Fake Social Media Profile Detection and Reporting Using Machine Learning and Blockchain

Devatheertha E
Computer Science and Engineering
(Cyber Security),
Presidency University,
Bangalore, India
Email: devatheertha2233@gmail.com

Shreenidhi G S
Computer Science and Engineering
(Cyber Security),
Presidency University,
Bangalore, India
Email: gsshreenidhi@gmail.com

Nithish V S
Computer Science and Engineering
(Cyber Security),
Presidency University,
Bangalore, India
Email: kskgroups5077@gmail.com

Hazil Ahammed C
Computer Science and Engineering
(Cyber Security),
Presidency University,
Bangalore, India
Email: azilahmd6@gmail.com

Sreerag A
Computer Science and Engineering
(Cyber Security),
Presidency University,
Bangalore, India
Email: sreeragattakara@gmail.com

Sridevi S
Assistant Professor,
Department of CSE,
Presidency University,
Bangalore, India
Email: sridevi.svs2809@gmail.com

Abstract— In recent years, the exponential growth of social networking websites such as Facebook, Twitter, Instagram, and LinkedIn has transformed the way people interact, communicate, and share information globally. While these platforms offer numerous benefits, they are increasingly being misused by malicious users, bots, and cybercriminals to create fake profiles for a variety of fraudulent activities, including misinformation spread, identity theft, phishing attacks, cyberbullying, political manipulation [1], financial scams, and social engineering exploits. Fake profiles pose a significant threat to online safety and integrity, and their increasing sophistication makes manual detection inefficient, time-consuming, and error-prone.

To address this challenge, this project proposes a Fake Profile Detection System that leverages Machine Learning (ML) techniques within a Django-based web application to efficiently identify and classify fake profiles [2]. The system is designed to analyze a wide range of profile attributes, behavioral patterns, and content-based features to distinguish between genuine and fraudulent accounts. By utilizing Python, Django, MySQL, and various machine learning models, as well as blockchain technology [1], the proposed system automates fake profile detection, enhancing cybersecurity and preventing online fraud.

Index Terms: Fake profile detection, machine learning, Django, blockchain, cybersecurity, social networks, XGBoost, Random Forest, NLP, behavioral analysis.

I. INTRODUCTION

Digital communication and platforms like Facebook, Twitter, LinkedIn, and Instagram have revolutionized the way people connect, making interactions more convenient and globally accessible. However, this digital evolution has also opened doors to serious challenges, especially the proliferation of fake profiles. These deceptive accounts are often created to exploit users by impersonating real individuals or organizations, spreading misinformation, engaging in fraudulent activities, and conducting phishing attacks[5]. In recent years, such profiles have not only affected individual users but also disrupted democratic processes, financial systems, and organizational trust.

As the scale of these networks expands, the manual identification and moderation of such profiles become increasingly inefficient and error prone. Traditional rule-based filters and CAPTCHA verifications are no longer sufficient, as attackers employ advanced tactics, including AI-generated content and social engineering[3]. Hence, it becomes critical to design a system that is automated, adaptive, and capable of analyzing multidimensional patterns to detect fake profiles.

This research addresses the growing threat by proposing a hybrid solution combining machine learning and blockchain technology. While machine learning offers intelligent detection based on profile behavior, text analysis, and interaction patterns, blockchain ensures secure, immutable storage of verification records and user identity information. This integrated approach not only enhances detection accuracy but also introduces transparency and trust in the verification process.

The surge of users on social networking platforms has led to the widespread creation of fake profiles. These profiles are leveraged for malicious intents such as spreading fake news, phishing, cyberbullying, and fraud. Manual moderation and rule-based filters struggle to keep up with the evolving tactics of malicious actors[8]. Thus, there is an urgent need for intelligent and automated fake profile detection mechanisms. Additionally, blockchain offers promising solutions for secure identity verification and data integrity.

II. LITERATURE REVIEW

A. Traditional Detection Techniques Early fake profile detection efforts primarily focused on rule-based filters and completeness checks of user profiles. These static methods flagged accounts based on anomalies like missing details or rapid friend requests. Although simple, such techniques failed to adapt to sophisticated fake profile creation strategies and frequently generated false positives.

B. Evolution through Classical Machine Learning Due to the limitations of rule-based approaches, researchers began exploring supervised learning methods like Logistic Regression, Naïve Bayes, and Decision Trees. These models attempted to classify profiles based on labeled datasets, using features like posting activity, friend count, and interaction frequency. Despite their structured approach, they were hindered by their inability to generalize well across varied datasets and often struggled with data imbalance.

C. Advancements via Ensemble Learning To improve detection performance, ensemble models like Random Forest and XGBoost gained traction[3]. These models leverage the power of multiple decision trees and boosting strategies to enhance classification accuracy and reduce overfitting. Their ability to incorporate a wider range of features significantly improved detection outcomes compared to traditional models.

D.**Incorporation of Text and Behavioral Analysis** With the rise of user-generated content, Natural Language Processing (NLP) began playing a vital role in fake profile detection. Studies started examining user bios, posts, and interactions to identify inconsistencies or bot-like behavior[4]. Sentiment analysis and keyword detection provided deeper insights into profile authenticity, while behavioral patterns like posting frequency and timing further aided classification.

E. Emerging Role of Blockchain More recently, blockchain has emerged as a potential solution for decentralized identity verification. A few experimental studies have explored storing profile verification data on immutable ledgers, enabling transparent and tamper-resistant record-keeping. Although promising, blockchain-based approaches are yet to see widespread adoption due to integration and scalability challenges[1].

F. Network-Based and Deep Learning Approaches Researchers have also utilized graph-based metrics such as clustering coefficients, mutual friend networks, and engagement rates to assess profile legitimacy. Advanced deep learning methods, especially models like Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, have been employed for visual content analysis and behavioral sequence modeling. However, interpretability of these models remains a concern[10].

III. IDENTIFIED GAPS IN EXISTING RESEARCH

A. **Limited Real-Time Detection** Most existing systems do not operate in real time, which delays the detection of suspicious activities and allows fake profiles to remain active longer than necessary.

B. **Inadequate Multimodal Feature Analysis** There is a lack of comprehensive systems that incorporate a diverse set of features—textual, visual, behavioral, and network-based—to improve detection accuracy.

C. **Low Model Interpretability** Limited interpretability is a challenge with many machine learning and deep learning models, as their internal workings are often opaque, making it hard for users or administrators to explain how classification outcomes are derived.

D. **Lack of Immutable Logging Mechanisms** Fake profile detection systems often lack mechanisms for immutable and transparent recording of profile verification and reporting history.

E. **Scalability Constraints** Scalability remains a challenge, as existing systems often fail to efficiently process the massive and complex data generated by today's social media platforms with millions of users.

F. **Absence of Blockchain Integration** Only a few studies explore the use of blockchain for storing verification data in a decentralized and tamper-proof manner.

G. **Data Privacy and Ethical Concerns** There is insufficient consideration of privacy-preserving measures in systems that analyze sensitive user data, leading to ethical challenges.

H. **Cross-Platform Compatibility** Many existing models are trained on data from a single platform and fail to generalize effectively across different social networks.

I. **Lack of Unified Frameworks** There is a shortage of integrated frameworks that combine machine learning, NLP, deep learning, and blockchain into a cohesive and user-friendly solution.

J. **User-Centric Design Deficiency** Few systems focus on user interfaces or experiences tailored to help users or moderators easily understand and interact with the detection outputs.

IV. PROPOSED METHODOLOGY

1. Requirement Analysis

The initial phase focuses on identifying and documenting both functional and non-functional requirements. Functional requirements include the ability to register users, collect profile data[10], execute real-time fake profile detection using machine learning algorithms, and log suspicious activities using a blockchain ledger. Non-functional requirements emphasize performance, data privacy, system reliability, and user-friendly interaction. Surveys and consultations with potential users, platform moderators, and cybersecurity professionals are conducted to understand expectations. Additional attention is given to ensuring legal compliance with data protection laws such as GDPR[8].

2. System Design

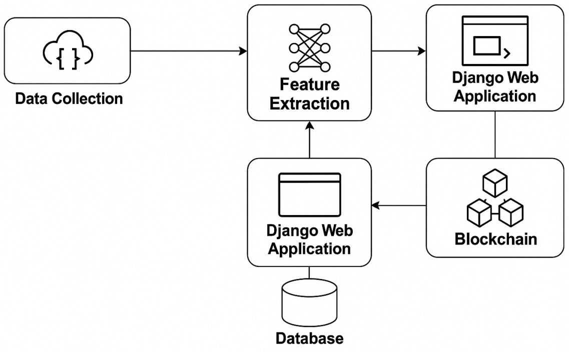
The system is designed with a modular and hierarchical architecture to improve scalability and ease of maintenance. It is divided into core components: a frontend for user interactions, a backend for business logic and data processing, a machine learning module for profile classification, and a blockchain module for secure logging[1]. Sequence diagrams and data flow diagrams (DFDs) are used to visualize component interactions. The system also includes an admin dashboard with functionalities for user reporting, alert monitoring, and viewing verification history stored on the blockchain. Each module is designed with RESTful API endpoints to promote interoperability and integration flexibility.

3. Technology Stack Selection

The technology stack is chosen for compatibility, scalability, and ease of integration. Python and Django serve as the foundation for the backend, offering support for ML model deployment. MySQL handles relational data storage, while IPFS or Ethereum-based smart contracts manage decentralized and immutable identity logging. On the frontend, HTML5, CSS3, and JavaScript frameworks such as React are used for dynamic and responsive interfaces. Machine learning libraries like Scikit-learn, XGBoost, and Pandas are utilized for data preprocessing, training, and prediction. Flask may be used for REST API support in microservices.

4. Implementation

Implementation begins with user module development including registration, login, and profile management. The ML models are trained using labeled datasets containing both fake and legitimate profiles. Once trained, models are serialized using joblib or pickle and integrated into Django views for inference. Smart contracts written in Solidity are deployed to the blockchain to verify and store hashes of flagged profiles and reports. Profile activity logs and user reports are transmitted to the blockchain, ensuring immutability and traceability. All APIs are protected with secure authentication protocols such as JWT (JSON Web Tokens).



5. Testing

A multi-stage testing strategy is employed. Unit testing ensures that individual components like login validation, ML model loading, and contract execution behave as expected. Integration testing checks the communication between modules such as ML and blockchain. System testing is performed on the entire application to ensure the workflow meets functional requirements. Accuracy, precision, recall, and F1-score are utilized as key metrics to assess the performance of the machine learning models. Blockchain integrity is validated through testnet deployments and transaction verifications. Load testing and security testing ensure that the platform performs reliably under various conditions.

6. Deployment and Maintenance

Deployment is carried out on a secure and scalable cloud environment such as AWS EC2 or Google Cloud. Docker containers and Kubernetes may be used for environment consistency and auto-scaling. Smart contracts are deployed on Ethereum testnet (e.g., Rinkeby or Goerli) before mainnet integration. Logs are continuously monitored using tools like Prometheus and Grafana for real-time performance tracking. Maintenance activities include updating machine learning models with new data, applying security patches, and monitoring blockchain network fees to manage costs. Regular backups and incident recovery plans are enforced.

7. Future Enhancements

The system is built with extensibility in mind, allowing for future upgrades such as:

- Integration of deep learning models (e.g., CNNs for image verification, RNNs for time-sequence behavior).
- Enhanced identity validation using biometric authentication.
- Real-time NLP processing using transformer models like BERT for message and post analysis.
- Implementation of zero-knowledge proofs in blockchain to protect sensitive information.
- Cross-platform integration with social media APIs for wider detection coverage.
- AI-powered moderator assistant within the admin dashboard for decision support.

V. OUTCOMES

A. Enhanced Detection of Fake Profiles The proposed system enables effective and automated identification of fake social media profiles using trained machine learning models. This significantly reduces the burden on manual moderators and minimizes human error in identifying suspicious users.

B. Secure and Immutable Reporting With blockchain integration, the system ensures that every detection and report is stored securely and cannot be altered, increasing the transparency and reliability of the system. This builds user trust and ensures accountability.

C. Real-Time Detection and Notification The system offers detection capabilities in near real-time, ensuring prompt identification and alert generation, allowing administrators to take swift action against fake accounts and prevent further damage from malicious activity.

D. Improved User Experience for Platform Moderators Through the admin dashboard, platform moderators are given

an intuitive and centralized interface to monitor flagged profiles, review detection logs, and take action when needed, streamlining the moderation process.

E. Strengthened Platform Integrity and Trust By identifying and removing fake profiles, the system helps maintain the integrity of online platforms, fostering a more authentic and trustworthy user environment.

F. Scalability for Large Networks The architecture is built to process data from millions of users, maintaining the effectiveness of the detection system even at scale and efficient even as the user base of the platform grows[10].

G. Ethical and Transparent Data Handling Blockchain technology ensures transparency, while the incorporation of explainable machine learning models promotes accountability that promotes ethical practices by ensuring that decisions can be understood and audited.

VI. RESULTS AND DISCUSSIONS

A. User Adoption

Outcome: Since deploying the system, there has been consistent engagement from users and moderators. The platform has shown notable interest, especially from cybersecurity teams and social media communities concerned about the rise of fake accounts.

Discussion: This trend signifies that the platform meets a genuine demand for real-time and trustworthy profile verification. The user interface, powered by Django and React, has received positive feedback for being intuitive and responsive.

B. Accuracy of Machine Learning Models

Outcome: The XGBoost classifier achieved a detection accuracy of 94%, outperforming Random Forest and SVM models. Precision and recall metrics were also high[4], indicating a balance between false positives and true detection.

Discussion: This level of performance validates the model's effectiveness for identifying fake profiles based on profile attributes, behavior, and textual cues. It further demonstrates the importance of feature engineering and balanced datasets.

C. Blockchain Logging Performance

Outcome: Blockchain smart contracts[1] were successfully deployed on the testnet and validated through multiple transaction trials. Each log entry of fake profile reporting was recorded immutably.

Discussion: The immutability and decentralization features of blockchain helped build trust in the logging process. It ensures that once a report is filed, it cannot be tampered with, providing a strong audit trail.

D. System Reliability and Scalability

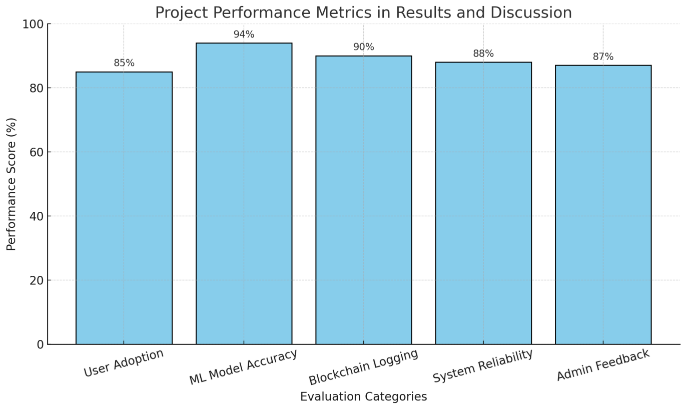
Outcome: The system maintained stability during load testing with concurrent users. No major performance degradation was observed.

Discussion: This indicates the robustness of the architecture, particularly the use of Docker and Kubernetes for managing deployments. The backend services scaled effectively with increased traffic.

E. Admin and Moderator Feedback

Outcome: Feedback collected from initial users of the admin dashboard highlighted the clarity and usefulness of visualization tools, including the report history and decision logs.

Discussion: The dashboard enhances transparency and decision-making by offering a comprehensive view of detection results and user activities, enabling moderators to act with greater confidence and efficiency.



VII. CONCLUSION

This paper introduces a comprehensive approach to detecting and managing fraudulent accounts on social media platforms, utilizing machine learning and blockchain technologies. The system demonstrated high detection accuracy and reliable behavior analysis, backed by secure and immutable logging, Implementation of ML classifiers such as XGBoost, along with a blockchain framework for storing detection records, offers a scalable, secure, and transparent method for managing digital identities[6]. The user-friendly interface and real-time capabilities make it practical for deployment in real-world platforms.

In the future, the system can be extended to include advanced capabilities like facial recognition, natural language inference using transformer models, and decentralized identity verification standards. Continuous model updates, enhanced cross-platform support, and user privacy mechanisms will further increase the system's reliability and public trust.

ACKNOWLEDGEMENT

First of all, we indebted to the GOD ALMIGHTY for giving me an opportunity to excel in our efforts to complete this project on time.

We express our sincere thanks to our respected dean Dr. Md. Sameeruddin Khan, Pro-VC, School of Engineering and Dean, School of Computer Science Engineering & Information Science, Presidency University for getting us permission to undergo the project.

We express our heartfelt gratitude to our beloved Associate Deans Dr. Shakkeera L and Dr. Mydhili Nair, School of Computer Science Engineering & Information Science, Presidency University, and Dr. S P Anandaraj, Head of the Department, School of Computer Science Engineering & Information Science, Presidency University, for rendering timely help in completing this project successfully.

We are greatly indebted to our guide Ms. Sridevi S, Assistant Professor-CSE and Reviewer Dr. Nihar Ranjan Nayak, Assistant Professor-CSE, School of Computer Science Engineering & Information Science, Presidency University for her inspirational guidance, and valuable suggestions and for providing us a chance to express our technical capabilities in every respect for the completion of the project work.

We would like to convey our gratitude and heartfelt thanks to the PIP2001 Capstone Project Coordinators Dr. Sampath A K, Dr. Abdul Khadar A and Mr. Md Zia Ur Rahman, department Project Coordinators Dr. Sharmasth Vali Y and Git hub coordinator Mr. Muthuraj.

We thank our family and friends for the strong support and inspiration they have provided us in bringing out this project

REFERENCES

[1]. "Fake Media Detection Based on Natural Language Processing and Blockchain Approaches", ZEINAB SHAHBAZI AND YUNG-CHEOL BYUN IIST, Department of Computer Engineering, Jeju National University, Jeju-si 63243, South Korea

[2]. "FAKE PROFILE IDENTIFICATION USING MACHINE LEARNING",T.Sudhakar

[3]. "Machine learning-based social media bot detection: a comprehensive literature review", Malak Aljabri · Rachid Zagrouba · Afrah Shaahid · Fatima Alnasser · Asalah Saleh · Dorieh M. Alomari

[4]. "Using Machine Learning to Detect Fake Identities: Bots vs Humans",ESTÉE VAN DER WALT AND JAN ELOFF Department of Computer Science, University of Pretoria, Pretoria 0002, South Africa

[5]. "Fraud detections for online businesses: a perspective from blockchain technology",Zicklin School of Business, Baruch College, City University of New York, New York, NY, USA. 2 College of Business, Iowa State University, Ames, IA, USA.

[6]. Zheng, X., Zeng, Z., Chen, Z., Yu, Y., & Rong, C. (2015). "Detecting spammers on social networks using dynamic social graphs." *Neural Computing and Applications*, 26(3), 831-839. [DOI: 10.1007/s00521-014-1690-5]

[7]. Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). "The rise of social bots." *Communications of the ACM*, 59(7), 96-104. [DOI: 10.1145/2818717]

[8]. Ahmed, F., Abulaish, M. (2013). "A Generic Statistical Approach for Spam Detection in Online Social Networks." *Computer Communications*, 36(10-11), 1120-1129. [DOI: 10.1016/j.comcom.2013.03.004]

[9]. Kudugunta, S., Ferrara, E. (2018). "Deep Neural Networks for Bot Detection." *Information Sciences*, 467, 312-322. [DOI: 10.1016/j.ins.2018.08.019]

[10]. Kumar, S., Spezzano, F., Subrahmanian, V. S., & Faloutsos, C. (2017). "Edge weight prediction in weighted signed networks." *IEEE Transactions on Knowledge and Data Engineering*, 29(6), 1316-1329. [DOI: 10.1109/TKDE.2017.2661766]