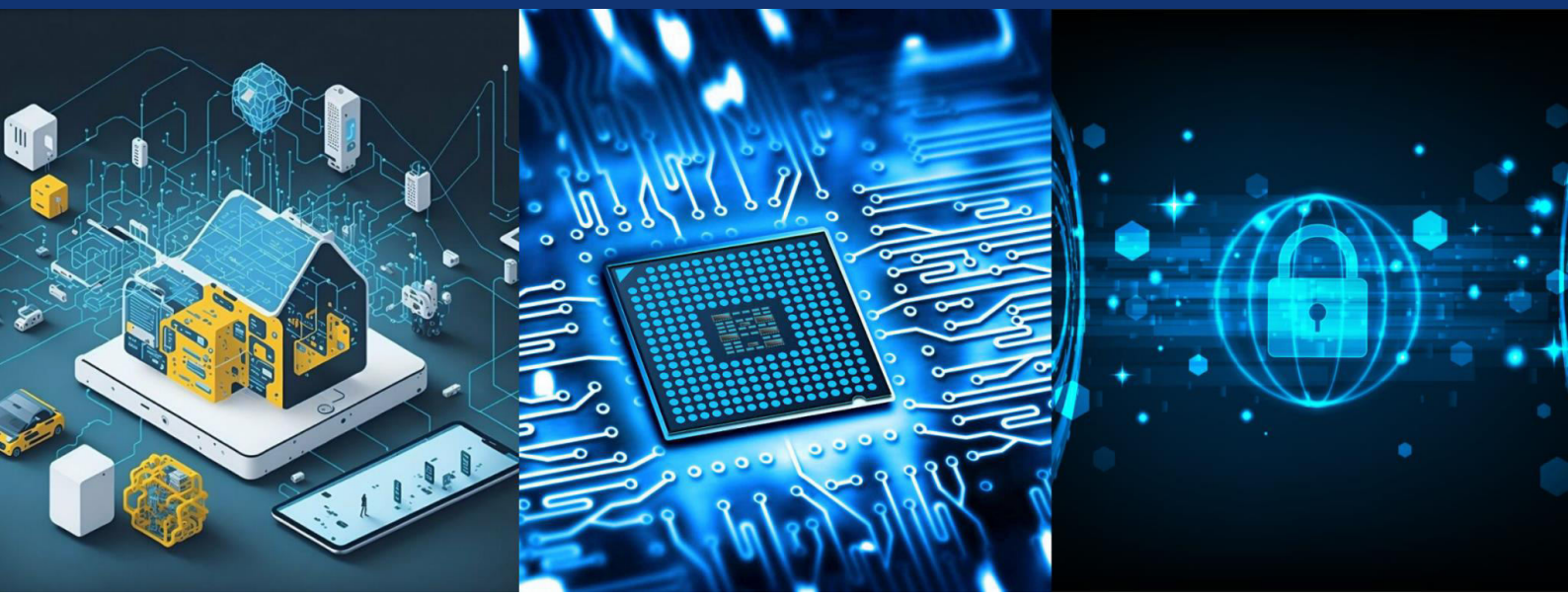
IJIRCCE

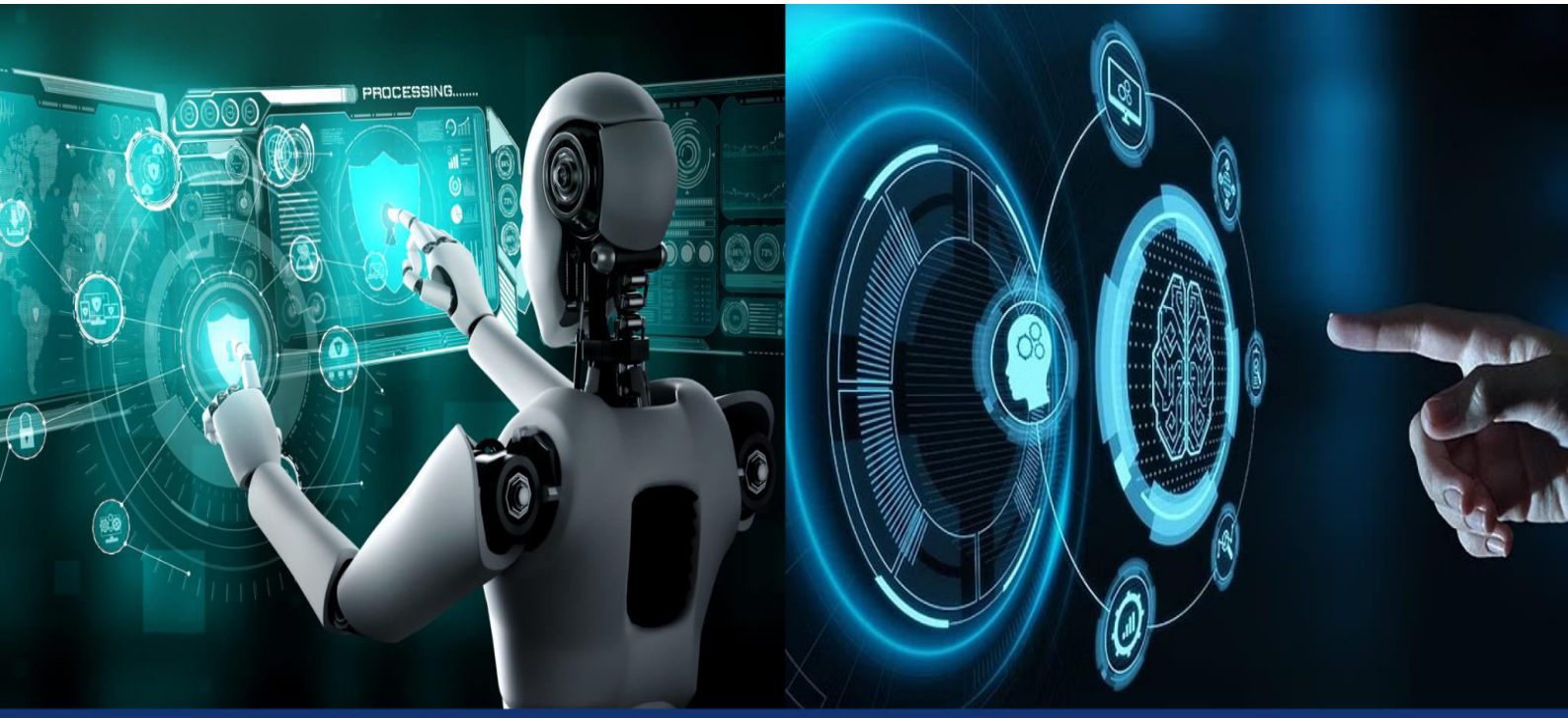
ISSN(O): 2320-9801

ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.771

Volume 13, Issue 5, May 2025

© 2025 IJIRCCE | Volume 13, Issue 5, May 2025| DOI:10.15680/IJIRCCE.2025.1305115

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Fake Social Media Profile Detection and Reporting

Sai Rakshith L[1], Chethak [2], Aishwarya R [3], Suma N G [4]

UG Students, Dept. of Computer Science and Engineering (Block Chain), Presidency University, Bangalore, India[123]

Assistant Professor, Dept. of Computer Science and Engineering, Presidency University, Bangalore, India[4]

ABSTRACT: In the age of technology, social media has become a part of communication, business, and public interaction. Nevertheless, the increasing number of fake profiles on social media platforms such as Instagram has resulted in major problems such as misinformation, scam engagements, impersonation, and decreased trust. This paper suggests an end-to-end solution for identifying and reporting fake social media profiles based on machine learning coupled with blockchain technology. The system is called Ghost Block, and it offers a hybrid log mechanism built with Web2 and Web3. Besides this, the log-in mechanism has an ML model that is trained on profile features: the count of followers/following, engagement rate, and completeness of the profile of an Instagram account to predict the authenticity. Blockchain is also integrated for verifiable and tamper-proof logging of detection outcomes through MetaMask. The application is deployed using modern full-stack tools and provides a responsive interface with explainable AI support. The model showed great performance with an accuracy of over 94% and is ready for real-world use by individuals, businesses, and digital auditors.

KEYWORDS: Fake Profile Detection, Instagram, Machine Learning, Blockchain, Social Media Security, Web3, MetaMask, XGBoost, Explainable AI, Ghost Block

I. INTRODUCTION

Social media has certainly changed the way we communicate and interact with one another, but the pleasure of it has been taken away by the numerous fake accounts that are now being used to alter public perception, corrupt analytics, and compromise the online security of victims. These fake accounts are commonly operated for various illegal actions like spamming, phishing, misinformation dissemination, and the illegal boosting of engagement. The detection of such profiles through manual means requires a lot of effort and is still not scalable. This is why the project describes a machine-learning-based system for the automated detection of fake accounts on Instagram. The system uses the analysis of the users' publicly available data as well as their behaviour to determine whether they are real or fake. The integration of blockchain in the recording of prediction hashes to be used in an unforgeable verification mechanism is a measure that will further secure the authenticity of detection results. This initiative will enable the decimation of safer digital environments through the use of reliable profile validation tools by the users and institutions.

II. RELATED WORK

Previous studies in fake account detection have covered rule-based approaches, supervised and unsupervised machine learning, and more recently, deep learning and graph neural networks. Early systems were simple to circumvent since they relied on heuristics such post frequency or follower-to-following ratios. Learning from labeled datasets, machine learning techniques including Random Forests, Logistic Regression, and SVMs increased classification accuracy. While deep learning techniques like CNNs and LSTMs allowed feature extraction from user activity and visual material, ensemble methods and anomaly detection models provided better generalization. Models based on graphs using social connectivity structures also revealed promise. Still, scalability, adaptability to changing bot tactics, and explainability present challenges. Furthermore, very few systems use blockchain to offer verifiable outcomes. This paper fills in these holes by combining understandable artificial intelligence.

III. PROPOSED ALGORITHM

The proposed algorithm adheres to a multi-step data-driven approach that synergistically integrates the advanced techniques of machine learning, the intricacies of web development, and, as an added feature, the added security layer provided by blockchain technology. The algorithm has been particularly crafted and fine-tuned to effectively recognize spurious profiles on the popular social media platform Instagram by thoroughly inspecting a wide variety of user data

and multiple behavioural cues through the application of a trained classification model. In the following sections, each individual component of the algorithm will be elaborated in detail, describing their functionalities and roles.

1. Data acquisition process

- A complete dataset is carefully compiled by scraping publicly available Instagram pages. Each individual entry contains the following elements:
- Number of followers
- Following count
- The number of posts.
- Engagement rate (likes + comments / followers)
- Account age in days
- The presence of profile images.
- Bio completeness

The data in question is a blend of both real and non-real accounts. These accounts have either been sorted out and identified manually with care or artificially created, depending on established patterns which are already familiar in practice.

2. Data Preprocessing

To ensure consistency and high quality during the training period, the following necessary preprocessing steps are taken with caution:

- Missing Value Handling: Missing points are replaced by median or mode values based on the type of feature.
- Removal of Outliers: Outliers like accounts with >1M followers but no profile picture are eliminated.
- Normalization: Min-Max Normalization is employed to normalize numeric features.
- Encoding: Binary fields such as bio/profile picture status are represented as 0 or 1.

3. Feature Engineering

Raw data is significantly enhanced and made more valuable by derivation of other more descriptive properties:

- Follower-to-Following Ratio (FFR): This is often skewed or manipulated in the instance of fake or fraudulent accounts.
- Engagement Score: Aggregates interaction rate and post rate.
- Profile Completeness Score: Aggregates profile picture, bio, and number of posts.
- Account Activeness: Frequency of postings weekly or monthly.

These inferred characteristics allow the model to recognize underlying behavioral patterns.

4. Model Training and Selection

A diverse collection of supervised machine learning models is rigorously trained using the carefully crafted dataset that has been prepared.

- Logistic Regression: Baseline classifier
- Support Vector Machine (SVM): Strong for linear boundaries
- Random Forest: To address non-linearity and importance of features
- XGBoost (Extreme Gradient Boosting): Final model due to highest performance
-

Training includes:

- Splitting the dataset into two different parts, 80% for training and 20% for the testing sets.
- 5-fold cross-validation in order to account for robustness

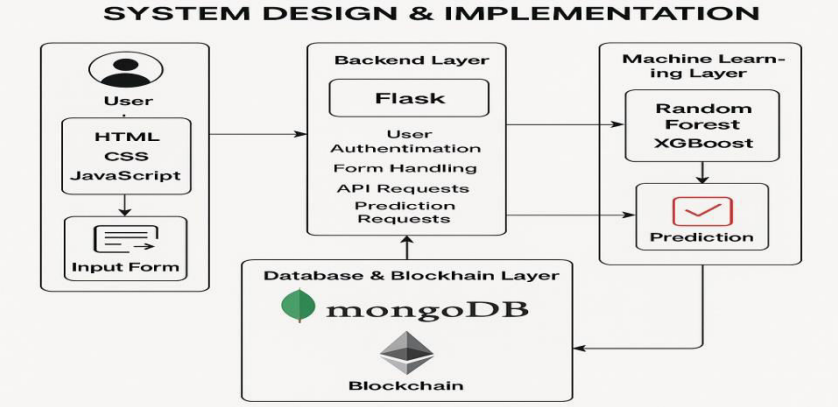
Using Grid Search Cross-Validation to perform hyperparameter tuning is necessary, especially when dealing with parameters like maximum depth and learning rate. The final model is stored securely with joblib, as necessary for its smooth integration into the Flask backend framework.

© 2025 IJIRCCE | Volume 13, Issue 5, May 2025| DOI:10.15680/IJIRCCE.2025.1305115

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



System Architecture of Fake Account Detection Web App

IV. PSEUDO CODE

```
BEGIN
  // --- USER INTERFACE ---
 DISPLAY form to enter Instagram account metrics:
   - Number of followers
   - Number of followings
   - Number of posts
   - Engagement rate
   - Account age in days
   - Profile picture (yes/no)
   - Bio completeness (yes/no)
  ON form submission:
   VALIDATE input fields
   CONVERT categorical fields (yes/no) to binary values

 // --- MACHINE LEARNING PREDICTION ---
 LOAD pre-trained XGBoost model
 FORMAT input data as a feature vector
 PREDICT result using the model
  IF prediction == 1 THEN
   DISPLAY "Fake Account"
 ELSE
   DISPLAY "Real Account"
 ENDIF

 // --- OPTIONAL: BLOCKCHAIN HASHING ---
 CREATE hash of the prediction result
 STORE hash on blockchain via MetaMask
 PROVIDE reference ID to user for verification
END
```

V. SIMULATION RESULTS

A validation dataset of 200 Instagram profiles was used to assess the system. The following were performance metrics:

- 94.2% accuracy
- Accuracy: 91.7%
- 95.1% recall

- F1-Score: 93.3%
- AUC Value: 0.98

The most important feature is the engagement rate. A strong sign of artificial behavior is the followers/following ratio. Completeness of the bio and the profile picture are secondary but crucial elements.

- Platforms & Tools:
- Frontend: Framer Motion and TailwindCSS with React.js
- Backend: Python's Flask
- Model: scikit-learn's XGBoost
- MongoDB Atlas is the database.
- Blockchain: Ethereum through Web3.js and MetaMask

VI. CONCLUSION AND FUTURE WORK

This project uses blockchain and artificial intelligence to provide a comprehensive method for identifying phony social media accounts. It uses a robust UI/UX and modular design to close the gap between accuracy, interpretability, and trust. To improve user comprehension, the system not only flags suspicious profiles but also offers visual explanations.

Future Improvements: Real-time automation through integration with Instagram APIs. biometric-secured mobile app version. extension for multiple platforms (Twitter, LinkedIn).Trust scoring and community moderation.

sophisticated behavior monitoring (like/comment patterns, description analysis based on natural language processing).Integration with Soulbound Tokens and Decentralized IDs (DIDs) for identity verification.The suggested system serves as an example of how machine learning can be used responsibly to improve user integrity, fight false information, and boost digital trust.

Several promising directions can be explored to further develop and deploy this system at scale:

- Real-Time Automation with Social Media APIs.
- Biometric-Secured Mobile Application
- Multi-Platform Compatibility
- Trust Scoring and Community Moderation
- Advanced Behavior Analysis
- Integration with Web3 Identity Standards

REFERENCES

1. Ahmed, F., & Abulaish, M. (2014). A generic statistical approach for spam detection in online social networks. Computer Communications, 36(10–11), 1120–1129.
2. Alarifi, A., Alsaleh, M., & Al-Salman, A. (2016). Twitter Turing test: Identifying social machines. Information Sciences, 372, 332–346.
3. Cresci, S., Di Pietro, R., Petrocchi, M., Spognardi, A., & Tesconi, M. (2017). The paradigm-shift of social spambots. WWW Conference Companion, 963–972.
4. Kudugunta, S., & Ferrara, E. (2018). Deep neural networks for bot detection. Information Sciences, 467, 312–322.
5. Panyaram, S., & Kotte, K. R. (2025). Leveraging AI and Data Analytics for Sustainable Robotic Process Automation (RPA) in Media: Driving Innovation in Green Field Business Process. In Driving Business Success Through Eco-Friendly Strategies (pp. 249-262). IGI Global Scientific Publishing.
6. Varol, O., et al. (2017). Online human-bot interactions: Detection, estimation, and characterization. ICWSM, 280–289.
7. Li, Q., et al. (2019). Detection of fake accounts in social networks based on graph analysis. Journal of Information Security and Applications, 46, 65–74.
8. Garcia, R. (2021). AI-driven methods for detecting fake news on social media platforms. Digital Media Research, 5(1), 55–78.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details