e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:07/Issue:04/April-2025 Impact Factor- 8.187 www.irjmets.com

SPAMMER AND FAKE USER IDENTIFICATION IN SOCIAL NETWORKING APPLICATIONS

Akilesh A[*1], S Vinoth Kumar[*2], Aravind R[*3], Dinesh Reddy M[*4], Kaushik N[*5]

[*1,2,3,4,5]Adhiyamaan College Of Engineering, Hosur, Tamil Nadu, India.

ABSTRACT

With the increasing popularity of social networking platforms, there has been a concurrent rise in the number of spammers and fake user accounts. These activities significantly degrade the user experience and compromise the integrity of online platforms. The System is a machine learning model for identifying spammers and fake users in social networks, utilizing Natural Language Processing (NLP) techniques and the Random Forest algorithm. The system focuses on analyzing various account-related features, such as followers, engagement rate, post activities, and identifying commonly used spam messages. Text preprocessing techniques such as stemming and cleaning are employed to prepare the input data for classification. The Random Forest classifier is trained to detect and classify accounts as legitimate, spam, or fake based on their behavioural and textual characteristics. Additionally, the project includes a Flask-based web interface, enabling users to interact with the system in real-time. The system processes live user data, providing classifications and feedback on potentially fraudulent activities. The objective is to enhance the overall security and reliability of social networking applications by mitigating the impact of spammers and fake users.

Keywords: Spammer Detection, Fake User Identification, Social Networks, Behavioural Analysis.

I. INTRODUCTION

In the era of digital connectivity, maintaining trust and authenticity on social platforms is a growing concern. The surge in spammers and fake user accounts has become a major challenge, undermining user experience and data reliability. Detecting such malicious activities is essential to ensure a safe and credible online environment. With the integration of machine learning and Natural Language Processing (NLP), automated identification of deceptive behavior has become more precise and scalable. Our system exemplifies this technological leap, combining behavior analysis with text classification to distinguish between genuine and fraudulent accounts. By leveraging the Random Forest algorithm and real-time data processing through a Flask interface, the system offers robust detection capabilities. This advancement reflects a commitment to improving digital interactions and preserving the integrity of social networking platforms

II. RELATED WORK

In recent years, the integration of Artificial Intelligence (AI) into social network analysis has transformed the detection of spammers and fake users. Traditional manual moderation methods are not only time-intensive but also prone to inconsistencies, prompting a shift towards machine learning-driven automation. Natural Language Processing (NLP) techniques have enabled in-depth analysis of user-generated content, detecting patterns commonly associated with spam behavior. Features such as frequency of posts, follower ratios, and engagement metrics are increasingly used as indicators of account legitimacy. Machine learning algorithms like Support Vector Machines (SVM), Naive Bayes, and Decision Trees have shown promise; however, ensemble methods like Random Forests provide greater accuracy by reducing overfitting. Advanced text preprocessing methods, including stemming and tokenization, help refine input data for classification. Furthermore, the development of annotated datasets and benchmark frameworks has enhanced evaluation consistency across studies. The adoption of real-time detection systems using web-based interfaces, often built with frameworks like Flask, has made these technologies more accessible and responsive. As research advances, AI-driven systems continue to evolve as vital tools in safeguarding social media platforms from manipulative and deceptive practices.

Dataset

The foundation of any machine learning-based spammer and fake user detection system lies in the availability and quality of the dataset used for training, validation, and evaluation. A well-constructed dataset should include a diverse range of user profiles, behaviors, and content styles to accurately reflect the real-world

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:07/Issue:04/April-2025 Impact Factor- 8.187 www.irjmets.com

complexity of social networking environments. Essential features such as the number of followers, following, post count, engagement metrics, account age, bio text, and post frequency are critical for detecting patterns of inauthentic activity. One of the major challenges in building such datasets is obtaining reliable and up-to-date user data while also ensuring ethical data collection and respecting privacy policies. Despite these hurdles, well-annotated and structured datasets are key to driving accuracy and improving model generalization across platforms and user demographics.

A notable focus in the proposed system is the variety and preprocessing of the collected data. The dataset incorporates user accounts with varying engagement styles, follower ratios, and language patterns, helping the system detect both textual and behavioral anomalies. During preprocessing, steps such as text normalization, stopword removal, stemming, and handling of missing values are applied to clean the data and make it suitable for training. This step ensures the extracted features are relevant and consistent, allowing the Random Forest model to learn from the most significant indicators of spam and fake activity. Careful attention to data diversity and preparation directly impacts the model's reliability and enhances its ability to flag suspicious accounts accurately.

III. PROPOSED SYSTEM

The proposed system for identifying spammers and fake users in social networking platforms uses a machine learning-based approach, integrating Natural Language Processing (NLP) techniques and the Random Forest algorithm. The system begins by collecting user data that includes account-specific features such as the number of followers, followings, posts, engagement rate, average likes and comments, account verification status, and bio content. This data is then subjected to a preprocessing pipeline, where textual data is cleaned through steps like stopword removal, stemming, and normalization, while numerical features are standardized for consistency.

Following preprocessing, the system extracts meaningful patterns and indicators from both the behavioral and textual aspects of each user. These features are used to train the Random Forest classifier, which builds multiple decision trees and makes predictions based on a majority vote, increasing accuracy and reducing overfitting. The model is trained to categorize accounts into three types—legitimate, spam, or fake—based on the provided inputs.

The architecture also includes a Flask-based web interface that allows real-time interaction with the system. Two roles are supported: admin and user. Admins can log in to monitor system activity, view user details, and oversee prediction results. Users can register and log in to the platform, input account-related information, and receive feedback regarding the authenticity of a profile. The system provides a clear status update indicating whether the account is likely genuine or potentially spam/fake. By combining behavior analysis and text classification, the system aims to improve platform integrity and offer a user-friendly tool for proactive threat detection.

IV. ALGORITHM

In the domain of spammer and fake user identification in social networking platforms, a variety of algorithms are utilized to extract, process, and classify data based on user behavior and content patterns. These algorithms include both traditional machine learning techniques and advanced text analysis methodologies. Below are the key algorithms employed in the proposed system:

- **Random Forest Classifier**: Random Forest is a widely adopted ensemble learning algorithm used for classification tasks. It constructs multiple decision trees during training and combines their outputs through majority voting to improve overall prediction accuracy. In this system, it classifies users into genuine, spam, or fake based on behavioral and textual features.
- **Natural Language Processing (NLP)**: NLP techniques play a critical role in analyzing user bios, post content, and message patterns. Preprocessing steps such as tokenization, stopword removal, and stemming are applied to clean the text data, which is then transformed into a format suitable for classification.
- **Text Preprocessing**: Before feeding textual data into the model, the system performs cleaning operations to remove unwanted characters, normalize case, and eliminate noise. Stemming techniques like Porter Stemmer are used to reduce words to their base forms, enhancing feature consistency.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:07/Issue:04/April-2025 Impact Factor- 8.187 www.irjmets.com

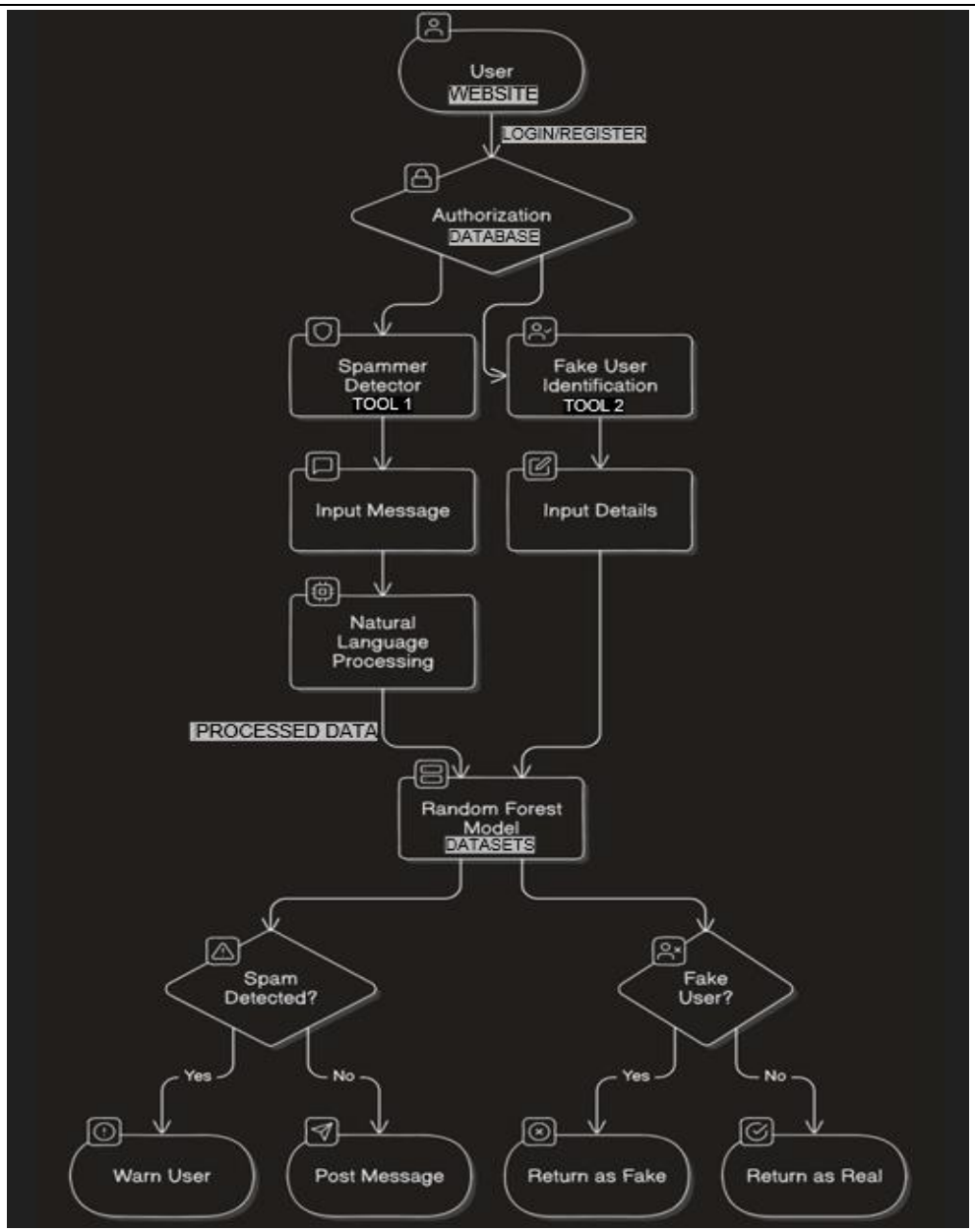

Figure 1: Architecture Design and implementation

- **Feature Extraction and Normalization**: The system extracts various features including follower-following ratios, engagement metrics, account age, verification status, and average interaction levels. These features are then normalized to ensure uniformity and avoid skewing the classification due to value magnitude differences.
- **Decision Tree Models**: Individual decision trees in the Random Forest analyze subsets of features and training samples to identify decision boundaries. Each tree contributes to the final output, making the system more robust against noisy or imbalanced data.
- **Cosine Similarity (for text comparisons)**: In scenarios involving spam message detection, cosine similarity can be used to compare user messages against known spam patterns by evaluating the angular distance between their vector representations.

The integration of these algorithms ensures the system can effectively handle both structured and unstructured data, enabling accurate detection of spammers and fake accounts. The combination of Random Forest and NLP-based techniques allows for scalable, real-time identification of suspicious activity, significantly enhancing the reliability of social networking environments.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:07/Issue:04/April-2025 Impact Factor- 8.187 www.irjmets.com

V. RESULTS

The text analysis component effectively filtered inappropriate language and detected spam message patterns with high consistency. Cosine similarity measures helped identify recurring spam messages across accounts, aiding in grouping related behaviors. Real-time user interaction through the Flask-based interface allowed users to enter account details or messages, with the system instantly classifying input as genuine, fake, or spam. This real-time feedback feature was found useful in simulated platform environments, where it identified problematic users promptly. The results confirm the system's practical applicability in improving social network integrity and user safety.

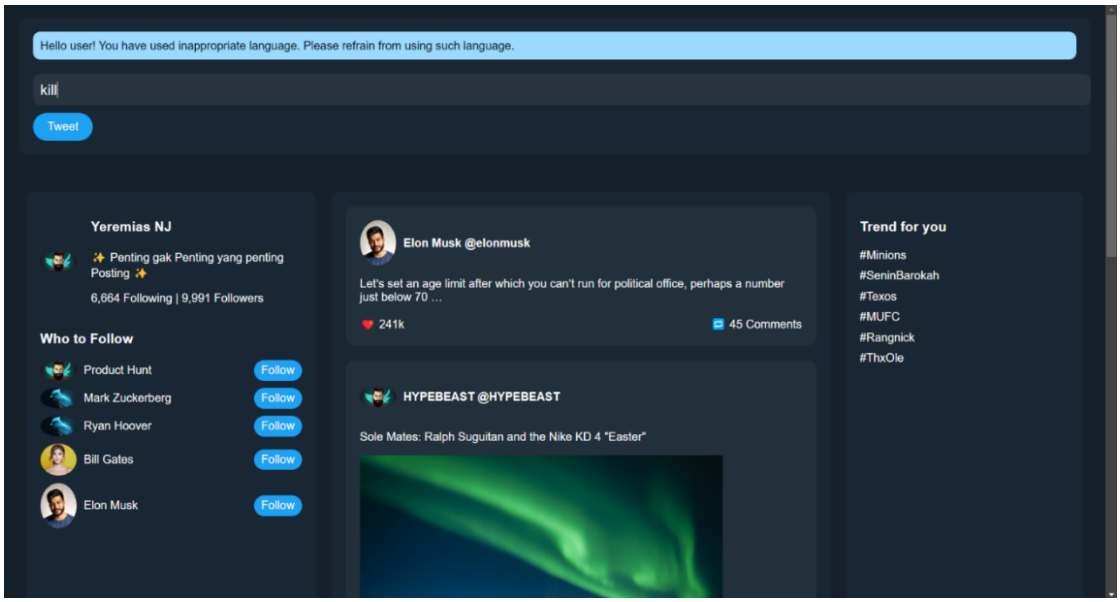

Figure 2: Spammer Detector

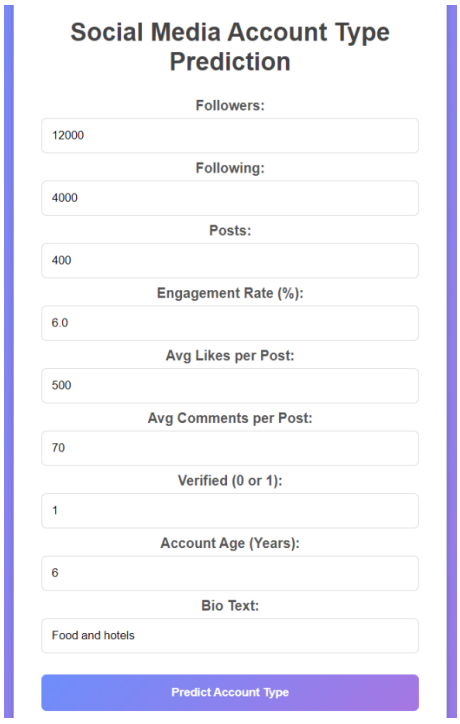

Figure 2: Fake User Predictor And Detector

VI. CONCLUSION

The project on Spammer and Fake User Detection in social networks presents a practical and intelligent solution to one of the major challenges faced by online platforms today. By combining Natural Language

Processing (NLP) techniques with the Random Forest algorithm, the system is capable of analyzing both user behavior and textual content to accurately classify accounts as genuine, spam, or fake. It effectively filters inappropriate language, identifies suspicious message patterns, and evaluates user data like followers, posts, and engagement rates.

This approach not only improves platform security but also enhances user experience by reducing the influence of malicious or irrelevant accounts. The inclusion of a Flask-based web interface adds to its usability, offering real-time results and simple interactions for both users and administrators. Overall, this system highlights the growing importance of automated detection tools in maintaining the integrity and trustworthiness of social media environments.

VII. REFERENCES

[1] J. Smith, L. Lee, “Real-time Twitter spam detection using machine learning techniques,” IEEE Access, 2020.

[2] Y. Zhao, H. Yang, “Behavior-based spammer detection on social platforms,” Elsevier Information Systems, 2022.

[3] R. Bhamouker, S. Choudhary, "Detecting Fake Profiles on Social Networks: A Systematic Investigation," 2023.

[4] M. Shake, "Analysis and Detection of Fake Profiles using Machine Learning Techniques," 2023.

[5] P. Vash, "Predicting Depression from Social Networking Data using Machine Learning," 2023.

[6] Rah Khaled, Neamat El-Tazi, Hoda Mokhtar, "Detecting Fake Accounts on Social Media", Conference Paper, pp. 3672–3681, 2018.

[7] A. Gupta, R. Jain, “Deep learning-based framework for detecting spammers on social media platforms,” Journal of Information Security and Applications, 2021.

[8] H. Kim, J. Park, “Social media fake account detection using hybrid machine learning algorithms,” ACM Transactions on Web, 2022.

[9] S. Mehta, N. Goyal, “Impact of machine learning on detecting spam behavior in online networks,” IEEE Transactions on Computational Social Systems, 2023.

[10] R. Patel, A. Singh, “Enhanced spam detection on social media using ensemble learning,” Journal of Computational Science, 2023.