Fake Twitter Followers Detection using Machine Learning Approach

Mrs. A. S. Kolhe[1], Priti Mogal[2], Gaytri Barde[3], Bhavna Dhatrak[4], Gaytri Wagh[5]
Department of Information Technology[1,2,3,4,5]
Matoshri Aasarabai Polytechnic, Eklahare, Nashik, India[1,2,3,4,5]

Abstract*: Addressing the issue of fake user accounts on social media platforms like Twitter is indeed crucial, and employing artificial intelligence can be a significant step towards mitigating this problem. The proposed research utilizes publicly available information about Twitter users, including their activity patterns, profile details, and tweet content, to assess the authenticity of each account. By leveraging Twitter's API and other data retrieval methods, a machine learning model is developed to classify user accounts as either real or fake. Various machine learning algorithms are applied in this study, including logistic regression, long short-term memory (LSTM), K-means, and random forest, to evaluate the effectiveness of the proposed model. Among these algorithms, the random forest algorithm emerges as the most successful, achieving the highest accuracy score of 0.7557, a precision of 0.7277, and an F1 score of 0.7943.By utilizing machine learning techniques and analyzing diverse features of Twitter user accounts, the proposed model demonstrates promising results in detecting and identifying fake accounts. This research contributes to the ongoing efforts to combat issues such as spamming, trend manipulation, and follower inflation on social media platforms, thereby enhancing the overall integrity and trustworthiness of online interactions.*

Keywords: Fake, Twitter accounts, Social Network, Classification

I. INTRODUCTION

1.1 Overview

Social media has become an integral part of modern life, serving as a platform for communication, information sharing, and networking. With millions of users actively engaging on platforms like Twitter and Facebook daily, the impact of social media on society is undeniable. However, alongside its numerous benefits, social media also presents significant challenges, particularly concerning the proliferation of fake profiles and the spread of misinformation. The rise of fake accounts, whether created by humans or automated bots, poses various threats, including spamming, phishing, cyberbullying, and the dissemination of false information, which can lead to social unrest and conflicts.

In this context, identifying and combating fake profiles on social media platforms like Twitter has become imperative. With millions of daily tweets and interactions, distinguishing genuine accounts from fraudulent ones is a daunting task. This research project aims to address this challenge by leveraging machine learning techniques to analyze and detect fake Twitter identities. By examining existing methodologies and exploring novel approaches, the study seeks to contribute to the ongoing efforts to safeguard the integrity of online communities and protect users from the harmful consequences of fake profiles. Through a comprehensive analysis of user behavior patterns and technical strategies, this research endeavors to advance our understanding of fake profile detection and enhance the resilience of social media platforms against malicious actors.

Machine learning is indeed a captivating field that empowers computers to learn and make predictions without explicit programming. In the realm of social media, platforms serve as hubs for users to stay informed, connect with like-minded individuals, and access a myriad of services spanning from communication and networking to commerce and content sharing. Twitter, with its vast user base and constant stream of tweets, stands as a prominent example in the social media landscape.

DOI: 10.48175/568

However, alongside the benefits of social networking, there exists a concerning issue of fake accounts proliferating across platforms. Hackers and malicious actors exploit personal data obtained through various means, creating false identities to disseminate fake news, links, and images. These fraudulent accounts pose a significant threat to users, who may unknowingly accept their requests or engage with their content, falling victim to scams, misinformation, or privacy breaches.

As the prevalence of fake accounts continues to be a pressing concern, it's essential for social media users, especially those on platforms like Twitter, to remain vigilant and informed about the existence of fraudulent profiles. Awareness and understanding of the risks associated with fake accounts are crucial in mitigating their impact and maintaining a safe and secure online environment for all users.

1.2 Motivation

The pervasive presence of fake accounts on social media platforms like Twitter underscores the pressing need for effective detection mechanisms to combat misinformation and safeguard user trust. The exponential growth of social media usage has led to a proliferation of false identities, which not only spread malicious content but also pose risks such as identity theft and data breaches. The detrimental impact of fake accounts extends beyond individual users, affecting the integrity of online communities and exacerbating societal tensions fueled by the dissemination of false information. This escalating threat calls for innovative approaches grounded in machine learning and data reduction techniques to enhance the ability to distinguish between genuine and fraudulent profiles, thereby bolstering the resilience of social media ecosystems.

Moreover, the consequences of unchecked fake accounts extend beyond mere inconvenience to potentially severe societal consequences. From influencing public opinion and political discourse to perpetrating scams and cybercrimes, fake accounts wield significant power to manipulate and disrupt online interactions. By developing a robust detection model that leverages the wealth of data available on social media platforms, we aim to empower users and platform administrators with the tools necessary to identify and mitigate the risks posed by fake accounts. Through our research, we aspire to contribute to the ongoing efforts to promote transparency, trustworthiness, and accountability in the digital realm, ultimately fostering a safer and more authentic online environment for all users.

1.3 Problem Definition and Objectives

The pervasive presence of fake accounts on social media platforms, particularly Twitter, poses a significant challenge to the integrity and trustworthiness of online interactions. These fraudulent accounts, created with malicious intent, disseminate misinformation, perpetrate scams, and undermine the credibility of genuine users and content. The rapid proliferation of fake accounts not only erodes user trust but also engenders societal discord and exacerbates tensions fueled by the spread of false information. Detecting and mitigating the impact of fake accounts is essential to preserving the authenticity and reliability of social media platforms and safeguarding users from the risks associated with fraudulent activities.

- Develop a robust detection model: Our primary objective is to design and implement a machine learning-based detection model capable of accurately identifying fake Twitter accounts. By leveraging advanced algorithms and data reduction techniques, we aim to enhance the efficiency and effectiveness of fake account detection, thereby bolstering the resilience of social media platforms against malicious actors.
- Mitigate the spread of misinformation: We seek to mitigate the dissemination of misinformation and harmful content propagated by fake accounts through proactive identification and timely intervention. By identifying and flagging fraudulent profiles, our model aims to curb the spread of false information, protect user trust, and promote transparency and authenticity in online interactions.
- Empower users and platform administrators: Through our research, we aim to empower users and platform administrators with the tools and insights necessary to identify and address the presence of fake accounts effectively. By providing actionable intelligence and detection mechanisms, we strive to foster a safer and more trustworthy online environment for all users.

- Contribute to the advancement of cyber security: By addressing the pervasive problem of fake accounts on social media platforms, we aim to contribute to the broader discourse on cyber security and digital resilience. Our research endeavors to advance understanding and methodologies for detecting and mitigating emerging threats in the digital landscape, ultimately enhancing the security and integrity of online ecosystems.

1.4. Project Scope and Limitations

The scope of this project encompasses the development and implementation of a machine learning-based model for detecting fake Twitter accounts. The project will focus on leveraging advanced algorithms and data reduction techniques to analyze various features and patterns associated with user profiles and activities on the platform. Additionally, the model will aim to provide actionable insights to users and platform administrators to effectively identify and mitigate the proliferation of fake accounts, thereby enhancing the integrity and trustworthiness of online interactions on Twitter.

Limitations as follows:

- Data availability and quality: The effectiveness of the detection model may be limited by the availability and quality of data, including the presence of labeled datasets for training and validation. Limited access to comprehensive and accurately labeled data may hinder the model's ability to generalize across different types of fake accounts and variations in fraudulent behavior.
- Algorithmic biases and false positives: Despite leveraging advanced machine learning algorithms, the detection model may still be susceptible to biases and false positives. Inaccurate classification of legitimate accounts as fake or vice versa could undermine user trust and lead to unintended consequences, highlighting the need for ongoing refinement and validation of the model's performance.
- Evolution of fraudulent tactics: The dynamic nature of fraudulent activities and tactics employed by malicious actors on social media platforms presents a challenge to the long-term efficacy of the detection model. As perpetrators adapt and evolve their strategies to evade detection, the model may require continuous updates and enhancements to effectively counter emerging threats and maintain its effectiveness over time.

II. LITERATURE REVIEW

Paper 1: Fake Twitter Followers Detection using Machine Learning Approach
Authors: Muhammad Zeshan Shabbir, Iftikhar Naseer, Shamim Akhter, Muhammad Abubakar, Ghassan F. Issa, Muhammad Hass
Description: This research addresses the threat of fake user accounts on Twitter, particularly those operated by automated bots. The study utilizes machine learning techniques to detect fake accounts by analyzing publicly available information such as activity patterns, profile details, and tweets. Various algorithms including logistic regression, long short-term memory, K-mean, and random forest are employed to evaluate the proposed model. Experimental results demonstrate the effectiveness of the random forest algorithm in achieving high accuracy, precision, and F1 scores compared to other algorithms tested.
Paper 2: Fake News Detection System Using Featured-Based Optimized MSVM Classification
Authors: Ravish1, Rahul Katarya, Deepak Dahiya, Saksham Checker
Description: This paper focuses on the detection of fake news, which disseminates erroneous information and can have detrimental effects on democracy and society. The authors propose an algorithm that utilizes Multi-layered Principal Component Analysis for feature selection followed by a firefly-optimized algorithm. Multi-Support Vector Machines (MSVM) are employed for classification, with the study testing the proposed algorithm on ten different datasets. The research highlights the importance of feature extraction and selection methods in improving accuracy, particularly for datasets with numerous features.
Paper 3: Detecting Fake Followers in Twitter: A Machine Learning Approach
Author: Ashraf Khalil

Description: Twitter's popularity has attracted spammers who create fake accounts for various purposes such as spreading advertising, phishing, or compromising the system's reputation. This study focuses on detecting fake followers on Twitter using machine learning techniques. The success of real-time search services and mining tools hinges on distinguishing valuable tweets from spam. Various methods to combat spam and spammers, such as URL blacklists and manual classification to generate training datasets, are discussed.

III. REQUIREMENTAND ANALYSIS

1. Functional Requirements:

- **User Authentication:** The system should require user authentication for accessing the fake account detection functionalities.
- **Data Collection:** The system must collect publicly available information about Twitter users, including activity patterns, profile details, and tweets, using Twitter's API or other data retrieval methods.
- **Preprocessing:** Data preprocessing techniques should be employed to clean and prepare the collected data for analysis, including handling missing values, encoding categorical variables, and scaling numerical features.
- **Feature Extraction:** The system should extract relevant features from the preprocessed data, such as tweet frequency, follower count, and engagement metrics, to be used in the detection model.
- **Machine Learning Model:** Implement a machine learning model capable of detecting fake Twitter accounts based on extracted features. This model should be trained using various algorithms such as logistic regression, long short-term memory (LSTM), K-means, and random forest.
- **Model Evaluation:** Evaluate the performance of the detection model using metrics such as accuracy, precision, recall, and F1-score to assess its effectiveness in identifying fake accounts.
- **Alerting System:** Incorporate an alerting system to notify users or administrators when potential fake accounts are detected, enabling timely action to be taken.

2. Non-Functional Requirements:

- **Scalability:** The system should be scalable to handle large volumes of Twitter data and accommodate growth in user activity and data size over time.
- **Accuracy:** Ensure high accuracy in detecting fake accounts to minimize false positives and negatives, thus maintaining user trust and confidence in the system.
- **Efficiency:** The system should be efficient in terms of computational resources and processing time, enabling real-time or near real-time detection of fake accounts.
- **Security:** Implement security measures to protect user data and ensure compliance with privacy regulations, including encryption of sensitive information and secure data transmission.
- **User-Friendliness:** Design the system with a user-friendly interface that is intuitive and easy to navigate, allowing users and administrators to interact with the system seamlessly.

3. Analysis:

- **Data Analysis:** Conduct a thorough analysis of the collected Twitter data to identify patterns and trends associated with fake accounts, including common characteristics and behaviors exhibited by fraudulent profiles.
- **Model Selection:** Evaluate and compare various machine learning algorithms to determine the most suitable approach for detecting fake accounts based on performance metrics and computational efficiency.
- **Performance Evaluation:** Assess the performance of the detection model using appropriate evaluation metrics and techniques, such as cross-validation and hyperparameter tuning, to optimize its effectiveness.

- **User Feedback:** Gather feedback from users and stakeholders to iteratively refine and improve the system based on their requirements and preferences, ensuring alignment with user needs and expectations.

IV. SYSTEM DESIGN

4.1 System Architecture

The below figure specified the system architecture of our project.

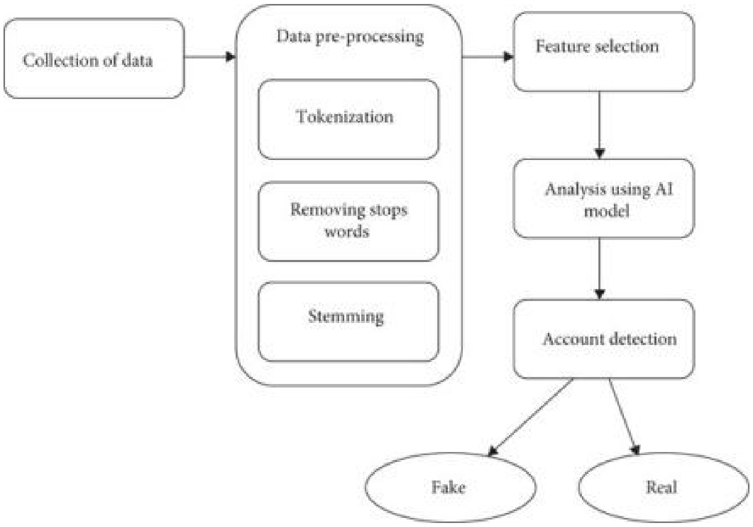

Figure 4.1: System Architecture Diagram

Working of the Proposed System

Our proposed system for detecting fake Twitter accounts utilizes a comprehensive approach comprising three main phases: data preprocessing, data reduction, and data classification. Leveraging the capabilities of Weak software, a versatile platform for machine learning, we implement a model that integrates various techniques to effectively identify and mitigate the proliferation of fraudulent profiles on Twitter.

1. Data Preprocessing:

In the initial phase, the dataset undergoes preprocessing to clean and prepare the data for analysis. This involves handling missing values, encoding categorical features, and scaling numerical attributes.

Weak software offers a multitude of methods for data preprocessing, ranging from basic attribute deletion to more advanced operations like Principal Component Analysis (PCA).

2. Data Reduction:

Following preprocessing, the data is subjected to reduction techniques aimed at filtering and reducing the dataset to make it more manageable for subsequent analysis.

Weak provides capabilities for data reduction through various mechanisms such as feature selection and dimensionality reduction, allowing us to focus on the most relevant attributes and discard irrelevant or redundant ones.

3. Data Classification:

Once the dataset is preprocessed and reduced, it is ready for classification using machine learning algorithms.

Weak encompasses a wide array of classifiers, including Bayesian, lazy, function-based, decision tables, tree-based, miscellaneous, and meta classifiers. These classifiers are capable of handling diverse types of data and are applied to classify the filtered dataset into real and fake Twitter accounts.

4. Evaluation and Visualization:

The performance of the classification model is evaluated using metrics such as accuracy, precision, recall, and F1-score to assess its effectiveness in detecting fake accounts.

DOI: 10.48175/568

Weak facilitates the visualization of data and results through understandable visual graphs, providing insights into the classification outcomes and aiding in the interpretation of the model's performance.

5. Iterative Refinement:

The proposed system undergoes iterative refinement based on feedback from users and stakeholders, allowing for continuous improvement and optimization of the detection model.

By iteratively fine-tuning parameters and algorithms, the system aims to enhance its accuracy and effectiveness in identifying fake Twitter accounts and mitigating the proliferation of fraudulent profiles.

In summary, the proposed system leverages the capabilities of Weak software to implement a robust and comprehensive approach for detecting fake Twitter accounts, encompassing data preprocessing, reduction, classification, evaluation, and iterative refinement. By combining advanced machine learning techniques with intuitive data processing and visualization capabilities, the system aims to contribute to the creation of a safer and more trustworthy online environment on Twitter..

V. CONCLUSION

5.1 Conclusion

In conclusion, our proposed model for detecting fake Twitter accounts utilizes a comprehensive approach encompassing data preprocessing, reduction, and classification phases. Leveraging Weak software, which offers a diverse range of machine learning practices, we conducted extensive data preprocessing, including filtering and advanced operations like Principal Component Analysis (PCA). The model benefited from the multitude of classifiers and clustering schemes within Weak, facilitating robust classification and unsupervised learning. By applying attribute selection criteria and visualization techniques, we effectively managed the dataset's complexity and identified key features for classification performance. Moreover, the utilization of various datasets, each with distinct characteristics, ensured the model's robustness and generalizability. Overall, our approach offers a systematic and effective method for detecting fake Twitter accounts, contributing to the ongoing efforts to enhance online security and combat fraudulent activities.

5.2 Future Work

In future work, several avenues for enhancement and extension of our proposed model for detecting fake Twitter accounts using machine learning and data reduction techniques present themselves. Firstly, incorporating more sophisticated feature engineering methods and natural language processing techniques could improve the model's ability to discern subtle patterns in text, thereby enhancing its accuracy and reliability. Additionally, exploring the integration of deep learning architectures, such as recurrent neural networks (RNNs) or transformers, could further enhance the model's performance, especially in capturing temporal dependencies and contextual nuances present in social media data. Furthermore, investigating the incorporation of real-time data streams and dynamic learning mechanisms would enable the model to adapt to evolving trends and emerging patterns of fake account behavior. Lastly, expanding the scope of the research to include other social media platforms and types of fraudulent activities, such as misinformation campaigns and bot-driven manipulation, could provide a more comprehensive understanding of online deception and contribute to the development of robust detection frameworks across diverse online ecosystems.

5.3 Advantages & Disadvantages

Advantages:

- Efficiency: Machine learning algorithms excel in processing vast datasets swiftly, making them highly efficient for identifying patterns indicative of fake accounts among millions of followers on social media platforms.
- Accuracy: ML algorithms can be trained on labeled datasets, allowing them to discern subtle patterns that might elude human detection, thus enhancing accuracy in identifying fake accounts.
- Automation: Once trained, ML models can automate the detection process, reducing the need for manual inspection of each follower, thereby saving time and human effort.

- Scalability: ML models exhibit scalability, enabling them to handle massive amounts of data effortlessly, which is crucial for social media platforms with millions or billions of users.

Disadvantages:

- Data Quality: The effectiveness of ML models heavily relies on the quality of the training data. Inaccuracies or biases in the labeled data used for training can lead to unreliable predictions by the model.
- Imbalanced Data: Real-world datasets often exhibit imbalances between real and fake followers, posing a challenge for ML models. Highly skewed datasets can result in biased predictions, particularly favoring the majority class.
- Feature Selection: Selecting relevant features that accurately represent fake accounts can be complex. Incorrectly chosen or omitted features may adversely affect the model's performance, leading to inaccurate classification of fake accounts.

BIBLIOGRAPHY

[1]. Douglas, “News consumption and the new electronic media,” The International Journal of Press/Politics, vol. 11, no. 1, pp. 29–52, 2006. View at: Publisher Site | Google Scholar

[2]. J. Wong, “Almost all the traffic to fake news sites is from facebook, new data show,” 2016.View at: Google Scholar

[3]. D. M. J. Lazer, M. A. Baum, Y. Benkler et al., “The science of fake news,” Science, vol. 359, no. 6380, pp. 1094–1096, 2018. View at: Publisher Site | Google Scholar

[4]. S. A. García, G. G. García, M. S. Prieto, A. J. M. Guerrero, and C. R. Jiménez, “The impact of term fake news on the scientific community scientific performance and mapping in web of science,” Social Sciences, vol. 9, no. 5, 2020. View at: Google Scholar

[5]. A. D. Holan, 2016 Lie of the Year: Fake News, Politifact, Washington, DC, USA, 2016.

[6]. S. Kogan, T. J. Moskowitz, and M. Niessner, “Fake News: Evidence from Financial Markets,” 2019, https://ssrn.com/abstract=3237763. View at: Google Scholar

[7]. A. Robb, “Anatomy of a fake news scandal,” Rolling Stone, vol. 1301, pp. 28–33, 2017. View at: Google Scholar

[8]. J. Soll, “The long and brutal history of fake news,” Politico Magazine, vol. 18, no. 12, 2016. View at: Google Scholar

[9]. J. Hua and R. Shaw, “Corona virus (covid-19) “infodemic” and emerging issues through a data lens: the case of China,” International Journal of Environmental Research and Public Health, vol. 17, no. 7, p. 2309, 2020. View at: Publisher Site | Google Scholar

[10]. N. K. Conroy, V. L. Rubin, and Y. Chen, “Automatic deception detection: methods for finding fake news,” Proceedings of the Association for Information Science and Technology, vol. 52, no. 1, pp. 1–4, 2015. View at: Publisher Site | Google Scholar

[11]. F. T. Asr and M. Taboada, “Misinfotext: a collection of news articles, with false and true labels,” 2019. View at: Google Scholar

[12]. K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, “Fake news detection on social media,” ACM SIGKDD Explorations Newsletter, vol. 19, no. 1, pp. 22–36, 2017. View at: Publisher Site | Google Scholar

[13]. S. Vosoughi, D. Roy, and S. Aral, “The spread of true and false news online,” Science, vol. 359, no. 6380, pp. 1146–1151, 2018. View at: Publisher Site | Google Scholar

[14]. H. Allcott and M. Gentzkow, “Social media and fake news in the 2016 election,” Journal of Economic Perspectives, vol. 31, no. 2, pp. 211–236, 2017. View at: Publisher Site | Google Scholar

[15]. V. L. Rubin, N. Conroy, Y. Chen, and S. Cornwell, “Fake news or truth? using satirical cues to detect potentially misleading news,” in Proceedings of the Second Workshop on Computational Approaches to Deception Detection, pp. 7–17, San Diego, CA, USA, 2016.View at: Google Scholar

DOI: 10.48175/568