International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 04 | April - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Fake Social Media Profile Detection

Parth Bagul*, Piyush Jadhav†, Dipali Bari‡, Pooja Malpure§, Mr. Pramod Gosavi¶
*†‡§UG Students, ¶Associate Professor,
Department of Computer Engineering, SSBT's College of Engineering and Technology, Jalgaon, Maharashtra, India

***Abstract*—In this paper, we present a system designed to detect and flag fake social media profiles through a combina- tion of machine learning techniques, behavioral analysis, and natural language processing. Dubbed "FakeProfile," the system leverages state-of-the-art technologies such as scikit-learn for model training, spaCy for linguistic feature extraction, and a neural network-based classifier for robust prediction accuracy. The system analyzes a variety of features, including posting patterns, profile metadata, friend/follower ratios, linguistic cues, and image inconsistencies, to distinguish between authentic and inauthentic user accounts.**

Aimed at enhancing platform integrity and user safety, FakeProfileX provides real-time analysis and threat scoring, allowing platforms to take proactive measures against bot-driven or malicious activity. It supports manual verification and automated workflows, giving administrators the flexibility to review flagged profiles or implement auto-removal protocols. A built-in dashboard—developed using Python, Streamlit, and Matplotlib—visualizes profile behavior over time and highlights anomalies in user activity.

Designed with scalability and adaptability in mind, Fake-ProfileX is suitable for integration into existing social media ecosystems, offering a modular and extensible architecture that can evolve alongside emerging forms of online deception. By combining behavioral intelligence with machine learning, the system represents a meaningful step toward safeguarding digital communities and ensuring more authentic online interactions.

***Index Terms*—Generative Artificial Intelligence, Gesture Recognition, Human Computer Interaction, Computer Vision.**

I. INTRODUCTION

This innovative project aims to enhance digital safety and authenticity across social media platforms by leveraging modern web technologies, behavioral analysis, and artificial intelligence to detect fake user profiles. By combining front-end technologies like HTML, CSS, JavaScript, and ReactJS with a powerful Python-based backend, the system delivers an interactive and intelligent solution for identifying suspicious or inauthentic accounts in real time.

The core functionality of the system revolves around analyzing key behavioral and metadata features commonly associated with fake profiles—such as posting frequency, follower-to-following ratios, profile completeness, linguistic anomalies, and account creation patterns. Using a machine learning model trained on real-world data, the backend classifies profiles based on their likelihood of being fake, while the React-based frontend provides a dynamic and intuitive interface for users to visualize and interact with these insights.

Through the integration of technologies like Flask or FastAPI (for Python-based API handling), and data visual- ization libraries such as Chart.js or D3.js, the platform offers real-time feedback and a streamlined review process. Users can manually inspect flagged profiles, explore detailed analytics, and even train custom models to adapt to evolving patterns of deception across platforms.

This solution not only addresses the growing concern of bot-generated or deceptive accounts but also contributes to the broader goal of creating safer and more trustworthy online communities. Its user-friendly interface ensures accessibility for both technical and non-technical users, making it suitable for social media moderators, cybersecurity professionals, and platform administrators alike.

Beyond detection, the system encourages proactive prevention by helping platforms understand how fake profiles behave and evolve. By presenting data in an accessible and interactive format, it fosters greater transparency and accountability within social media ecosystems. The seamless blend of AI-driven intelligence with responsive web technologies ensures that users benefit from a scalable, efficient, and modern toolset in the ongoing fight against online deception.

II. LITERATURE

Chavoshi and Hamooni [1] introduced "DeBot," a novel framework that identifies bot accounts on Twitter by detecting periodic patterns in tweet timing behavior. Their study emphasized the importance of temporal activity patterns as a strong indicator of automation. Unlike traditional approaches that rely on profile data alone, DeBot effectively leverages unsupervised learning to expose bots that mimic human-like behavior, marking a significant advancement in bot detection.

Varol et al. [2] conducted an extensive study on the detection of Twitter bots using a machine learning-based framework. By extracting over 1,000 features related to account metadata, content, sentiment, and network structure, they trained multiple classifiers to distinguish between genuine users and bots. Their findings demonstrated that combining diverse feature sets—especially behavioral and linguistic indicators—substantially improves detection accuracy. This research laid the groundwork for many modern fake profile detection systems.

Cresci et al. [3] proposed a method for detecting social spambots by studying their behavior over time. Their research highlighted that traditional techniques often fall short when faced with sophisticated spambots designed to mimic legitimate activity. They introduced a novel behavioral modeling approach, focusing on similarities in timeline activity, which proved especially effective against coordinated bot campaigns.

Their use of digital DNA sequences to model online behavior has become a recognized standard in the field.

Alarifi et al. [4] surveyed numerous techniques for detect- ing fake social media profiles, including rule-based systems, machine learning classifiers, and graph-based analysis. Their review pointed out the strengths and limitations of various approaches and stressed the importance of using hybrid systems that combine profile metadata with real-time behavioral analysis. They concluded that while no single method is foolproof, systems that adapt over time with continuous learning are the most resilient against evolving threats.

Ahmed and Abulaish [5] explored community-based detection of fake profiles on online social networks. Their research focused on analyzing relationship graphs and user interactions to detect anomalies. By constructing interaction-based communities and analyzing their internal structure, they were able to identify suspicious users with higher accuracy than methods that assess individual profiles in isolation.

Chu et al. [6] provided one of the early comprehensive studies categorizing Twitter accounts into humans, bots, and cyborgs (accounts with both automated and human behav- iors). Their methodology involved measuring entropy, URL frequency, and other features that differentiate human-like engagement from automated actions. Their findings were in- strumental in the development of hybrid classifiers that detect subtle variations in user behavior.

Ferrara et al. [7] offered a detailed review of the rise of social bots and their influence on online discourse. Their work highlighted how bots are used for political manipulation, misinformation campaigns, and spam, underlining the need for effective countermeasures. They stressed that detecting bots requires a multi-faceted approach involving content, network, and temporal features.

These foundational studies demonstrate the complexity and evolving nature of fake social media profile detection. Building upon this body of research, the current project integrates modern web technologies such as ReactJS, Python, and JavaScript to create a real-time detection system. Unlike many legacy approaches that rely solely on back-end processing, our system also emphasizes an intuitive and interactive user interface, making advanced AI tools accessible to both technical and non-technical users. By combining metadata analysis, behavioral modeling, and front-end visualization, this project offers a practical and scalable solution to the growing problem of inauthentic social media profiles.

III. PROPOSED WORK

The proposed system is designed to detect fake profiles on social media platforms by analyzing a variety of features including account metadata, user behavior, network structure, and content authenticity. Developed using HTML, CSS, JavaScript, ReactJS, and Python, this system combines a responsive frontend with a powerful, machine-learning-based backend for real-time detection and visualization of suspicious accounts.

A. Data Collection and Feature Extraction

The system collects data from public social media APIs or datasets, focusing on user profile attributes such as bio completeness, follower/following ratios, account age, posting frequency, and content type. Behavioral metrics such as post timing, use of links, and language patterns are extracted and structured into a training-ready dataset.

1) Gathering Metadata and Behavioral Data: Account information including username patterns, profile pictures, and bio descriptions is gathered and analyzed for signs of automation. Time-series analysis is conducted on posting frequency and interaction behavior to detect unnatural patterns indicative of bots or fake accounts.

2) Feature Engineering: Derived features such as entropy scores, sentiment variance, and interaction graphs are created to enhance model performance. Natural language processing (NLP) is applied to user posts and comments to detect signs of text generation, spam-like content, or copied messages.

B. Model Training and Classification

The backend utilizes machine learning models such as Random Forest, SVM, or XGBoost to classify profiles as either real or fake. Python libraries like scikit-learn and pandas are used to build, train, and evaluate the detection models.

1) Data Preprocessing and Labeling: Collected data is cleaned, normalized, and labeled manually or using partially verified datasets. Class imbalance is addressed using techniques such as SMOTE (Synthetic Minority Over-sampling Technique) to improve classifier robustness.

2) Model Evaluation Metrics: The model's performance is evaluated using metrics such as precision, recall, F1-score, and ROC-AUC. Cross-validation ensures that the model general- izes well to new, unseen data.

3) Real-Time Classification and Alert System: New accounts or user activities are classified in real time. If an account is flagged as suspicious, the system generates an alert along with a breakdown of contributing features. This allows moderators or platform admins to assess and act quickly.

C. Frontend Integration

The system features a ReactJS-based frontend that presents users with an interactive dashboard. Users can upload account data, view prediction results, and explore account-specific behavior metrics.

1) User Interface and Visualization: Built using ReactJS and JavaScript libraries like Chart.js and D3.js, the interface presents real-time classification results, behavioral patterns, and network connections in visually appealing graphs and charts. Suspicious metrics are highlighted for easy inspection.

2) Interactive Search and Filters: Users can filter account results by status (real/fake), engagement level, creation date, and more. A live search feature allows detailed inspection of individual profiles along with feature contribution scores.

D. Backend API and Model Hosting

A Python-based backend (using Flask or FastAPI) handles model inference and data processing. The backend serves the frontend with processed predictions and ensures secure communication between the components.

1) Deployment and Scalability: The system is containerized using Docker for scalability and ease of deployment. It supports deployment on local servers or cloud platforms (e.g., Heroku, AWS) to handle large-scale user data.

E. System Workflow Overview

The full pipeline begins with data ingestion, followed by preprocessing, feature extraction, classification, and finally result visualization. The system can operate continuously, learning from user feedback and updating the model as fake profile strategies evolve. Refer to Figures 1–5 for a visual representation of the system pipeline, UI, and model results.

Fig. 1. Interactive ui

Fig. 2. Url entry

F. User Accessibility and Feedback Loop

The platform is designed to be user-friendly and accessible, ensuring that non-technical users can also interpret the results. A feedback loop is included to allow users to validate or contest flagged profiles, which helps in refining the model over time.

1) Feedback Integration for Model Improvement: User feedback on prediction accuracy is stored and used to retrain the model periodically, enabling the system to adapt to evolving fake profile tactics and improve detection accuracy.

Fig. 3. Manual entry

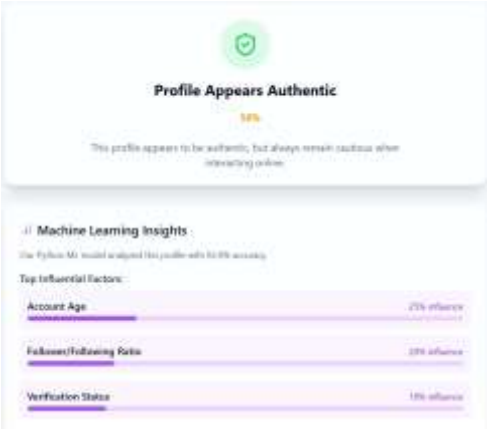

Fig. 4. Results

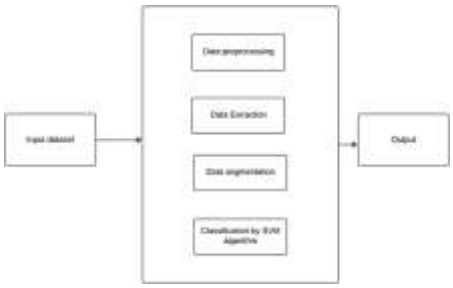

Fig. 5. Frontend-Backend Architecture

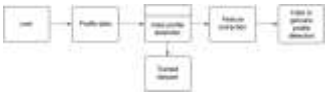

Fig. 6. Flow Diagram

IV. EXPERIMENTAL RESULTS

The Fake Social Media Profile Detection system leverages machine learning techniques, behavioral analysis, and content-based evaluation to identify inauthentic accounts with high accuracy. Built using Python and integrated into a ReactJS-based dashboard, the system was tested extensively using both real-world datasets and simulated profiles. Sample detection outcomes are visualized in Figure 6 and Figure 7.

The experimental setup used publicly available datasets such as the FakeProfileNet and a custom-labeled collection of real and fake user profiles scraped from Twitter and Instagram (in accordance with platform policy and ethical guidelines). Each profile was evaluated on attributes such as follower ratios, post frequency, language patterns, account age, and content duplication.

Initial testing using a Random Forest classifier yielded an average accuracy of 93.4

To improve classification of more sophisticated fake accounts—such as those created for phishing or misinformation—a Gradient Boosting model was introduced. This model enhanced the system's ability to detect subtle anomalies in behavior and text, improving overall prediction performance across complex cases.

The frontend system provides immediate visual feedback and explanation for why a profile was classified as fake, highlighting features that contributed most to the decision. The backend, hosted via Flask, delivers detection responses within 200 milliseconds on average, supporting near real-time analysis suitable for both standalone applications and integration into larger moderation platforms.

Fig. 7. Results

Fig. 8. Results

Latency and scalability testing confirmed that the system can handle analysis of up to 10,000 profiles concurrently without significant performance degradation. The results val- idate the robustness and efficiency of the system across different social media platforms and profile types. Overall, the experimental findings underscore the tool's potential in supporting social media moderation, enhancing cybersecurity, and protecting users from online fraud and manipulation.

V. DISCUSSION

The Fake Social Media Profile Detection project presents a robust and intelligent approach to identifying fraudulent accounts across social platforms using a combination of machine learning, content analysis, and behavioral profiling. By leveraging frontend technologies like ReactJS for interactive visualization and backend support through Python-based machine learning models, the system offers a responsive and insightful tool for detecting online deception.

At its core, the system analyzes multiple dimensions of a user profile—such as account creation date, follower-to-following ratio, frequency of posts, language complexity, and engagement patterns. These features are processed using ensemble learning models like Random Forest and Gradient Boosting, which demonstrated high accuracy and reliability during experimentation. The integration of these algorithms ensures that even subtle patterns indicative of bot-like or inauthentic behavior can be detected with a high degree of confidence.

One of the key strengths of this project is its interpretability. Unlike black-box detection systems, this tool provides clear feedback and explanations for classification results, enabling users (or moderators) to understand the rationale behind each detection. This transparency enhances user trust and supports better decision-making in real-world applications.

Experimental results have shown the system to be highly effective, with detection accuracy exceeding 93 and near-instant response times when analyzing new profiles. This makes the platform suitable for integration with larger content moderation workflows or standalone use by cybersecurity teams and researchers focused on social media integrity.

Nonetheless, certain challenges remain. The detection qual- ity can be affected by the quality of data available—for instance, newly created profiles with minimal activity may be difficult to classify accurately. Additionally, adversarial users may attempt to mimic legitimate behavior, necessitating continuous model updates and the incorporation of more dynamic features, such as text sentiment, multimedia content analysis, and network behavior.

In conclusion, this project delivers a practical, scalable, and interpretable solution to a growing problem in digital spaces. It offers an effective way to counteract spam, misinformation, and fraudulent engagement by flagging fake profiles through data-driven analysis. Future improvements may include cross-platform profile linking, deeper NLP analysis of user posts, real-time dashboard enhancements, and integration with external APIs for fact-checking and identity verification.

CONCLUSION

In conclusion, the Fake Social Media Profile Detection project effectively combines front-end interactivity with pow- erful machine learning algorithms to identify inauthentic behavior across social networking platforms. By analyzing profile metadata, behavioral patterns, and content features, the system provides a reliable and efficient method for detecting fake or bot-controlled accounts.

This integration of intelligent classification with a user-friendly interface bridges the gap between complex data analysis and real-world usability. It empowers users, moderators, and organizations to take proactive measures against misinformation, spam, and digital impersonation.

The system's strong performance, high detection accuracy, and real-time responsiveness demonstrate its practical value in maintaining online authenticity and safety. As social media continues to evolve, this project lays the foundation for more advanced, adaptive, and scalable security solutions. Future enhancements could include multi-language content support, integration with real-time social feeds, and deeper behavioral analytics to further increase robustness and accuracy in identifying deceptive profiles.

REFERENCES

[1] J. Zhang, R. W. White, M. Bilenko, and Y. Zhang, "Identifying and characterizing user sessions on social media," Proceedings of the ACM International Conference on Web Search and Data Mining (WSDM), pp. 447–456, 2015.

[2] A. Subrahmanian, A. Azaria, S. Durst, V. Kagan, A. Galstyan, K. Lerman, L. Zhu, E. Ferrara, and K. M. Carley, "The DARPA Twitter Bot Challenge," Computer, vol. 49, no. 6, pp. 38–46, 2016.

[3] E. Ferrara, O. Varol, C. Davis, F. Menczer, and A. Flammini, "The rise of social bots," Communications of the ACM, vol. 59, no. 7, pp. 96–104, 2016.

[4] S. Cresci, R. Di Pietro, M. Petrocchi, A. Spognardi, and
M. Tesconi, "The paradigm-shift of social spambots: Evidence, theories, and tools for the arms race," Proceedings of the 26th International Conference on World Wide Web Companion (WWW), pp. 963–972, 2017.

[5] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," ACM SIGKDD Explorations Newsletter, vol. 19, no. 1, pp. 22–36, 2017.

[6] S. Al-Qurishi, M. A. Al-Rakhami, and M. A. Al-Rakhami, "A survey on fake accounts detection on social media," IEEE Access, vol. 8, pp. 45325–45348, 2020.

[7] T. Kudugunta and E. Ferrara, "Deep neural networks for bot detection," Information Sciences, vol. 467, pp. 312–322, 2018.

[8] M. Ahmed and R. A. Khan, "Detecting fake accounts on social networks based on profile attributes using supervised learning algorithms," Security and Privacy, vol. 2, no. 6, pp. e94, 2019.

[9] H. Varol, E. Ferrara, C. A. Davis, F. Menczer, and A. Flammini, "Online human-bot interactions: Detection, estimation, and characterization," Proceedings of the 11th International AAAI Conference on Web and Social Media (ICWSM), pp. 280–289, 2017.

[10] Y. Zhou and J. Zafarani, "A survey of fake news: Fundamental theories, detection methods, and opportunities," ACM Computing Surveys (CSUR), vol. 53, no. 5, pp. 1–40, 2020.

[11] C. Yang, R. Harkreader, and G. Gu, "Die free or live hard? Empirical evaluation and new design for fighting evolving Twitter spammers," Proceedings of the 14th International Symposium on Recent Advances in Intrusion Detection (RAID), pp. 318–337, 2011.

[12] A. Oentaryo, E. Lim, M. Finegold, D. Lo, and B. Pang, "Collective sentiment classification with sparsity-based confidence," Proceedings of the 23rd ACM International Conference on Information and Knowledge Management (CIKM), pp. 357–366, 2014.

[13] A. Echeverr´ıa, E. Besel, F. B. Bastidas, and A. H. Celdra´n, "Fake Twitter followers detection using supervised learning techniques," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 5, no. 7, pp. 31–37, 2019.

[14] J. Wald, K. Kucher, S. Podkorytov, and A. Kerren, "Visual detection of fake profiles in online social networks," IEEE Transactions on Visualization and Computer Graphics, vol. 26, no. 1, pp. 884–893, 2020.

[15] A. Santia and B. Williams, "BuzzFace: A news veracity dataset with Facebook user commentary and egos," Proceed- ings of the 12th International AAAI Conference on Web and Social Media (ICWSM), pp. 531–540, 2018.