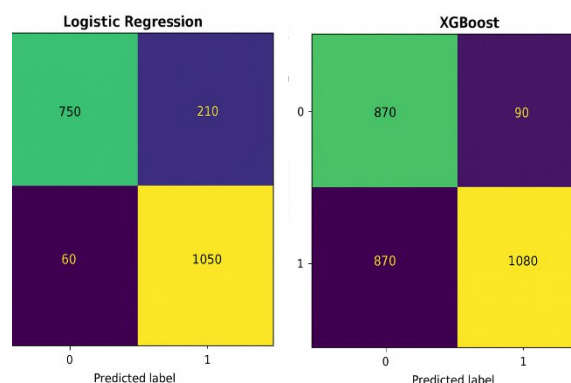


Fig. 5: Confusion Matrices for Both Models

V. CONCLUSION

This study presents an integrated machine learning-based approach for the detection of social media bots by leveraging both user metadata and structural graph features derived from a user-user mention network. By constructing a mention graph and applying a greedy modularity-based community detection algorithm, users were grouped into communities that provided additional structural insights beyond conventional profile attributes. Features such as in-degree, out-degree, clustering coefficient, and community membership were combined with profile data and used to train and evaluate Logistic Regression and XGBoost classifiers. Among the two models, XGBoost achieved superior performance with an accuracy of 94% and a ROC-AUC score of 0.969, outperforming the baseline logistic regression

DOI:10.15680/IJARETY.2026.1302066

model. While metadata-based features dominated in importance, the addition of graph-based features, particularly community labels, contributed to contextual differentiation. This validates the potential of incorporating basic network structure for improving classification in bot detection tasks. The study primarily focuses on static graph structures and metadata from a single dataset, which may not fully capture the temporal dynamics or platform-specific variations in bot behavior. Future work will focus on expanding the graph feature set, employing dynamic network analysis, and integrating explainability techniques like SHAP for deeper model interpretability.

REFERENCES

- [1] M. R. Sethurajan and N. K., "An adept approach to ascertain and elude probable social bots attacks on Twitter and Twitch employing machine learning approach," *MethodsX*, vol. 11, Art. no. 102430, 2023.
- [2] L. H. X. Ng, D. C. Robertson, and K. M. Carley, "Stabilizing a supervised bot detection algorithm: How much data is needed for consistent predictions?," *Online Social Networks and Media*, vol. 28, Art. no. 100198, 2022.
- [3] M. Nasser *et al.*, "A systematic review of multimodal fake news detection on social media using deep learning models," *Results in Engineering*, vol. 26, Art. no. 104752, 2025.
- [4] A. Ferrara, O. Varol, C. Davis, F. Menczer, and A. Flammini, "The rise of social bots," *Communications of the ACM*, vol. 59, no. 7, pp. 96–104, 2016.
- [5] A. Z. Kenyeres, *Social Media Bot Detection*, Master's thesis, Dept. Comput. Sci. and Electr. Eng., Luleå Univ. of Technology, Sweden, 2021.
- [6] R. Shukla, A. Sinha, and A. Chaudhary, "TweezBot: An AI-driven online media bot identification algorithm for Twitter social networks," *Electronics*, vol. 11, no. 5, p. 743, 2022.
- [7] O. Varol, E. Ferrara, C. A. Davis, F. Menczer, and A. Flammini, "Online human-bot interactions: Detection, estimation, and characterization," in *Proc. Int. AAAI Conf. Web and Social Media (ICWSM)*, vol. 11, no. 1, pp. 280–289, 2017.
- [8] A. Dehghan *et al.*, "Detecting bots in social networks using node and structural embeddings," *J. Big Data*, vol. 10, no. 119, 2023.
- [9] A. Bessi and E. Ferrara, "Social bots distort the 2016 U.S. Presidential election online discussion," *First Monday*, vol. 21, no. 11, 2016.
- [10] S. Lopez-Joya, J. A. Diaz-Garcia, M. D. Ruiz, and M. J. Martin-Bautista, "Exploring social bots: A feature-based approach to improve bot detection in social networks," *arXiv preprint arXiv:2411.06626*, 2024.
- [11] Z. Ellaky and F. Benabbou, "Political social media bot detection: Unveiling cutting-edge feature selection and engineering strategies in machine learning model development," *Scientific African*, vol. 25, Art. no. e02269, 2024.
- [12] A. Subrahmanian *et al.*, "The DARPA Twitter bot challenge," *Computer*, vol. 49, no. 6, pp. 38–46, 2016.
- [13] Q. Guyan, Y. Liu, J. Liu, and P. Zhang, "PEGNN: Peripheral-enhanced graph neural network for social bot detection," *Expert Systems With Applications*, vol. 278, Art. no. 127294, 2025.
- [14] J. Pastor-Galindo, F. Gómez Mármol, and G. Martínez Pérez, "Profiling users and bots in Twitter through social media analysis," *Information Sciences*, vol. 613, pp. 161–183, 2022.
- [15] A. V. S. Kumar, N. S. Kumar, and R. K. Devi, "An efficient hybrid Hopfield convolutional neural network for detecting spam bots in Twitter platform," *Int. J. Cognitive Computing in Engineering*, vol. 6, pp. 569–587, 2025.
- [16] A. Egli, T. Lynn, P. Rosati, and G. Sinclair, "Bad robot? The benevolent use of automated software and social bots by influencers in the #antivaxx discourse on Twitter," *Online Information Review*, vol. 49, no. 8, pp. 44–61, 2025.
- [17] A. Z. Kenyeres, "Social Media Bot Detection," M.S. thesis, Dept. Comput. Sci. and Electr. Eng., Luleå Univ. of Technology, Sweden, 2021.
- [18] A. Bhardwaj, S. Bharany, and S. Kim, "Fake social media news and distorted campaign detection framework using sentiment analysis and machine learning," *Heliyon*, vol. 10, Art. no. e36049, 2024.
- [19] V. Suarez-Lledo, E. Ortega-Martin, J. Carretero-Bravo, B. Ramos-Fiol, and J. Alvarez-Galvez, "Unraveling the use of disinformation hashtags by social bots during the COVID-19 pandemic: Social networks analysis," *JMIR Infodemiology*, vol. 5, Art. no. e50021, 2025.
- [20] J. Kim and M. C. Horning, "Exploring automated bot accounts and their impact in Twitter discussions on climate change," *Technology in Society*, vol. 75, Art. no. 102344, 2023.

International Journal of Advanced Research in Education and Technology

ISSN: 2394-2975

Impact Factor: 8.152