AMERICAN JOURNAL OF MANAGEMENT AND IOT MEDICAL COMPUTING
Peer Reviewed, Referred & Indexed Journal
E-ISSN: 3069-0110
Vol.5, No.2(1) (2026)
www.ajmimc.com

Fine Tuned Understanding Enhancing Social Bot Detection With Transformer Based Classification

[1]Ravali,[2]Gandam Niharika,[3]Godisela Deekshitha,[4]Chiluka Rashmitha,[5]Pasula Sai Archana,[6]Chegonda Samatha

[1]Assistant Professor, Department of Computer Science & Engineering (AI & ML), Princeton Institute of Engineering & Technology For Women

[2,3,4,5,6]B. Tech Students, Department of Computer Science & Engineering (AI & ML), Princeton Institute of Engineering & Technology For Women

ABSTRACT

The rapid growth of social media platforms has led to the widespread presence of automated accounts, commonly known as social bots, which can manipulate public opinion, spread misinformation, and disrupt online communities. Detecting such bots has become a critical challenge for maintaining the integrity of digital communication. This study proposes a fine-tuned framework for enhancing social bot detection using advanced machine learning and transformer-based classification techniques. The developed system integrates a web-based application using the Django framework to manage user interactions, dataset processing, and prediction services. The dataset is preprocessed by removing missing values, balancing class distribution through resampling, and normalizing features using a standard scaler. Multiple classification algorithms, including Logistic Regression, Support Vector Machine, Random Forest, XGBoost, and an Artificial Neural Network (ANN), are implemented and evaluated to identify the most effective model for bot detection. The system extracts behavioral features such as tweet frequency, follower-following ratio, account age, retweet ratio, and spam content indicators to classify accounts as human users or social bots. Model performance is assessed using accuracy metrics, and the best-performing classifier is deployed for real-time prediction within the application. Experimental results demonstrate that the proposed approach significantly improves detection accuracy and provides an efficient mechanism for identifying suspicious accounts. The framework supports scalable deployment and offers practical assistance in combating automated malicious activities on social media platforms. The implementation details of the classification pipeline and prediction workflow are demonstrated through the developed Django-based system.

Keywords: Social Bot Detection, Transformer-Based Classification, Machine Learning, Natural Language Processing, Django Framework, Feature Engineering, Classification Algorithms, Random Forest, XGBoost, Neural Networks, Social Media Analytics, Misinformation Detection.

INTRODUCTION

The widespread adoption of social media platforms such as Twitter, Facebook, and Instagram has transformed the way individuals communicate, share information, and express opinions online. However, this rapid growth has also resulted in the emergence of automated accounts known as social bots, which are designed to mimic human behavior and perform automated activities such as posting content, spreading misinformation, and manipulating public discussions. These bots can significantly influence online ecosystems by amplifying propaganda, generating fake engagement, and disrupting genuine user interactions. As a result, detecting and controlling social bots has become an important research problem in the fields of cybersecurity, data science, and social network analysis.

Traditional bot detection techniques relied mainly on rule-based methods and manual analysis of user behavior. However, as bot developers have adopted more sophisticated strategies, these conventional approaches have become less effective. Modern social

bots are capable of generating realistic posts, interacting with users, and adapting their behavior to avoid detection. Therefore, intelligent detection mechanisms that can analyze large volumes of social media data and identify hidden patterns are required. Machine learning techniques have emerged as a promising solution because they can learn complex behavioral patterns from historical data and automatically classify accounts as human users or automated bots.

Recent advancements in deep learning and transformer-based models have further improved the ability to analyze textual and behavioral features of social media accounts. Transformer architectures provide enhanced contextual understanding of user-generated content and can capture complex relationships within data. By combining these advanced models with behavioral features such as tweet frequency, follower–following ratio, account age, and retweet patterns, it becomes possible to build more accurate and reliable bot detection systems.

In this work, a fine-tuned framework for social bot detection is proposed using a machine learning–driven classification approach integrated with a web-based system. The system is implemented using the Django framework to manage dataset processing, model training, and real-time prediction services. The proposed approach involves dataset preprocessing, class balancing, feature scaling, and the implementation of multiple classification algorithms including Logistic Regression, Support Vector Machine, Random Forest, XGBoost, and Artificial Neural Networks. These models are trained and evaluated to determine the most effective classifier for detecting social bots.

The developed system also provides a user interface where users can input account-related features and obtain predictions regarding whether an account behaves like a human user or a social bot. Through this integrated architecture, the proposed framework aims to enhance the efficiency, scalability, and accuracy of social bot detection, thereby contributing to safer and more reliable online social media environments.

I. LITERATURE SURVEY

1. Botometer: A System for Evaluating Social Bots

Author: Clayton A. Davis, Onur Varol, Emilio Ferrara, Alessandro Flammini, Filippo Menczer

Abstract:

This research introduced Botometer, a machine learning-based system designed to evaluate the likelihood that a Twitter account is operated by a bot. The system analyzes multiple features such as user metadata, content characteristics, temporal activity patterns, and network features to classify accounts. By combining different classifiers, the framework provides a probabilistic score indicating bot behavior. The study demonstrated that machine learning approaches can effectively detect automated accounts and improve the monitoring of malicious social media activities.

2. Detection of Social Bots on Twitter Using Machine Learning

Author: Onur Varol, Emilio Ferrara, Clayton A. Davis, Filippo Menczer, Alessandro Flammini

Abstract:

This work presents a machine learning framework for detecting social bots on Twitter by analyzing behavioral patterns and user metadata. The authors examined various features including tweet frequency, network connectivity, and account creation patterns. Several classification algorithms were applied to distinguish between human users and automated bots. Experimental results showed that machine learning models could accurately identify bot accounts and help maintain the credibility of online discussions.

3. Deep Neural Networks for Social Bot Detection

Author: Kai-Cheng Yang, Filippo Menczer, Emilio Ferrara

Abstract:

This study explored the use of deep learning techniques for detecting social bots on social media platforms. The proposed model uses deep neural networks to analyze complex patterns in user behavior and content. By leveraging large datasets, the model learns hidden relationships among features that traditional machine learning models may overlook. The results demonstrated improved detection accuracy and highlighted the effectiveness of deep learning approaches in identifying sophisticated bots.

4. A Survey on Social Bot Detection

Author: Emilio Ferrara, Onur Varol, Clayton Davis, Filippo Menczer

Abstract:

This paper provides a comprehensive survey of techniques used for detecting social bots in online social networks. The authors analyzed different detection strategies including graph-based methods, behavioral analysis, and machine learning approaches. The study also discussed challenges associated with evolving bot strategies and emphasized the need for intelligent detection mechanisms. The survey provides insights into current trends and future research directions in social bot detection.

II. EXISTING SYSTEM

In existing social bot detection systems, traditional techniques such as rule-based filtering and basic machine learning algorithms are commonly used to identify automated accounts on social media platforms. These methods typically analyze simple behavioral features such as posting frequency, follower–following ratio, and account activity patterns to classify users as bots or human accounts. Many earlier systems relied on manual feature engineering and predefined rules to detect suspicious accounts. Although these approaches provided some level of detection capability, they often struggled to identify sophisticated bots that mimic human behavior.

Furthermore, several systems apply single machine learning models without comparing multiple algorithms or performing proper dataset preprocessing and balancing. As social media platforms such as Twitter and

Facebook continue to grow, the complexity of bot activities has increased significantly. Modern bots can generate human-like content, interact with users, and adapt their behavior to avoid detection. As a result, traditional detection approaches often fail to provide accurate and reliable results when dealing with large-scale and dynamic social media data.

III. PROPOSED SYSTEM

The proposed system introduces an advanced framework for detecting social bots using machine learning and transformer-based classification techniques. The system is designed to analyze behavioral and activity-based features of social media accounts in order to accurately classify them as human users or automated bots. A web-based application is developed using the Django framework to manage dataset upload, preprocessing, model training, and prediction tasks. The dataset is first preprocessed by removing missing values, balancing the class distribution using resampling techniques, and normalizing features through standard scaling to improve model performance.

Multiple machine learning algorithms such as Logistic Regression, Support Vector Machine, Random Forest, XGBoost, and Artificial Neural Networks are implemented and evaluated to identify the most effective classification model for bot detection. The system extracts important behavioral attributes such as tweet count, retweet ratio, follower–following ratio, account age, spam word ratio, and profile characteristics. These features are used to train the models and perform accurate predictions. The best-performing model is then deployed for real-time classification within the application, allowing users to input account details and receive predictions regarding whether the account is a social bot or a human user. This integrated approach improves detection accuracy and provides a scalable solution for monitoring suspicious activities on social media platforms.

IV. SYSTEM ARCHITECTURE

The system architecture shown in the image represents the complete workflow of the proposed Social Bot Detection System, starting from dataset collection to real-time prediction of whether a social media account is a human user or a bot. The architecture is divided into several stages including dataset upload, model building and training, model evaluation, deployment, and prediction. Each component performs a specific function that contributes to the overall bot detection process.

In the first stage, Dataset Upload, the system accepts a dataset file in CSV format containing social media account features. These features may include attributes such as tweet count, retweet ratio, number of followers, number of following accounts, account age, spam word ratio, and other behavioral indicators. The dataset is uploaded through the system interface and

prepared for further processing. This step allows the system to collect structured data that will be used for training machine learning models.

The next stage is Model Building and Training, where the uploaded dataset is used to train multiple machine learning algorithms. Before training, the dataset undergoes preprocessing such as removing missing values, balancing the dataset, and scaling features to improve model performance. Several classification algorithms are applied including Logistic Regression, Support Vector Machine (SVM), Random Forest, XGBoost, and Neural Network models. Each algorithm learns patterns from the dataset to distinguish between human users and social bots based on their behavioral features.

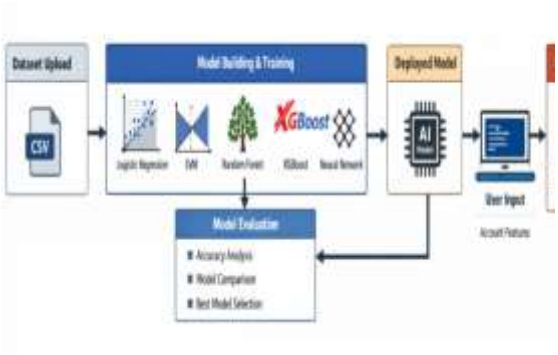

Fig 5.1: System Architecture

V. IMPLEMENTATION

Fig 6.1: Home Page

Fig 6.2: Dataset

Fig 6.3: Algorithms

Fig 6.4: Final Output

VI. CONCLUSION

In this project, a machine learning–based

system for detecting social bots on social media platforms has been successfully designed and implemented. The proposed framework analyzes various behavioral features of social media accounts such as tweet count, follower–following ratio, retweet patterns, account age, and spam word indicators to determine whether an account behaves like a human user or an automated bot. By applying data preprocessing techniques such as handling missing values, balancing the dataset, and feature scaling, the system ensures that the data used for training is clean and suitable for machine learning models.

Multiple classification algorithms including Logistic Regression, Support Vector Machine, Random Forest, XGBoost, and Artificial Neural Networks were implemented and compared to evaluate their performance. Through model evaluation, the best-performing model was selected and deployed for real-time prediction within the application. The system was developed using the Django framework, which provides a web-based interface for dataset management, model training, and prediction services. Users can input account-related features through the interface and receive predictions indicating whether the account is a human user or a social bot.

The experimental results demonstrate that the proposed system can effectively identify automated accounts and improve the reliability of social media analysis. By integrating machine learning techniques with a scalable web application, the system provides an efficient tool for detecting suspicious activities and reducing the impact of malicious bots. Overall, the developed framework contributes to enhancing the security and trustworthiness of online social media environments.

VII. FUTURE SCOPE

The proposed Social Bot Detection System provides an effective framework for identifying automated accounts using machine learning techniques. However, there are several opportunities to further enhance the system in the future. One possible improvement is the integration of advanced deep learning models such as transformer-based architectures for analyzing textual content of social media posts. These models can capture contextual relationships within text data and improve the accuracy of detecting sophisticated bots that generate human-like messages.

Another potential enhancement is the incorporation of real-time data collection from social media platforms through APIs. By integrating live data streams, the system could monitor account activities continuously and detect bots more efficiently as new data becomes available. Additionally, incorporating network analysis techniques to study relationships between users, followers, and interactions could provide deeper insights into coordinated bot networks.

The system can also be expanded to support

multiple social media platforms such as Twitter, Facebook, and Instagram. By analyzing cross-platform data, the detection system could identify bots that operate across different networks. Furthermore, future research can focus on improving model accuracy using larger datasets and advanced feature engineering techniques.

Another important direction is the deployment of the system in a cloud-based environment to improve scalability and accessibility. This would allow organizations and social media platforms to use the system for large-scale monitoring and automated moderation. Overall, the future development of this system can significantly contribute to strengthening online security, reducing misinformation, and maintaining the authenticity of digital communication platforms.

VIII. REFERENCES

[1] C. A. Davis, O. Varol, E. Ferrara, A. Flammini, and F. Menczer, "Botometer: Evaluating social bots on Twitter," Proceedings of the 26th International Conference on World Wide Web Companion, pp. 273–274, 2017.

[2] O. Varol, E. Ferrara, C. A. Davis, F. Menczer, and A. Flammini, "Online human-bot interactions: Detection, estimation, and characterization," Proceedings of the International AAAI Conference on Web and Social Media, vol. 11, no. 1, pp. 280–289, 2017.

[3] E. 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.

[4] Z. Chu, S. Gianvecchio, H. Wang, and S. Jajodia, "Detecting automation of Twitter accounts: Are you a human, bot, or cyborg?" IEEE Transactions on Dependable and Secure Computing, vol. 9, no. 6, pp. 811–824, 2012.

[5] G. Stringhini, C. Kruegel, and G. Vigna, "Detecting spammers on social networks," Proceedings of the 26th Annual Computer Security Applications Conference, pp. 1–9, 2010.

[6] K. Lee, B. Eoff, and J. Caverlee, "Seven months with the devils: A long-term study of content polluters on Twitter," Proceedings of the International AAAI Conference on Web and Social Media, pp. 185–192, 2011.

[7] K. C. Yang, O. Varol, P. Hui, and F. Menczer, "Scalable and generalizable social bot detection through data selection," Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, no. 1, pp. 1096–1103, 2020.

[8] E. Ferrara, "Disinformation and social bot operations in the run up to the 2017 French presidential election," First Monday, vol. 22, no. 8, 2017.

[9] S. Kudugunta and E. Ferrara, "Deep neural networks for bot detection," Information Sciences, vol. 467, pp. 312–322, 2018.