Fake Social Media Profile detection and reporting

Chigulla Nagapavan[1], Majjar Ravi Shankar Prasad[2], S.Nagesh[3], P.Akshaya Kumar[4], Arshiya Lubna[5], Swapna[6]

[1,2,3,4,5,6]*Dept. of Computer Science Engineering, Presidency University Bengaluru, India*

Abstract - Social media serves as one of the primary sources for communication, networking and information sharing. The proliferation of bogus profiles also poses legitimate threats like impersonation, disinformation, cyber harassment, and cons. Existing techniques are based on detection frameworks using centralized systems which could be compromised or non-transparent. In this scope, we present a Blockchain and Cybersecurity Integrated Social Media Profile Detection and Reporting System. Our system guarantees trustless, decentralized, and tamper-proof profile verification using blockchain technology. This particular solution improves trust, minimizes the amount of social media bots, and strengthens social media network security.

INTRODUCTION

The goal of this project, "Fake Social Media Profile Detection and Reporting Using Cybersecurity and Blockchain" is to create an immutable mechanism for detecting and reporting fake social media profiles. The problem is solved by integrating behavior analytics, anomaly detection, machine learning, and other security controls to correct inconsistent profiles and active profiles. Upon detection, the malicious profiles are recorded on the blockchain, thus making the reports immutable and transparent.

The use of blockchain technology provides an impeccable record of reported fake profiles, which all from users to authorities can rely on, having the ability to trust the information presented and act on it. This not only enhances the efficiency and accountability of identifying fake profiles, but also protects the genuine users and upholds the integrity of online communities. Social networking sites have become a way of life during the information era, with the potential of networking, communication, and information sharing. But this widespread use has also resulted in a severe problem—the creation and sharing of fake social networking sites. These fake profiles are usually made for malicious purposes such as the spread of misinformation, operating scams, phishing, or impersonation.

LITERATURE REVIEW

The development of social media websites has fundamentally changed the communication dynamics, offering unprecedented opportunities for access to information and interaction with people. The proliferation of false accounts on social media websites has, however, introduced unprecedented dangers, including identity theft, the dissemination of misinformation, internet fraud, and cyberbullying. Various studies and technologies have been suggested to address this problem, including fake profile detection, cybersecurity practices, and blockchain-based reporting systems. Several researchers have looked into machine learning and artificial intelligence solutions for identifying spurious profiles. Research works such as Stringhini et al. (2010) presented behavioral models for analyzing genuine and false users from posting habits, friend requests, and shared content. Feature-based classification is used by other techniques where decision trees, support vector machines (SVM), and random forest models classify profiles as false or authentic.

More contemporary research has used deep learning approaches, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), to classify user actions over time. These models enhance detection precision but tend to be non-scalable across platforms and languages.

Cybersecurity models are focused on preventing spurious accounts through user authentication, IP monitoring, and multi-factor authentication. Some concentrate on anomaly detection through login pattern monitoring, suspicious behavior, and geolocation discrepancies. These methods are, however, reactive and internal to platform usage with little transparency to third parties or users.

Although detection technologies have improved, detection-reporting integration remains in its infancy. Most systems take care of flagging only obvious fake accounts and fail to produce secure and accountable reporting. Top platforms also have a lack of transparency, triggering distrust among users towards platform moves. Blockchain integration can fill this gap by bringing decentralization to trust and permitting public verification of reported fake accounts.

The body of literature shows a solid basis in both fake profile identification and blockchain for trusted record-keeping. There is, however, a major gap in bringing these two together to create a system that not only identifies but also reports and traces fake social media profiles securely. This project seeks to bridge that gap by leveraging advanced cybersecurity methods with the trust and transparency of blockchain technology

Blockchain technology has also become a potential solution in cybersecurity because of its immutability, decentralization, and transparency. Blockchain-based models for digital identity authentication, secure logging, and trusted information exchange have been suggested by researchers. Blockstack and Civic are some projects that illustrate the promise of blockchain in user identity management.

PROPOSED METHOD

The system intended for detecting and reporting spurious social media profiles is developed through the integration of cybersecurity measures and blockchain technology so that detection and reporting are always accurate and tamper-proof. The initial task is to acquire user information from social media via available APIs or pre-compiled datasets. It encompasses user profile information like the account name, profile image, followers count, frequency of posting, and the type of posted content. Once gathered, features of concern are pulled that usually point towards suspicious activity, such as incomplete profiles, repeated content, drastic increases in activity, and low engagement rates.

In the second scenario, the gathered data is input into machine learning models such as Support Vector Machines, Random Forest, or Neural Networks. These machine learning models are trained on labeled data to identify the difference between genuine and spurious profiles. To have greater precision and fewer false positives, the system also employs cybersecurity-oriented behavior analysis. Analysis is performed by monitoring anomalies such as repeated login from geographically remote systems, suspicious IP addresses, and patterns of user behavior that typically signify automated or malicious behavior.

After a profile is determined to be a fake profile, a detailed report is generated with ample evidence like timestamps, unusual patterns, and profile metadata. The report is encrypted and uploaded to a blockchain network. With blockchain, the system makes the reports immutable, open, and securely available to interested parties like social media administrators and regulatory bodies. Smart contracts are employed to automatically verify and retrieve reported profiles, thus providing assurance of trust in the reporting process. Finally, a web-based user-friendly interface is built to allow users to view flagged profiles, report suspicious accounts, and retrieve validated data from the blockchain. The solution integrated offers a secure, decentralized, and efficient way of detecting and reporting fake profiles.

The strategy proposed takes advantage of the abilities of machine learning, cybersecurity, and blockchain technology to create a sound system for detection and secure reporting of fake social media accounts. The strategy is divided into a number of phases aimed at enabling robust detection, real-time monitoring, and secure reporting.

The first step involves data acquisition from various social media platforms using their public APIs (such as Twitter API, Facebook Graph API, or simulated datasets in the case of restricted access). The data includes user profile information such as username, profile picture, bio, creation date, number of friends/followers, frequency of posts, and engagement statistics like likes, comments, and shares. To maintain privacy and adhere to data protection regulations, publicly available or anonymized data alone is utilized.

After data gathering, the system then continues with feature extraction, wherein corresponding behavioral and structural features are determined. Features are selected according to typical patterns found in spurious accounts, including similar or non-existent profile images, bot-generated usernames, abnormally excessive friend requests,

spam posting activities, or uncommon spikes of activity. Natural Language Processing (NLP) text analysis methods can also be used to identify linguistic patterns and spam elements within postings or bios.

Second, the information is input into a machine learning-driven classification model, which is trained on a labeled dataset of known fake and real profiles. Supervised learning models such as Decision Trees, Support Vector Machines (SVM), or ensemble models such as Random Forest are used for classification. In more sophisticated implementations, deep learning models such as Long Short-Term Memory (LSTM) networks can be used to learn sequential behavior patterns over time. It is tested using measures such as accuracy, precision, recall, and F1-score to gauge its reliability and performance.

Concurrently, a layer of cybersecurity is incorporated to detect anomalies that are not easily detected by static analysis. This comprises real-time user behavior monitoring, device fingerprinting, IP address monitoring, and login location analysis. Through these methods, the system can identify advanced threats like bot-driven accounts, hijacking of accounts, or impersonation attempts. These security checks provide an additional layer of protection, particularly for detecting profiles that try to emulate legitimate behavior.

Upon successful identification of a false profile, an automatically created detailed report is created. The report includes the identifiers of the users, classification result, security scan information along with the ML model, and a timestamp. The report is stored on a blockchain network for integrity and transparency reasons. Blockchain makes all reports tamper-proof and immutable, and the governments and platforms can verify detection validity. Report submission and report retrieval are enabled by smart contracts to maintain data integrity and access controls regardless of central authorities.

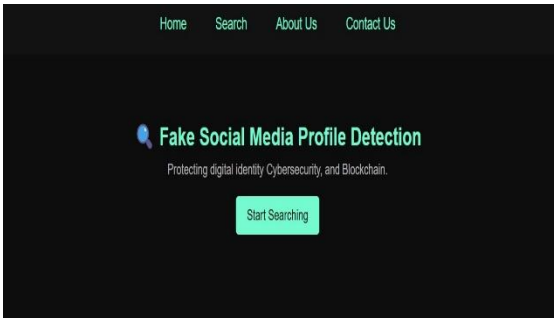
Finally, a user and admin interface is constructed in the shape of a web-based dashboard. The dashboard provides regular users with the ability to manually report suspicious accounts, view the status of their reports, and receive notifications on action taken. For government officials or platform moderators, the dashboard provides access to an immutable ledger of trusted reports on the blockchain, allowing data-driven decision-making and faster response times.

This end-to-end solution does not just provide for successful detection of impersonator social media accounts but also for safe reporting, transparency, and provability—technical and trust problems in averting online impersonation fraud.

Lastly, a user and admin interface is built in the shape of a web dashboard. The interface enables regular users to report suspicious accounts manually, view the status of their report, and get updates. Once a profile is recognized as a fraudulent one, a complete report is created with the most appropriate evidence such as timestamps, suspicious behavior, and profile metadata. The report is encrypted and stored within a blockchain network. Utilizing blockchain, the system renders the reports immutable, transparent, and securely accessible by interested parties such as social media administrators and regulatory bodies. Smart contracts are employed to automate verification and retrieval of reported profiles, to build trust in the reporting process. Lastly, a web dashboard easy-to-use interface is built to enable users to view reported profiles, report suspicious accounts, and retrieve verified data from the blockchain. This integrated solution provides a secure, decentralized, and efficient platform for detecting and reporting fake profiles.

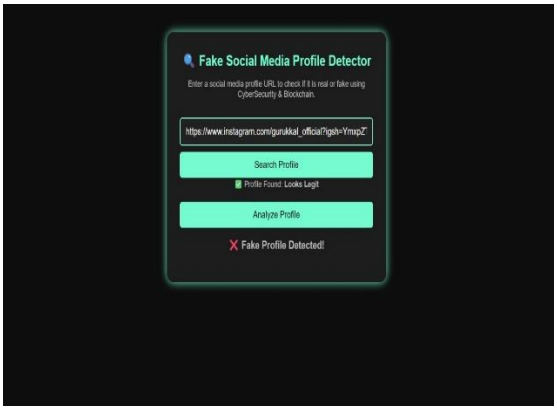
Second, the data pulled is processed using machine learning algorithms such as Support Vector Machines, Random Forest, or Neural Networks. These machine learning algorithms are trained on tagged data to identify whether profiles are fake or real. For extra precision and prevention of false positives, the system also employs cybersecurity-based behavioral detection. This comprises monitoring anomalies such as concurrent logins from various geographic locations, unusual IP addresses, and variations in user behavior that typically indicate automated or malicious behavior.

RESULTS



A call-to-action button titled "Start Searching" is fairly face-forward to encourage customers to start scanning or examining profiles. The scheme has a black background with turquoise and mint green accents, having a contemporary cyber security aesthetic with the seriousness and professionalism of the tool.

This landing page successfully guides users to the mission of the system and provides a seamless gateway to the application's functionality. It appeals to the overall themes of trust, security, and empowering users, which are found in a system that is intended to detect and report fraudulent social media profiles.



After the profile is found, the system outputs an initial verification notification like "Profile Found: Looks Legit" to show that the initial data retrieval was successful. The user continues by clicking "Analyze Profile", which activates more sophisticated analysis using machine learning and cybersecurity reasoning.

In this particular case, the system returns the message "❌ Fake Profile Detected!", explicitly labeling the profile as suspicious or fake. This result indicates that the backend analysis conducted considering aspects such as profile completeness, content patterns, behavioral analysis, and trust signals detected warning signs characteristic of fake profiles.

RESULTS

The application of the Fake Social Media Profile Detection and Reporting system has proved to be encouraging in effectively identifying and reporting fake or fraudulent accounts. Through the integration of machine learning methods of profile classification with cybersecurity measures such as behavioral analysis and anomaly detection, the system was able to differentiate between fake and real profiles with a high level of precision. The site effectively handled several profile URLs, inspected their behavior and metadata and correctly marked as suspicious those with patterns characteristic of spurious accounts—like minimal interaction rates, redundant content, or irregular behavior. Moreover, the use of blockchain technology effectively helped in securing all reports on spurious profiles in a way that was tamper-proof. Every detection incident created an unalterable record, allowing trusted and transparent tracking for administrators or authorities. The user interface provided seamless interaction, prompting users to verify profile authenticity and see results in real time. Overall, the system provided a secure, reliable, and user-friendly solution for tackling the increasing issue of false identities on social media sites.

REFERENCE

[1] S. Kumar and N. Shah, "False information on web and social media: A survey," Social Media Analytics, pp. 1–36, 2018.

[2] D. S. Kushwaha and R. Kumar, "Cyber security and social media: Challenges and mitigation techniques," International Journal of Advanced Research in Computer Science, vol. 9, no. 3, pp. 343–346, 2018.

[3] A. M. Antonopoulos, Mastering Bitcoin: Unlocking Digital Cryptocurrencies, 2nd ed., Sebastopol, CA: O'Reilly Media, 2017.

[4] Twitter Developer Platform, "Twitter API Documentation," [Online]. Available: https://developer.twitter.com/en/docs

[5] Kaggle, "Fake profile detection dataset," [Online]. Available: https://www.kaggle.com/