International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 2, Mar-Apr-2026, ISSN (Online): 2395-566X

Instagram Fake Account Detection Using Machine Learning

Himani Atul Khamkar [1] , Riddhika Dattaram Zolage [2] , Prof. Sanjay Eknath Gawli [3]
Department of Computer Engineering University of Mumbai
G. M. Vedak College, Tala, India

Abstract- — Social media platforms such as Instagram are widely used for communication, networking, and content sharing. How- ever, the rapid growth of these platforms has also led to a signifi- cant increase in fraudulent or fake accounts. These accounts are often involved in activities such as spamming, phishing, spreading misinformation, and manipulating engagement metrics. Due to the large number of users and the dynamic behavior of social media platforms, manual identification of fake accounts becomes difficult and inefficient. This research proposes a machine learning-based approach to detect fake Instagram accounts using profile-based features. Various attributes such as follower-following ratio, number of posts, engagement behavior, profile completeness, and other pro- file characteristics are analyzed. The Random Forest classification algorithm is used to distinguish between real and fake accounts. The performance of the model is evaluated using metrics such as accuracy, precision, recall, and F1-score. The experimental results demonstrate that the proposed approach can effectively identify fake accounts and contribute to improving the reliability and security of social media platforms.

Index Terms—Fake account detection, Instagram, machine learning, social networks.

I. INTRODUCTION

Social media platforms have become essential tools for communication, networking, and information sharing. Among these platforms, Instagram has gained immense popularity due to its visually oriented content and high user engagement. However, the rapid growth of Instagram has also resulted in the emergence of fake and fraudulent accounts. These accounts are often created to perform malicious activities such as spreading spam, phishing, promoting scams, or artificially increasing engagement metrics.

The large number of users and the dynamic nature of online behavior make manual detection of fake accounts extremely challenging. Fake accounts can imitate real user behavior, making it difficult to identify them using simple rule-based techniques. Therefore, automated detection systems are re- quired to analyze user patterns and detect suspicious activities effectively.

Machine learning techniques provide an efficient approach for detecting fake accounts by analyzing different profile attributes and behavioral patterns. In this study, a machine learning-based framework is proposed that analyzes features such as follower count, following count, number of posts, biography length, and presence of profile images. The Random Forest classification algorithm is used to classify accounts as either real or fake. The performance of the model is evaluated using accuracy, precision, recall, and F1-score. Experimental results demonstrate that the proposed approach can effectively detect fake accounts and improve user trust and platform integrity.

II. LITERATURE REVIEW

Fake and automated accounts have rapidly increased with the growing popularity of Instagram and other social network- ing platforms. These accounts are often used for spreading spam, manipulating public opinion, promoting fraudulent ac- tivities, and artificially increasing engagement. As a result, detecting fake profiles has become an important research area in social media analytics and cybersecurity.

Early research mainly focused on rule-based detection meth- ods using simple profile features such as number of followers, following count, posting frequency, and profile completeness. Although these methods were easy to implement, they were not very effective against advanced fake accounts that imitate genuine user behavior.

With the advancement of machine learning, researchers started applying supervised classification algorithms to iden- tify fake profiles. Algorithms such as Random Forest, Na¨ıve Bayes, Logistic Regression, and Support Vector Machines have been widely used for this task. These models are capable of learning complex patterns from labeled datasets and often provide higher detection accuracy compared to traditional rule- based approaches.

Several studies have also incorporated behavioral and content-based features to improve detection performance. These include analysis of posting patterns, hashtag usage, engagement rates, comment behavior, and activity intervals. Network-based features such as follower-following relation- ships have also been explored to detect suspicious account clusters.

Despite the promising performance of existing approaches, challenges such as evolving attacker strategies and dataset imbalance still remain. Therefore, developing efficient and adaptive detection models remains an important area of re-search. This study contributes to this field by applying a feature-based Random Forest classifier using real-world profile data collected from public datasets and real-time sources.

III. DATASET

The primary dataset used in this study is the Instagram Fake and Real Accounts Dataset obtained from Kaggle. This dataset contains labeled Instagram profiles categorized as either real or fake accounts. It includes several publicly available features such as profile picture presence, username length, full name attributes, biography length, privacy status, number of posts, follower count, and following count.

These attributes provide useful information for training ma-chine learning models to recognize patterns that differentiate genuine users from fraudulent accounts. The dataset is publicly available at the following link:

https://www.kaggle.com/datasets/rezaunderfit/instagram-fake-and-real-accounts-dataset

In addition to the Kaggle dataset, real-time profile infor- mation was collected using the Instaloader tool. Instaloader is a Python library that allows extraction of publicly available Instagram profile data such as follower count, number of posts, and profile metadata. This helps incorporate real-world data for testing and validating the proposed system.

IV. DATA PREPROCESSING

Real and fake Instagram profiles are included in the col- lected dataset. Before training the machine learning model, data preprocessing is performed to ensure data quality and consistency. Missing values are handled by removing incom-plete records or assigning default values when appropriate. Duplicate records are also removed to prevent biased training results.

Categorical features such as account privacy status and pres-ence of profile pictures are converted into numerical format so that they can be processed by the machine learning algorithm. Additionally, numerical features such as follower count and following count are normalized to maintain a consistent scale across the dataset.

The dataset is then divided into training and testing sets using an 80:20 split. The training set is used to train the Random Forest classifier, while the testing set is used to evaluate the model's performance.

V. FEATURE EXTRACTION

Feature extraction focuses on profile-based attributes that help distinguish real users from fake accounts. Important features extracted from the dataset include the number of followers, number of accounts followed, total number of posts, biography length, username length, presence of a profile picture, and account privacy status.

Fake accounts often have unusual patterns such as extremely high following counts, low follower counts, very few posts, or incomplete profile information. Therefore, analyzing these attributes helps the machine learning model detect suspicious patterns.

All extracted features are encoded and normalized before being provided as input to the Random Forest classifier. This ensures that the model can effectively analyze the data and identify meaningful patterns related to fraudulent account behavior.

VI. PROPOSED ARCHITECTURE

- The proposed system follows a feature-based supervised machine learning architecture using the Random Forest algo- rithm. The system consists of multiple stages beginning with data collection and ending with account classification.

- Data Collection: Data is collected from a Kaggle dataset and real-time profile information extracted using Instaloader.

- Data Preprocessing: The collected data is cleaned by handling missing values and removing duplicate entries. Categorical attributes are converted into numerical for-mat.

- Feature Extraction: Relevant profile-based features such as follower count, following count, posts, biogra- phy length, and profile picture presence are extracted.
- Model Training and Evaluation: A Random Forest classifier is trained using the processed dataset. The trained model is evaluated using accuracy, precision, recall, and F1-score.

Fig. 1. Proposed System Architecture

VII. MACHINE LEARNING ALGORITHM

The Random Forest algorithm is used in this research to classify Instagram accounts as real or fake. Random Forest is an ensemble learning technique that builds multiple decision trees and combines their predictions to improve classification accuracy.

Random Forest is particularly suitable for this problem because it can handle structured tabular data, reduce overfitting through ensemble learning, and identify the most important features influencing predictions.

Python libraries used in the implementation are:

- Pandas: Used for data manipulation, preprocessing, and handling datasets in DataFrame format.
- Scikit-learn: Provides machine learning algorithms in- cluding the Random Forest classifier.
- Instaloader: Used to collect publicly available Insta- gram profile data.
- Joblib: Used for saving and loading the trained machine learning model.
- Streamlit: Used to build a simple web interface to test the trained model.
- Venv: Used to create a virtual environment for managing project dependencies.

VIII. IMPLEMENTATION AND RESULTS

Python is used in a local development environment to implement the proposed fake Instagram account detection system. The dataset collected from Kaggle and Instaloader is first preprocessed to remove duplicates and handle missing values. Numerical features are normalized, and categorical attributes such as account privacy and profile image avail- ability are converted into numerical format to improve model performance.

Table I
Performance Evaluation Of Random Forest Classifier

Metric	Value
Accuracy	96.75%
Precision	96.98%
Recall	96.50%
F1 Score	96.74%

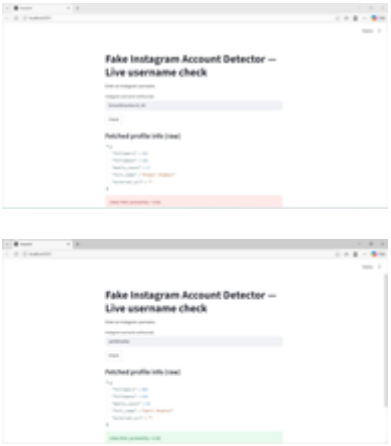

Fig. 2. Streamlit output showing fake and real account predictions

IX. CONCLUSION

This study presented a machine learning-based approach for detecting fake Instagram accounts using profile-based features. Data was collected from a publicly available Kaggle dataset and real-time profile information extracted using Instaloader. After preprocessing and feature extraction, a Random Forest classifier was trained to classify Instagram profiles as real or fake.

Experimental results demonstrate that the proposed system can effectively detect suspicious accounts using simple yet meaningful profile attributes. The trained model was also deployed using a Streamlit interface, enabling real-time testing and practical usability.

Overall, the proposed approach provides an efficient and scalable solution for detecting fake accounts on Instagram and can contribute to improving the security and reliability of social media platforms.

REFERENCES

1. B. K. Bhavya and K. Nikhitha, "Detecting Fake Accounts on Social Me- dia – Instagram," Bachelor of Engineering Project Report, Sathyabama Institute of Science and Technology, Chennai, India, Apr. 2023.
2. N. Kadam and S. K. Sharma, "Social Media Fake Profile Detection Using Data Mining Technique," Journal of Advances in Information Technology, vol. 13, no. 5, pp. 518–523, Oct. 2022.
3. J. Singh and M. Z. Khan, "Detection of Fake Profile in Social Media," Journal of Emerging Technologies and Innovative Research (JETIR), vol. 6, no. 6, pp. 38–41, Jun. 2019.
4. P. Deshmukh et al., "Fake Social Media Profile Detection," International Journal of Scientific & Technical Education, vol. 47, pp. 190–197, Oct. 2024.
5. S. Bhambar et al., "Detecting Fake Accounts on Social Media Using Neural Network," IRJMETS, vol. 4, no. 5, pp. 3450–3455, May 2022.