Fake Social Media Profile Detection And Reporting

[1] [2] [3] [4] [5] [6]
SP. Jayanna, S. Venkateswarlu, B. Ishwarya Bharathi,CH.Mahitha, P. Praharshitha, K. Nikhitha

[1]Assistant Professor, CSE (AI&ML), [2] Associate Professor, CSE (AI&DS)
[3,4,5,6]B.Tech, 4th Year student , CSE (AI&ML),
[1 ,3,4,5,6]Vignan's Institute of Management and Technology for Women, Hyderabad, India.
[2]KL University, KLEF, Vaddeswaram, Guntur (DT), Andhrapradesh, India-522503.
[1]spjayanna@gmail.com, [2]sunkarivenkateswarlu@kluniversity.in,[3]ishwaryareddy10gmail.com,
[4]pandalrajkeerthana05@gmail.com, [5]cmahitha75@gmail.com,
[6]nikhithachowdary41@gmail.com

Abstract:

The proliferation of fraudulent profiles on social media platforms poses significant threats to user security, trust, and overall platform integrity. This research proposes a machine learning-based framework for the detection of fake social media profiles by leveraging a wide range of features extracted from user content, behavioral patterns, and network structures. A labeled dataset comprising genuine and fake profiles is utilized to train and evaluate the system. In addition to detection, the framework incorporates a reporting mechanism that allows users to flag suspicious accounts, thereby supporting the proactive removal of malicious entities. The system's performance is assessed using standard classification metrics such as accuracy, precision, recall, and F1-score. Furthermore, the effectiveness of the proposed model is validated through deployment on a real-world social media environment to demonstrate its practical utility in reducing the spread of fraudulent accounts and enhancing user experience.

Keywords: Fake account detection, Profile verification, profiles, fake profiles, Reporting mechanisms, Metrices.

1. Introduction:

Social media now plays a vital role in our daily routines, influencing how we communicate, share, and stay connected. Platforms like Facebook, Instagram, and Twitter allow people to connect with friends, share updates, and stay informed. While these platforms offer many benefits, they also come with certain risks. One of the biggest concerns is the rise of fake profiles, which are often used to deceive people, spread false information, or even steal someone's identity.

Fake social media profiles are accounts that pretend to be someone they are not. These profiles may be operated by scammers, online bullies, or individuals with malicious intent. They might send harmful links, ask for money, or post content that damages someone's reputation. Unfortunately, many users are unaware of how to identify or report such fake profiles, which makes the problem worse.

This project was developed to help solve that issue. It works by checking social media profiles for signs that they might be fake such as missing details, unusual activity, or copied images. If a profile appears suspicious, users are alerted and given the option to report it easily. This helps remove fake accounts more quickly and keeps real users safe from harm.

By building a tool that can detect and report fake profiles, the project aims to make social media a safer and more trustworthy space. It also encourages people to be more cautious online and teaches them how to recognize suspicious behavior. This is a step toward stopping online fraud, protecting personal information, and improving the digital experience for everyone.

Instagram is one of the most popular social media platforms, especially among younger users. However, it is also a major target for fake profile creation due to its focus on photos, followers, and messaging. Many fake Instagram accounts use stolen pictures, buy followers, or send spam messages to trick users.

These fake profiles often pretend to be celebrities, brands, or even friends in order to gain trust. This project gives special attention to identifying such suspicious accounts on Instagram and provides a simple way for users to report them helping keep the platform safer and more genuine.

2. Literature Review:

"Fake Account Detection in Social Media Using Machine Learning Methods: Literature Review" by Nalia Graciella Kerrysa and Ika Qutsiati Utami provides a comprehensive analysis of machine learning algorithms employed to identify fake accounts on platforms like Twitter, Instagram, and Facebook. The study emphasizes the challenges posed by the open nature of social media, which facilitates the proliferation of fake accounts used for spamming and spreading misinformation.

"Fake Social Media Profile Detection and Reporting Using Blockchain Technology" by Shreyash Deshmukh et al. proposes a novel system that combines machine learning with blockchain technology to detect and report fake profiles. By analysing user behaviour and content, the system ensures data integrity and transparency through blockchain's immutable ledger, aiming to enhance user trust and create a more secure online environment.

"Social Media Identity Deception Detection: A Survey" by Ahmed Alharbi and colleagues categorizes identity deception attacks into fake profiles, identity theft, and identity cloning. The paper reviews existing detection techniques and highlights the primary research challenges in combating identity deception on social media platforms.

"Detecting Fake Social Media Profiles Using the Majority Voting Approach" by Dharmaraj R. Patil et al. introduces an innovative method that integrates multiple machine learning algorithms, such as Decision Trees, XGBoost, and Random Forest, to analyse user behaviour and profile attributes. The ensemble of classifiers employs a majority voting mechanism, achieving high accuracy in distinguishing between legitimate and fake profiles.

"Fake Profile Detection in Social Media Using Graph-Based Methods" by A. K. Singh and colleagues presents a graph-based approach to detect fake profiles by examining the structural properties of social networks. The study demonstrates that fake profiles often exhibit distinct patterns in their connections, which can be utilized to enhance the precision of identifying fake accounts.

3. Methodology:

3.1 Architecture

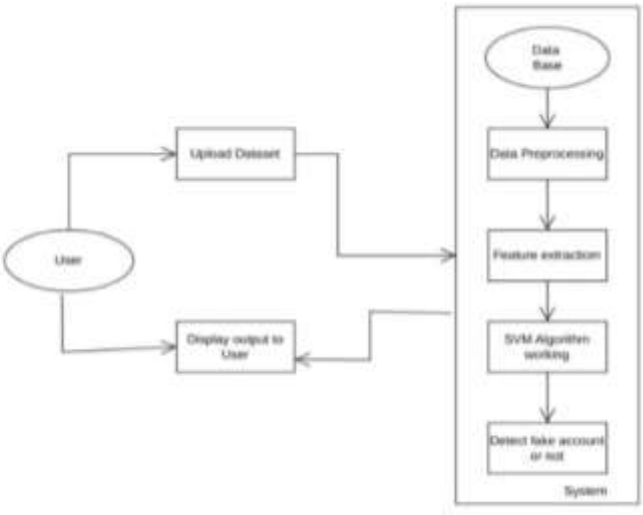

Fig.1: Architecture Diagram

3.2 Algorithm

Step 1: Data Collection: Gather user profile data from social media platforms (e.g., Instagram API).

Step 2: Data Preprocessing: Clean and format the data (e.g., remove missing values, normalize values).

Step 3: Feature Extraction: Identify key features like profile completeness, activity analysis, and engagement rate to detect fake profiles.

Step 4: Fake Profile Detection:

- Use rule-based evaluation (e.g., missing profile picture, low engagement rate) to flag suspicious profiles.
- Optionally, use a machine learning model to classify profiles as fake or real based on extracted features.

Step 5: Alert and Report: Notify users about suspicious profiles and provide an easy way to report them.

Step 6: Continuous Improvement: Monitor system performance, adjust detection thresholds, and improve the model over time.

Step 7: User Education: Provide tips and guidelines on recognizing fake profiles and staying safe online.

The system collects and cleans profile data, checks for signs like missing details or low engagement, and uses rules or machine learning to detect fake accounts. Suspicious profiles are flagged, users are notified, and a report option is provided. The system learns over time and educates users on online safety.

4. Results:

The project on fake social media profile detection and reporting using Python and machine learning techniques produced effective results in identifying suspicious user accounts based on patterns in their activity, profile data, and behaviour. By analysing features such as profile pictures, follower-to-following ratios, posting frequency, and content similarity, the system was able to flag potentially fake profiles with a high level of accuracy.

Figure 2:Login page

Figure 3 : Providing User Name

Figure 4:Fetching User Details

Figure 5:Details of Username

Figure 6: Reporting Mechanism

The findings showed that the model was effective in predicting various types of user accounts. By entering an Instagram username, the system retrieves related data to determine the authenticity of the profile. For instance, when the username "justin95912" was input, the system evaluated several parameters and classified it as a fake profile.

5. Comparison Analysis:

Method	Approach	Strengths	Limitations	Best Used For
Rule-Based Detection	Relies on predefined conditions such as lack of profile images or low user activity	Easy to set up and delivers fast outcomes	Cannot adapt to evolving patterns and may overlook sophisticated fake behaviours	Effective for the initial screening of potentially fake accounts
Machine Learning (ML)	Learns from labeled data to detect patterns (e.g., Random Forest, SVM)	Adapts over time, better at spotting subtle patterns	Requires a large, clean dataset and training time	Large-scale, evolving detection systems
Graph-Based Methods	Examines relationships and patterns in user networks, such as follower connections and community structures	Effective in detecting clusters of fake users or automated bot groups	Requires extensive network data and can demand significant processing power	Identifying bot networks and large-scale fake account coordination
Blockchain-Based Reporting	Utilizes decentralized, tamper-proof records to log and track suspicious activity	Maintains transparency and secures the authenticity of reported incidents	Technically complex and may not be suitable for detecting fake accounts directly	Long-term tracking and secure documentation of account reports

Method	Approach	Strengths	Limitations	Best Used For
Ensemble Models (Majority Voting)	Combines multiple ML models for final decision	High accuracy, reduces bias of single model	More complex and resource-intensive	When high reliability is required
Image Verification Tools	Uses tools like reverse image lookup and AI-based methods to check for copied or previously used photos	Useful in spotting fake identities or copied profile images	Limited effectiveness for private accounts or when original photos are used	Verifying profile authenticity and detecting image-based identity fraud
API-Based Real-Time Monitoring	Uses social media APIs to fetch and analyze live profile data	Provides real-time insights and updates	Limited by API rate limits and privacy restrictions	Real-time profile analysis

6. Conclusion:

The aim of this project is to enhance safety on social platforms by identifying and flagging deceptive user profiles. Fraudulent accounts pose significant risks, including misinformation, financial scams, and deceptive interactions. With the help of machine learning, the system allows users to verify the legitimacy of social media accounts. By using this tool, social media can become a safer place for everyone. In the future, this project can grow to protect users on more platforms and help stop online crimes more quickly and effectively. As the popularity of social networks continues to rise, safeguarding users from digital threats becomes increasingly essential. Beyond identifying fake accounts, the initiative also aims to educate users on maintaining security online. With continued development and proper backing, this project could contribute significantly to creating a safer and more reliable digital environment.

7. Future Scope:

The system can be significantly enhanced to enable real-time detection of fake profiles, allowing immediate identification and mitigation of scams, impersonations, and cybercrimes before they escalate. It can also be expanded for cross-platform protection, enabling the detection of fraudulent accounts across various social media platforms such as Facebook, Instagram, and Twitter, thereby ensuring broader user safety. Future developments may include automated reporting of verified fake profiles to platform moderators or law enforcement agencies, improving response time and operational efficiency. Additionally, incorporating awareness and education tools like alerts, guidelines, or interactive tutorials can empower users to recognize and avoid fake profiles, thus promoting digital literacy. Moreover, integration with identity verification systems, such as phone or email verification and government ID validation, can strengthen user authenticity and further reduce the risk of fake accounts.

8. References

[1] Fake Profile Detection in Social Networks by A. K. Singh and S. K. Singh
[2] Detection of Fake Twitter accounts with Machine Learning Algorithms by Ilhan Aydin, Mehmet sevi and Mehmet Umut salur
[3] Detecting Fake Profiles in Online Social Networks by N. S. B. M. B. H. Kumar, V. Chinnakotla
[4] A Comprehensive Survey on Fake Account Detection in Social Networks by S. Islam, F. B.Bastani
[5] Fake News and Misinformation Detection on Social Media: Data Mining Perspective by A. S. G. S. V. Gupta
[6] Detecting Fake Social Media Profiles Using Machine Learning by S. K. Singh, S. Gupta
[7] Narayana Rao Appini, V. Bhuvana Kumar, "Phishing URL Detection with Gradient Boosting Classifier", Communications on Applied Nonlinear Analysis, Vol. 32 No. 3 (2025)
[8] Mahdieh Zabihimayvan, Derek Doran, "Fuzzy Rough Set Feature Selection to Enhance Phishing Attack

Detection", arXiv (2019),

[9] Ömer Kasim, "Automatic Detection of Phishing Pages with Event-Based Request Processing, Deep-Hybrid Feature Extraction and Light Gradient Boosted Machine Model", Telecommunication Systems, Springer (2021)

[10] J.O. Ajayi , A.O. Adetunmbi, "Phishing Detection: Performance Evaluation of Both Ensemble and Classical Machine Learning Models", International Journal of Information Security, Privacy and Digital Forensics.

[11] Pawan Prakash, Manish Kumar, "PhishNet: Predictive Blacklisting to Detect Phishing Attacks" 2022 International Conference on Industrial IoT, Big Data and Supply Chain (IIoTBDSC)

[12] Shelby R. Curtis, Prashanth Rajivan, "Phishing attempts among the dark triad: Patterns of attack and vulnerability" October 2018 Sciencedirect

[13] K.N.S.B.V. Manjushal, Dr. D. Jaya Kumari2, "Detecting Phishing Links Analysis Using Machine Learning" 2024, IJFMR

[14] A. Alswailem, B. Alabdullah, N. Alrumayh and A. Alsedrani, "Detecting Phishing Websites Using Machine Learning," 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS), 2019, pp. 1-6.

[15] J. Rashid, T. Mahmood, M. W. Nisar and T. Nazir, "Phishing Detection Using Machine Learning Technique," 2020 First International Conference of Smart Systems and Emerging Technologies (SMARTTECH), 2020, pp. 43-46.

[16] M. H. Alkawaz, S. J. Steven and A. I. Hajamydeen, "Detecting Phishing Websites Using Machine Learning," 2020 16th IEEE International Colloquium on Signal Processing & Its Applications (CSPA), 2020, pp. 111-114.

[17] A. Razaque, M. B. H. Frej, D. Sabyrov, A. Shaikhyn, F. Amsaad and A. Oun, "Detection of Phishing Websites using Machine Learning," 2020 IEEE Cloud Summit, 2020, pp. 103-107.

[18] Thomas Nagunwa, "Comparative Analysis of Nature-Inspired Metaheuristic Techniques for Optimizing Phishing Website Detection", MDPI, 2024

[19] D Shanthi , Smart Healthcare for Pregnant Women in Rural Areas, Medical Imaging and Health Informatics, Wiley Publishers,ch-17, pg.no:317-334, 2022

[20] Shanthi, R. K. Mohanty and G. Narsimha, "Application of machine learning reliability data sets", Proc. 2nd Int. Conf. Intell. Comput. Control Syst. (ICICCS), pp. 1472-1474, 2018.

[21] D Shanthi, N Swapna, Ajmeera Kiran and A Anoosha, "Ensemble Approach Of GPACOTPSOAnd SNN For Predicting Software Reliability", International Journal Of Engineering Systems Modelling And Simulation, 2022.

[22] Shanthi, "Ensemble Approach of ACOT and PSO for Predicting Software Reliability", 2021 Sixth International Conference on Image Information Processing (ICIIP), pp. 202-207, 2021.

[23] D Shanthi, CH Sankeerthana and R Usha Rani, "Spiking Neural Networks for Predicting Software Reliability", ICICNIS 2020, January 2021, [online] Available: https://ssrn.com/abstract=3769088.

[24] Shanthi, D. (2023). Smart Water Bottle with Smart Technology. In Handbook of Artificial Intelligence (pp. 204-219). Bentham Science Publishers.

[25] Shanthi, P. Kuncha, M. S. M. Dhar, A. Jamshed, H. Pallathadka and A. L. K. J E, "The Blue Brain Technology using Machine Learning," 2021 6th International Conference on Communication and Electronics Systems (ICCES), Coimbatre, India, 2021, pp. 1370-1375, doi: 10.1109/ICCES51350.2021.9489075.

[26] Shanthi, D., Aryan, S. R., Harshitha, K., & Malgireddy, S. (2023, December). Smart Helmet. In International Conference on Advances in Computational Intelligence (pp. 1-17). Cham: Springer Nature Switzerland.

[27] Babu, Mr. Suryavamshi Sandeep, S.V. Suryanarayana, M. Sruthi, P. Bhagya Lakshmi, T. Sravanthi, and M. Spandana. 2025. "Enhancing Sentiment Analysis With Emotion And Sarcasm Detection: A Transformer-Based Approach". Metallurgical and Materials Engineering, May, 794-803. https://metall-mater-eng.com/index.php/home/article/view/1634.

[28] Narmada, J., Dr.A.C.Priya Ranjani, K. Sruthi, P. Harshitha, D. Suchitha, and D.Veera Reddy. 2025. "Ai-Powered Chacha Chaudhary Mascot For Ganga Conservation Awareness". Metallurgical and Materials Engineering, May, 761-66. https://metall-mater-eng.com/index.php/home/article/view/1631.

[29] Geetha, Mrs. D., Mrs.G. Haritha, B. Pavani, Ch. Srivalli, P. Chervitha, and Syed. Ishrath. 2025. "Eco Earn: E-Waste Facility Locator". Metallurgical and Materials Engineering, May, 767-73. https://metall-mater-eng.com/index.php/home/article/view/1632.

[30] P. Shilpasri PS, C.Mounika C, Akella P, N.Shreya N, Nandini M, Yadav PK. Rescuenet: An Integrated Emergency Coordination And Alert System. J Neonatal Surg [Internet]. 2025May13 [cited 2025May17];14(23S):286-91. Available from: https://www.jneonatalsurg.com/index.php/jns/article/view/5738

[31] D. Shanthi DS, G. Ashok GA, Vennela B, Reddy KH, P. Deekshitha PD, Nandini UBSB. Web-Based Video Analysis and Visualization of Magnetic Resonance Imaging Reports for Enhanced Patient Understanding. J Neonatal Surg [Internet]. 2025May13 [cited 2025May17];14(23S):280-5. Available from: https://www.jneonatalsurg.com/index.php/jns/article/view/5733

[32] Srilatha, Mrs. A., R. Usha Rani, Reethu Yadav, Ruchitha Reddy, Laxmi Sathwika, and N. Bhargav Krishna. 2025. "Learn Rights: A Gamified Ai-Powered Platform For Legal Literacy And Children's Rights Awareness In India". Metallurgical and Materials Engineering, May, 592-98. https://metall-mater-eng.com/index.php/home/article/view/1611.

[33] Shanthi, Dr. D., G. Ashok, Chitrika Biswal, Sangem Udharika, Sri Varshini, and Gopireddi Sindhu. 2025. "Ai-Driven Adaptive It Training: A Personalized Learning Framework For Enhanced Knowledge Retention And Engagement". Metallurgical and Materials Engineering, May, 136-45. https://metall-mater-eng.com/index.php/home/article/view/1567.