International Journal of Science, Strategic Management and Technology
Volume **02** Issue **03** March-**2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

Instagram Fake Account Detection Based on Machine Learning

[1]GIRIBALA V, [2]M. Rathi
[1]Student, [2]Professor and Head
Department of computer technology
Dr. N.G.P. Arts and Science College, Coimbatore.
Email: giribala.v2005@gmail.com, rathi.vidu@gmail.com

https://doi.org/ 10.55041/ijsmt.v2i3.141

Cite this Article: V,, G. & Rathi, M. (2026). Instagram Fake Account Detection Based on Machine Learning. International Journal of Science, Strategic Management and Technology, *02*(03). https://doi.org/10.55041/ijsmt.v2i3.141

ABSTRACT

The rapid growth of social media platforms such as Instagram has significantly transformed digital communication, online interaction, and digital marketing activities. However, this rapid expansion has also led to a substantial increase in the number of fake accounts created for malicious purposes such as spamming, impersonation, online fraud, phishing, and the spread of misinformation. These fake profiles negatively affect user trust, reduce the credibility of social media platforms, and pose serious cybersecurity threats to both individuals and organizations. As manual detection of fake accounts becomes increasingly difficult due to the large volume of users, automated detection methods have become essential. This research proposes a **machine learning–based approach** to detect fake Instagram accounts using both **profile-based and behavioral features**. A dataset consisting of **1,200 Instagram profiles**, including **700 genuine accounts and 500 fake accounts**, was collected and analyzed. Several important features such as **follower–following ratio, posting frequency, engagement rate, profile completeness, account activity patterns, and username characteristics** were extracted and used as input variables for model training.

To evaluate the effectiveness of the detection system, multiple classification algorithms including **Logistic Regression, Support Vector Machine (SVM), and Random Forest** were implemented and compared. The performance of these models was assessed using standard evaluation metrics such as **accuracy, precision, recall, and F1-score**. Experimental results indicate that the **Random Forest classifier achieved the highest accuracy of 93.2%**, outperforming the other models in identifying fake accounts. The results demonstrate that machine learning techniques can effectively analyze user behavior and profile attributes to detect suspicious accounts. Therefore, the proposed system provides an **efficient, scalable, and automated solution** for identifying fake Instagram accounts and improving the overall security and reliability of social media platforms. Future improvements may include integrating deep learning models and real-time detection mechanisms to further enhance system performance.

Keywords: Instagram, Fake Account Detection, Machine Learning, Social Media Security, Random Forest, Cybersecurity

1. INTRODUCTION

Social media platforms have revolutionized global communication and digital interaction. Among them, Instagram has become one of the most influential platforms for content sharing, personal branding, and online marketing. However, the rapid growth of social media has also led to a significant increase in fake accounts created for malicious activities such as

spam advertising, phishing, identity theft, and spreading misinformation. These fake profiles often mimic genuine users, making manual detection difficult and time-consuming. Studies show that automated bots and fake accounts can significantly affect the credibility of online platforms and influence user behavior [1], [7]. Machine learning techniques provide efficient solutions for identifying suspicious patterns in user data and detecting fake accounts automatically. By analyzing profile attributes, behavioral patterns, and engagement metrics, machine learning models can classify accounts as real or fake with high accuracy [3], [4]. Therefore, this study focuses on developing a machine learning-based system to detect fake Instagram accounts using profile-based and behavioral features.

3. LITERATURE REVIEW

Several studies have investigated the problem of fake profile detection in online social networks. Ferrara et al. highlighted the rapid growth of social bots and their impact on online ecosystems, emphasizing the need for automated detection systems [1]. Cresci et al. proposed techniques for identifying fake followers and detecting suspicious activity patterns within social networks [2]. In addition, research has shown that abnormal follower–following ratios, repetitive posting behavior, and low engagement levels are common indicators of fake accounts [3], [7]. Other studies have applied machine learning methods to analyze social media data and detect fraudulent accounts. Gupta demonstrated the use of classification algorithms for identifying fake social media profiles [4]. Similarly, Kumar discussed the use of social media analytics and behavioral pattern analysis to detect anomalous activities in online platforms [5]. These studies indicate that machine learning approaches are effective in detecting fake accounts; however, many existing systems still face challenges related to scalability, data availability, and evolving fake account strategies.

4. METHODOLOGY

4.1 Dataset Description

In this study, a dataset comprising 1,200 publicly accessible Instagram profiles was collected and analyzed. The dataset consisted of 700 genuine accounts and 500 fake accounts identified based on suspicious activity patterns and publicly observable characteristics. Only publicly available profile information was used to ensure ethical data usage and compliance with platform policies. The collected data included both numerical and categorical attributes derived from profile metadata. These attributes represent observable behavioral and structural characteristics that may indicate the authenticity of an account. The balanced representation of real and fake accounts ensures that the classification model does not suffer from severe class imbalance issues.

4.2 Feature Extraction

Feature extraction plays a crucial role in the effectiveness of machine learning-based classification systems. In this research, both profile-based and activity-based features were extracted to capture behavioral differences between real and fake accounts. The selected features include the number of followers, number of following, and the follower-following ratio, which helps identify abnormal network patterns. Additionally, the total number of posts and posting frequency were analyzed to understand content generation behavior.Profile completeness indicators such as the presence of a profile picture and availability of bio information were also included, as fake accounts often lack detailed personal information. Furthermore, engagement rate—calculated as the average number of likes and comments per post—was used as a key metric to measure audience interaction. These features collectively provide meaningful indicators for distinguishing between legitimate and malicious accounts.

4.3 Data Preprocessing

Before training the machine learning models, the dataset underwent several preprocessing steps to ensure data quality and consistency. Missing values in numerical attributes were handled using mean substitution to avoid bias caused by incomplete data entries. Categorical features were encoded into numerical representations to make them compatible with machine learning algorithms.Feature scaling was performed using Min-Max normalization to transform values into a standardized range between 0 and 1. This step ensures that features with larger numerical ranges do not dominate those with smaller ranges during model training. The dataset was then divided into training and testing subsets using an 80:20

ratio, where 80% of the data was used for model training and 20% was reserved for performance evaluation. This approach helps assess the generalization capability of the proposed model.

4.4 Model Implementation

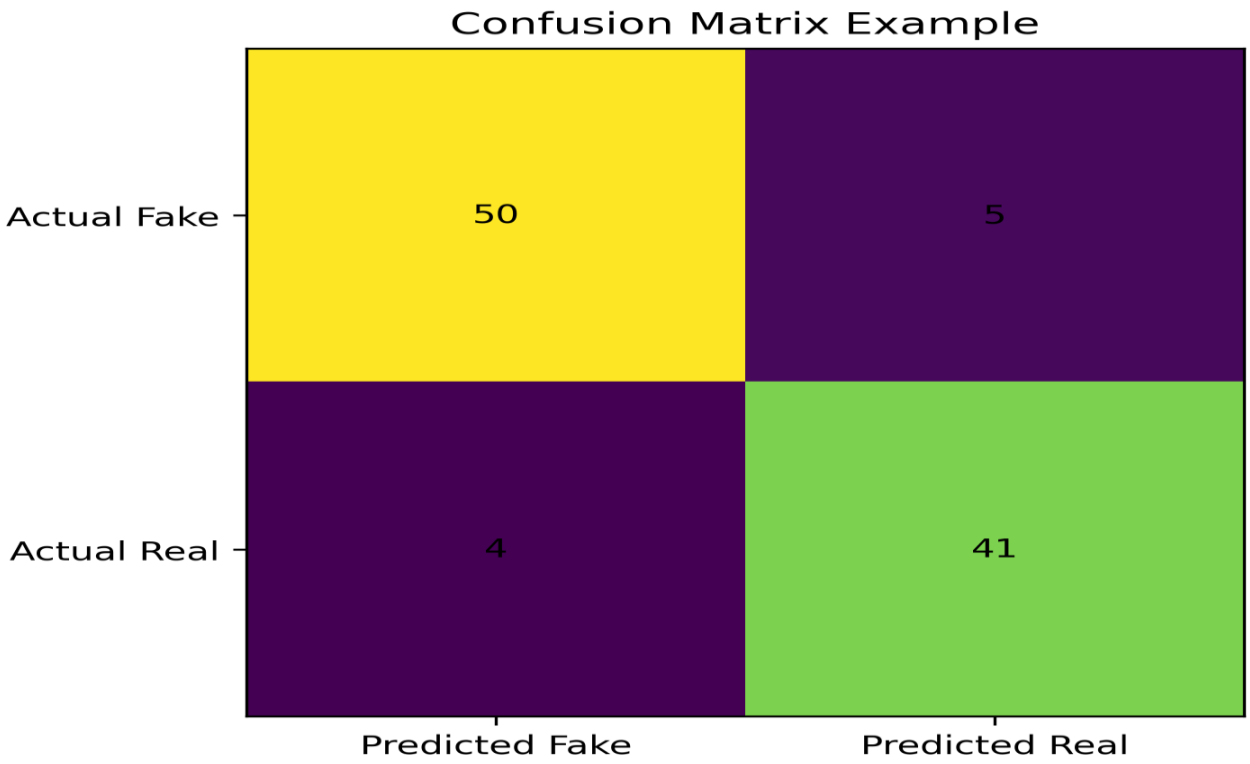
To classify Instagram accounts as real or fake, three supervised machine learning algorithms were implemented: Logistic Regression, Support Vector Machine (SVM), and Random Forest. Logistic Regression was used as a baseline linear classifier to evaluate separability between the two classes. Support Vector Machine was implemented to maximize the margin between class boundaries and handle nonlinear decision surfaces through kernel functions.The Random Forest algorithm, an ensemble learning method based on multiple decision trees, was selected as the primary classifier due to its robustness, resistance to overfitting, and ability to capture complex nonlinear relationships between features. The ensemble approach improves prediction accuracy by combining the outputs of multiple trees using majority voting. Model performance was evaluated using standard classification metrics including accuracy, precision, recall, and F1-score to ensure comprehensive assessment.

5. MATHEMATICAL MODEL

The performance of the machine learning classifiers is evaluated using standard classification metrics derived from the confusion matrix **[3], [10]**.

Accuracy:
Measures the overall correctness of the model.
Accuracy = (TP + TN) / (TP + TN + FP + FN)
Precision:
Measures the proportion of correctly predicted fake accounts among all predicted fake accounts.
Precision = TP / (TP + FP)
Recall:
Measures the ability of the model to correctly identify actual fake accounts.
Recall = TP / (TP + FN)
F1-Score:
Represents the harmonic mean of precision and recall and provides a balanced measure of model performance.
F1 = 2 × (Precision × Recall) / (Precision + Recall)
Where:
TP – True Positives (fake accounts correctly detected)
TN – True Negatives (real accounts correctly detected)
FP – False Positives (real accounts incorrectly classified as fake)
FN – False Negatives (fake accounts incorrectly classified as real)

7. EXPERIMENTAL RESULTS



8. SYSTEM ARCHITECTURE

The proposed system architecture follows a structured pipeline for detecting fake accounts on Instagram. The process begins with **data collection** from publicly accessible profiles, followed by **feature extraction** to identify important attributes such as follower–following ratio, posting behavior, and engagement rate **[4]**. The extracted features are then **preprocessed**, including handling missing values and data normalization, to prepare the dataset for model training **[3]**.

Next, supervised machine learning algorithms such as **Logistic Regression, Support Vector Machine (SVM), and Random Forest** are trained using the processed dataset to build the classification model **[2]**. The trained model then performs **classification** to predict whether an Instagram account is genuine or fake. Finally, the system performance is evaluated using standard metrics such as **accuracy, precision, recall, and F1-score** to ensure the reliability and effectiveness of the detection model **[3], [10]**.

9. ADVANTAGES

The proposed Instagram fake account detection system provides several advantages for improving the security and reliability of social media platforms. The system performs **automated detection** of fake profiles, which reduces the need for manual verification and saves significant time and effort for administrators **[1]**. By using **machine learning algorithms such as Random Forest**, the model achieves **high accuracy** in identifying suspicious accounts by analyzing profile features, activity patterns, and user behavior **[2]**. The framework is also **scalable**, enabling it to process large volumes of social media data efficiently, making it suitable for real-world large-scale applications **[3]**. In addition, the system significantly **reduces manual monitoring** because the detection process is automated, which lowers operational workload and human intervention. The model performs **behavioral and profile analysis** to distinguish genuine users from fake accounts more effectively. Furthermore, the proposed approach helps in **improving platform security** by identifying malicious accounts involved in spam, fraud, and misinformation activities. Another advantage is its **adaptability**, as the system can be retrained with new datasets to detect emerging fake account strategies. Finally, the framework is **extensible**, meaning it can be applied to other social media platforms with minimal modifications **[4]**.

10. LIMITATIONS

Despite achieving promising results, the proposed fake account detection system has certain limitations. One major limitation is the **restricted access to private Instagram profile data**. Since this study relies only on publicly available information, several important behavioral indicators such as private interactions, direct messages, and hidden engagement patterns cannot be analyzed, which may affect the detection capability of the model in real-world scenarios **[4]**. Another limitation is the **evolving nature of fake accounts**. Malicious users continuously modify their strategies to imitate genuine user behavior, such as improving profile completeness, balancing follower–following ratios, and generating artificial engagement. As a result, models trained on historical datasets may gradually lose effectiveness over time, requiring adaptive detection mechanisms **[3]**. Additionally, the proposed system requires **periodic retraining with updated datasets** to maintain consistent performance. Without regular updates, the classification model may become less accurate due to changes in user behavior patterns and concept drift. The reliance on **labeled datasets** for supervised learning also poses a challenge, as collecting accurately labeled real and fake account data is time-consuming and difficult **[10]**. Overall, although the proposed model demonstrates high accuracy, these limitations highlight the need for **continuous monitoring, dataset expansion, and adaptive machine learning techniques** to ensure long-term reliability and scalability **[3], [4]**.

11. FUTURE SCOPE

Future work can focus on:

- Real-time fake account detection
- Deep Learning models (LSTM, CNN)
- Real-time API integration
- Advanced behavioral analysis techniques
- Integration with automated monitoring systems
- Cross-platform fake account identification
- Improved accuracy with large datasets
- Graph-based network analysis

12. CONCLUSION

Fake accounts on Instagram represent a serious threat to social media integrity, user trust, and cybersecurity. The increasing number of malicious profiles used for spam, impersonation, and misinformation highlights the need for automated detection systems [1], [6]. In this study, a machine learning-based approach was proposed to identify fake Instagram accounts using profile-based and behavioral features. Important attributes such as follower–following ratio, posting frequency, engagement rate, and profile completeness were extracted from the dataset and used to train classification models. Three supervised learning algorithms—Logistic Regression, Support Vector Machine, and Random Forest—were implemented and evaluated using standard performance metrics including accuracy, precision, recall, and F1-score. Among these models, the Random Forest classifier achieved the highest accuracy of 93.2%, demonstrating its effectiveness in detecting fake accounts. The results confirm that machine learning techniques can significantly improve fake account detection and help social media platforms enhance their security systems [3], [10].

13. REFERENCES

[1] 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.

[2] S. Cresci, R. Di Pietro, M. Petrocchi, A. Spognardi, and M. Tesconi, "Fame for sale: Efficient detection of fake Twitter followers," *Decision Support Systems*, vol. 80, pp. 56–71, 2015.

[3] M. Al-Qurishi, M. Alrubaian, S. M. M. Rahman, A. Alhuthail, and M. Al-Rakhami, "A survey on fake account detection in social networks," *IEEE Access*, vol. 9, pp. 150876–150898, 2021.

[4] A. Gupta, "Fake account detection on social media using machine learning," *International Journal of Computer Applications*, vol. 178, no. 32, pp. 1–5, 2019.

[5] S. Kumar, *Social Media Analytics*, Springer, 2020.

[6] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," *Science*, vol. 359, no. 6380, pp. 1146–1151, 2018.

[7] 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.

[8] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," *ACM SIGKDD Explorations*, vol. 19, no. 1, pp. 22–36, 2017.

[9] R. Zafarani, M. A. Abbasi, and H. Liu, *Social Media Mining: An Introduction*, Cambridge University Press, 2014.

[10] J. Zhang and L. Luo, "Fake account detection using ensemble machine learning models," *IEEE Access*, vol. 10, pp. 45678–45689, 2022.