Prediction of Fake Instagram Profile Using Machine Learning

Aswin S[1], Sanjeev M S[2], Ajanesh R[3], Ms. M.Varsha[4]
[1,2,3]*B.Sc. Information Technology, Department of Information Technology, Nehru Arts and Science College, Coimbatore India*
[4]*Assistant Professor, Department of Information Technology, Nehru Arts and Science College, Coimbatore India*

Abstract - Fake profiles on social media platforms like Instagram are increasingly used for spam, scams, and misinformation. This paper proposes a machine learning-based system to automatically identify fake Instagram accounts using profile-level features such as follower count, engagement ratio, account age, and posting behavior. Multiple classification algorithms including Random Forest, Logistic Regression, and Support Vector Machine are evaluated to determine the most accurate model. The proposed system aims to assist platforms and users in detecting suspicious accounts efficiently and improving online trust and safety.

I.INTRODUCTION

Social media platforms have become essential for communication and business. However, the rise of fake accounts poses serious threats including fraud, impersonation, and spam. Manual detection is inefficient due to the massive number of users. Therefore, an automated intelligent system is required to detect fake profiles accurately. This project focuses on building a machine learning model that analyzes account characteristics and predicts whether a profile is genuine or fake following imbalance, posting frequency, and profile completeness, the system learns distinguishing characteristics of fraudulent accounts. Experimental results show that the proposed approach achieves high classification performance, demonstrating its effectiveness as a proactive tool for enhancing social media security

a. Existing System and Limitations

The current methods used by social media platforms to identify fake Instagram profiles primarily rely on manual reporting mechanisms and rule-based filters. Users can report suspicious accounts, after which moderators review them based on predefined guidelines. Additionally, automated systems often use simple heuristics such as detecting excessive follow requests, repetitive content, or spam links. While these approaches help in identifying obvious fake accounts, they depend heavily on static rules and human intervention, which makes the process slow and less efficient when dealing with the massive scale of social media data.

b. Workflow

1. Data collection (public profile features dataset)
2. Data preprocessing and feature selection
3. Model training using multiple algorithms
4. Performance evaluation (accuracy, precision, recall)
5. Deployment as a prediction tool

c.Key Features Used

- Followers count
- Following count
- Posts count
- Engagement rate
- Username patterns

III. METHODOLOGY

The methodology of this project follows a systematic machine learning workflow to ensure accurate detection of fake Instagram profiles. Initially, a dataset containing various profile attributes is collected and preprocessed by handling missing values, removing duplicates, and normalizing numerical features. Relevant features are then selected and engineered, such as calculating the follower-to-following ratio and engagement indicators, to improve model performance. The processed dataset is split into

training and testing sets, and multiple classification algorithms are trained to learn patterns that differentiate fake and genuine accounts. The models are evaluated using performance metrics such as accuracy, precision, recall, and F1-score, and the most effective model is chosen for deployment. This structured approach ensures reliability, reproducibility, and scalability of the detection system.

IV. PROPOSED SYSTEM

The proposed system uses supervised machine learning to classify Instagram profiles.
Workflow
Data preprocessing and feature selection
Data collection (public profile features dataset) Model training using multiple algorithms Performance evaluation (accuracy, precision, recall) Deployment as a prediction tool
Key Features Used`
Followers count Following count Posts count Engagement rate Username patterns

V. SYSTEM ARCHITECTURE

The system architecture is designed as a modular pipeline that processes Instagram profile data from input to prediction. Initially, profile information such as follower count, following count, posts, and engagement metrics is provided as input to the system. This data passes through a preprocessing module where cleaning, normalization, and feature extraction are performed to prepare it for analysis. The processed features are then fed into the trained machine learning model, which analyzes patterns and classifies the profile as either fake or genuine. Finally, the prediction results are displayed through a user interface or stored for further monitoring. This layered architecture ensures scalability, easy integration, and efficient real-time detection of suspicious profiles.

VI. APPLICATIONS

The proposed fake Instagram profile detection system has a wide range of practical applications in enhancing online security and trust. Social media platforms can integrate this system to automatically identify and remove fraudulent accounts, thereby reducing spam, scams, and impersonation activities. Businesses and influencers can use the tool to verify the authenticity of followers and protect their brand reputation from fake engagement. Additionally, cybersecurity agencies can leverage the system for monitoring suspicious online behavior and preventing digital fraud. By providing accurate and automated detection, the solution contributes to creating a safer and more reliable social media environment for users and organizations alike.

VII. ADVANTAGES AND CHALLENGES

The proposed system offers several advantages, including automated detection of fake profiles, improved accuracy compared to rule-based methods, and the ability to scale for large volumes of social media data. By leveraging machine learning, the system can continuously learn from new patterns and adapt to evolving fraudulent behaviors, reducing manual effort and enhancing platform security. However, the system also faces certain challenges, such as the need for high- quality and balanced datasets to avoid biased predictions. Privacy considerations and data access limitations may restrict the availability of useful features, and sophisticated fake accounts that closely mimic genuine user behavior can still pose difficulties. Addressing these challenges is essential to further improve the reliability and effectiveness of the detection system.

VIII. FUTURE SCOPE

The future scope of this project includes enhancing the detection system by incorporating advanced deep learning techniques and larger, more diverse datasets to improve prediction accuracy. Real-time monitoring can be implemented to identify suspicious accounts instantly as they are created or become active. The system can also be extended to analyze behavioral patterns such as interaction networks, comment sentiment, and activity timelines for more comprehensive detection. Additionally, integrating the model into web or mobile applications and expanding it to support multiple social media platforms would increase its practical usability and impact. Continuous updates and model retraining will ensure the system remains effective against evolving fake profile strategies.

XI. CONCLUSION

In conclusion, this project demonstrates the effectiveness of machine learning techniques in detecting fake Instagram profiles by analyzing profile attributes and behavioral patterns. The proposed system provides a scalable and automated solution that improves detection accuracy compared to traditional manual and rule- based approaches. By leveraging data-driven insights, the model can identify suspicious accounts efficiently and contribute to reducing spam, fraud, and impersonation on social media platforms. Although challenges such as data availability and evolving fake account strategies remain, the results show that intelligent detection
systems can significantly enhance online trust and security. Future improvements and continuous model updates will further strengthen the system's reliability and real-world applicability.

REFERENCES

[1] M. Fire, R. Goldschmidt, and Y. Elovici, "Online social networks: Threats and solutions," IEEE Communications Surveys & Tutorials, vol. 16, no. 4, pp. 2019–2036, 2014.

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

[3] S. Cresci, R. Di Pietro, M. Petrocchi, A. Spognardi, and M. Tesconi, "The paradigm-shift of social spambots: Evidence, theories, and tools for the arms race," in Proceedings of the International World Wide Web Conference (WWW), 2017, pp. 963–972.

[4] A. Gupta and R. Kaushal, "Improving spam detection in online social networks," in Proceedings of the IEEE International Conference on Communication Systems (ICCS), 2015, pp. 17–21.

[5] F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida, "Detecting spammers on Twitter," in Proceedings of the Collaboration, Electronic Messaging, Anti-Abuse and Spam Conference (CEAS), 2010.

[6] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed. San Francisco, CA, USA: Morgan Kaufmann, 2012.

[7] Scikit-learn Developers, "Scikit-learn: Machine learning in Python." [Online]. Available: https://scikit-learn.org

[8] Instagram Help Center, "Community guidelines and fake account policies." [Online]. Available: https://help.instagram.com